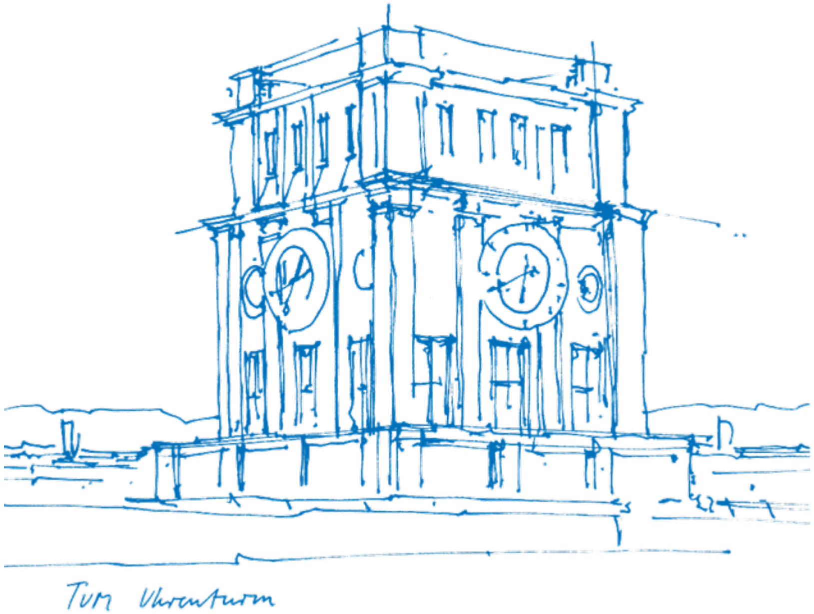
Cost-Aware Optimized Front-Door Experimental Design

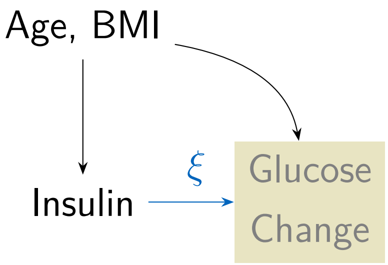
Leopold Mareis
Department of Mathematics
Technical University of Munich (TUM)

(joint work with Mathias Drton)



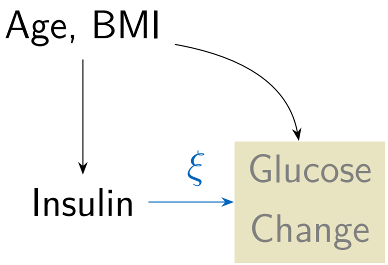
Selective Measurement in Clinical Trials

Example: Effect ξ of insulin on blood glucose in ICU patients (MIMIC-IV, $n = 999$).



Selective Measurement in Clinical Trials

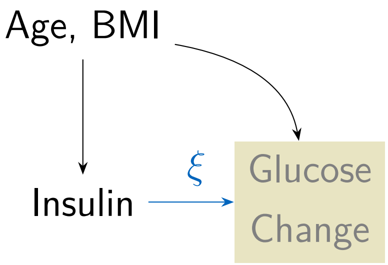
Example: Effect ξ of insulin on blood glucose in ICU patients (MIMIC-IV, $n = 999$).



Variable	Cost	Always?
Age, BMI	free	✓
Insulin dose	free	✓
Blood glucose change	**costly**	×

Selective Measurement in Clinical Trials

Example: Effect ξ of insulin on blood glucose in ICU patients (MIMIC-IV, $n = 999$).

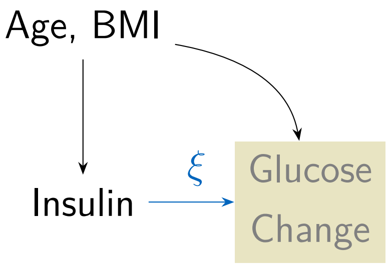


Variable	Cost	Always?
Age, BMI	free	✓
Insulin dose	free	✓
Blood glucose change	**costly**	×

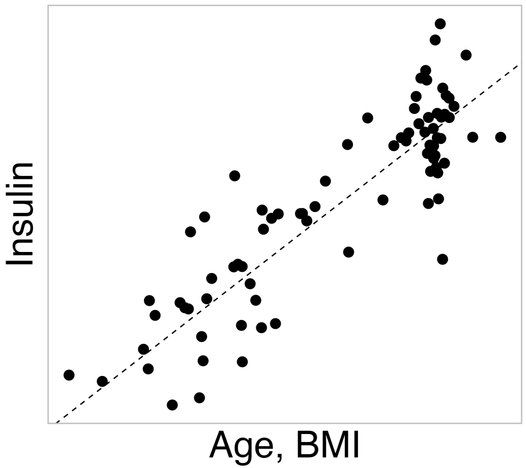
Inefficiency: Follow-up for all patients is expensive.

Selective Measurement in Clinical Trials

Example: Effect ξ of insulin on blood glucose in ICU patients (MIMIC-IV, $n = 999$).



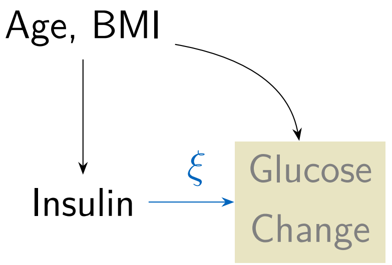
Variable	Cost	Always?
Age, BMI	free	✓
Insulin dose	free	✓
Blood glucose change	**costly**	×



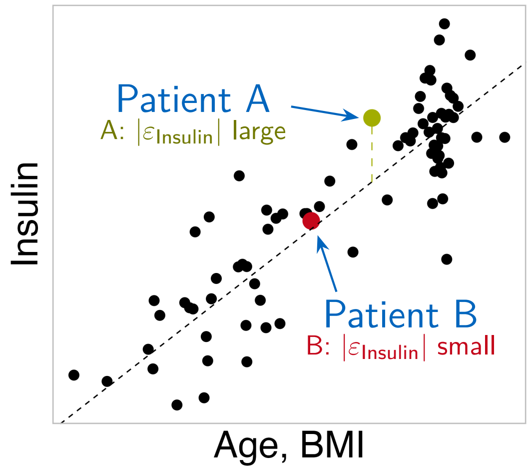
Inefficiency: Follow-up for all patients is expensive.

Selective Measurement in Clinical Trials

Example: Effect ξ of insulin on blood glucose in ICU patients (MIMIC-IV, $n = 999$).



Variable	Cost	Always?
Age, BMI	free	✓
Insulin dose	free	✓
Blood glucose change	**costly**	×

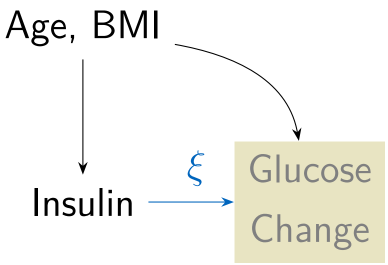


✓ measure X_{Glucose} × skip follow-up

Inefficiency: Follow-up for all patients is expensive.

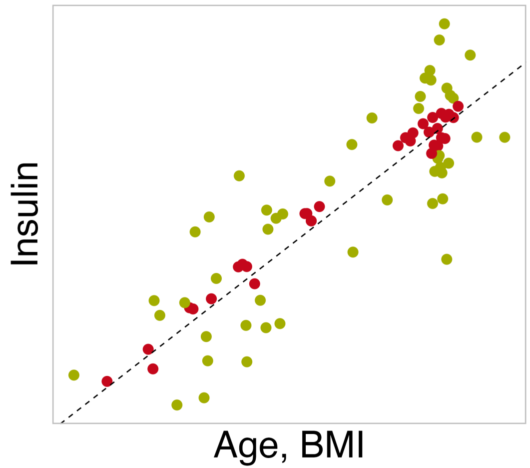
Selective Measurement in Clinical Trials

Example: Effect ξ of insulin on blood glucose in ICU patients (MIMIC-IV, $n = 999$).



Variable	Cost	Always?
Age, BMI	free	✓
Insulin dose	free	✓
Blood glucose change	**costly**	×

Inefficiency: Follow-up for all patients is expensive.

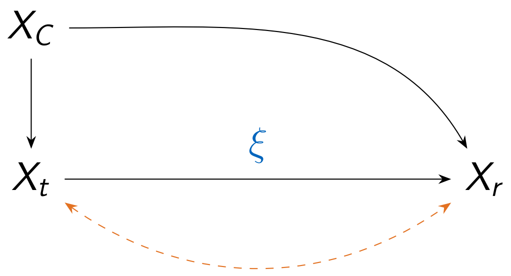


✓ measure X_{Glucose} × skip follow-up

$\Rightarrow$ Optimize sequential measurement decisions driven by propensity $\pi(X)$.

Front-Door Estimation

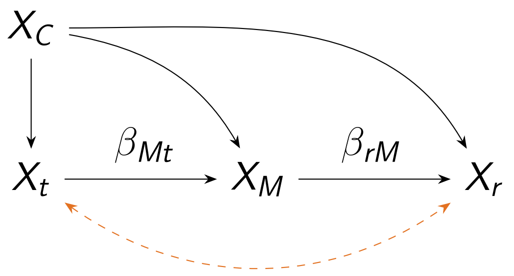
Goal: Estimate the causal effect ξ of a treatment X_t on a response X_r under confounding.



Front-Door Estimation

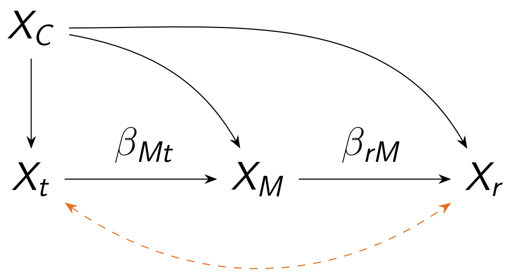
Goal: Estimate the causal effect ξ of a treatment X_t on a response X_r under confounding.

Fix: Mediator X_M



Front-Door Estimation

Goal: Estimate the causal effect ξ of a treatment X_t on a response X_r under confounding.



Fix: Mediator X_M

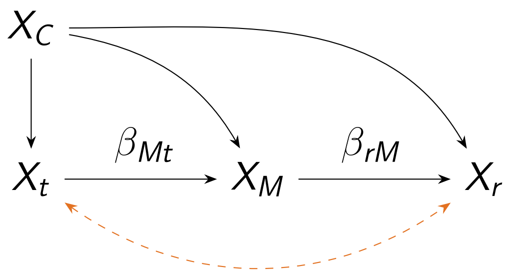
SEM: Full data $X = (X_C, X_t, X_M, X_r)$ solves

$$\begin{pmatrix} X_C \\ X_t \\ X_M \\ X_r \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{tC} & 0 & 0 & 0 \\ \beta_{MC} & \beta_{Mt} & 0 & 0 \\ \beta_{rC} & 0 & \beta_{rM} & 0 \end{pmatrix} \begin{pmatrix} X_C \\ X_t \\ X_M \\ X_r \end{pmatrix} + \begin{pmatrix} \varepsilon_C \\ \varepsilon_t \\ \varepsilon_M \\ \varepsilon_r \end{pmatrix},$$

$$\mathcal{P}_\varepsilon = \mathcal{P}_{\varepsilon_C} \otimes \mathcal{P}_{(\varepsilon_t, \varepsilon_r)} \otimes \mathcal{P}_{\varepsilon_M}.$$

Front-Door Estimation

Goal: Estimate the causal effect ξ of a treatment X_t on a response X_r under confounding.



Fix: Mediator X_M

SEM: Full data $X = (X_C, X_t, X_M, X_r)$ solves

$$\begin{pmatrix} X_C \\ X_t \\ X_M \\ X_r \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{tC} & 0 & 0 & 0 \\ \beta_{MC} & \beta_{Mt} & 0 & 0 \\ \beta_{rC} & 0 & \beta_{rM} & 0 \end{pmatrix} \begin{pmatrix} X_C \\ X_t \\ X_M \\ X_r \end{pmatrix} + \begin{pmatrix} \varepsilon_C \\ \varepsilon_t \\ \varepsilon_M \\ \varepsilon_r \end{pmatrix},$$

$$\mathcal{P}_\varepsilon = \mathcal{P}_{\varepsilon_C} \otimes \mathcal{P}_{(\varepsilon_t, \varepsilon_r)} \otimes \mathcal{P}_{\varepsilon_M}.$$

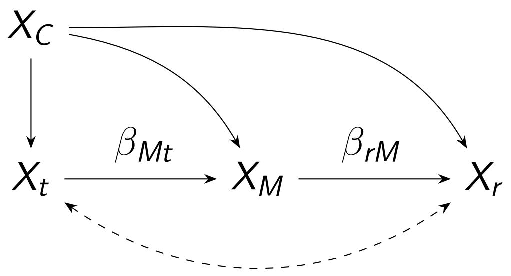
Identification:

$$\xi = \beta_{Mt}\beta_{rM}$$

$$\beta_{Mt} = \Sigma_{Mt.C}/\Sigma_{tt.C}$$

$$\beta_{rM} = \Sigma_{rM.C}\Sigma_{MM.C}^{-1}$$

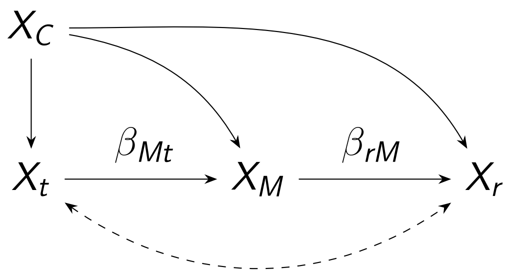
Partial Observation Model $\mathcal{M}_\pi$



Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

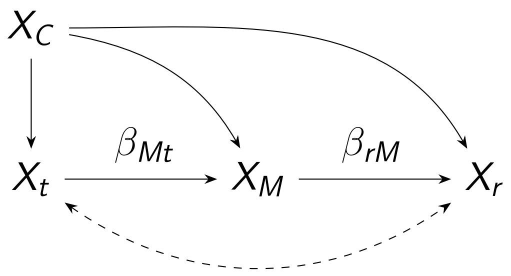
Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	−0.7	·
∞	1.2	0.8	1.5	−0.3
1	−1.1	0.2	·	·
2	1.4	−0.9	0.6	·



Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	-0.7	·
∞	1.2	0.8	1.5	-0.3
1	-1.1	0.2	·	·
2	1.4	-0.9	0.6	·

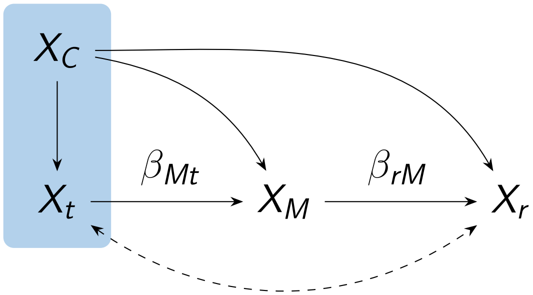


Δ	$G_\Delta(X)$	Cost
1	X_{Ct}	c_0
2	X_{CtM}	$c_0 + c_1$
∞	X_{CtMr}	$c_0 + c_1 + c_2$

Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	-0.7	·
∞	1.2	0.8	1.5	-0.3
1	-1.1	0.2	·	·
2	1.4	-0.9	0.6	·

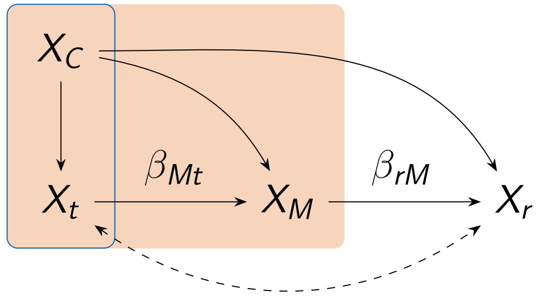


Δ	$G_\Delta(X)$	Cost
1	X_{Ct}	c_0
2	X_{CtM}	$c_0 + c_1$
∞	X_{CtMr}	$c_0 + c_1 + c_2$

Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	−0.7	·
∞	1.2	0.8	1.5	−0.3
1	−1.1	0.2	·	·
2	1.4	−0.9	0.6	·



Measure X_M with probability $\pi_1(X_{Ct})$

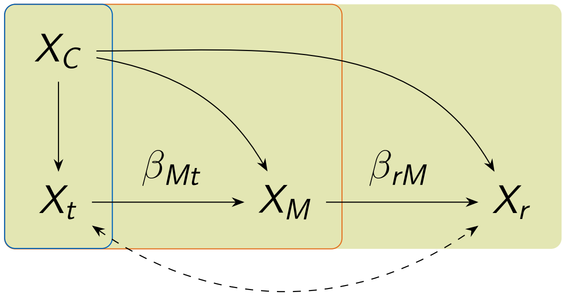
π_1

Δ	$G_\Delta(X)$	Cost
1	X_{Ct}	c_0
2	X_{CtM}	$c_0 + c_1$
∞	X_{CtMr}	$c_0 + c_1 + c_2$

Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	−0.7	·
∞	1.2	0.8	1.5	−0.3
1	−1.1	0.2	·	·
2	1.4	−0.9	0.6	·



Measure X_r with probability $\pi_2(X_{CtM})$

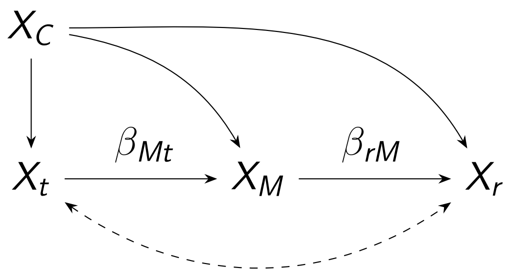
π_1 π_2

Δ	$G_\Delta(X)$	Cost
1	X_{Ct}	c_0
2	X_{CtM}	$c_0 + c_1$
∞	X_{CtMr}	$c_0 + c_1 + c_2$

Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	-0.7	$\cdot$
∞	1.2	0.8	1.5	-0.3
1	-1.1	0.2	$\cdot$	$\cdot$
2	1.4	-0.9	0.6	$\cdot$



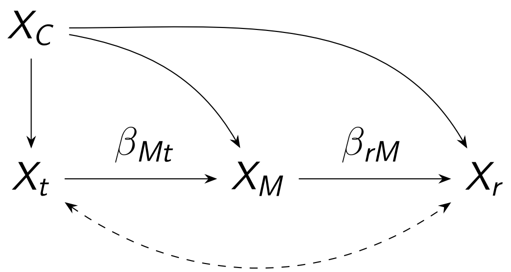
Δ	$G_\Delta(X)$	Cost
1	X_{Ct}	c_0
2	X_{CtM}	$c_0 + c_1$
∞	X_{CtMr}	$c_0 + c_1 + c_2$

π_1
π_2

Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	−0.7	·
∞	1.2	0.8	1.5	−0.3
1	−1.1	0.2	·	·
2	1.4	−0.9	0.6	·



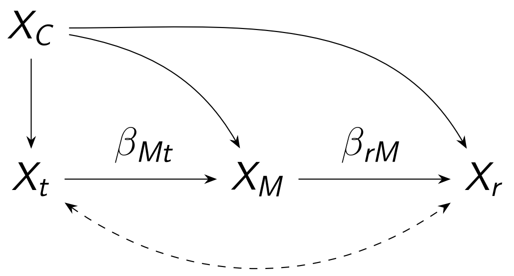
- **Model class $\mathcal{M}_\pi$:** Observed data distributions of $(\Delta, G_\Delta(X))$ governed by $\pi = (\pi_1, \pi_2)$.
- **Full-data model $\mathcal{M}_1$:** $\pi \equiv (1, 1)$, so $G_\Delta(X) \equiv X_{CtMr}$.

	Δ	$G_\Delta(X)$	Cost
π_1	1	X_{Ct}	c_0
π_2	2	X_{CtM}	$c_0 + c_1$
	∞	X_{CtMr}	$c_0 + c_1 + c_2$

Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	-0.7	·
∞	1.2	0.8	1.5	-0.3
1	-1.1	0.2	·	·
2	1.4	-0.9	0.6	·



- **Model class $\mathcal{M}_\pi$:** Observed data distributions of $(\Delta, G_\Delta(X))$ governed by $\pi = (\pi_1, \pi_2)$.
- **Full-data model $\mathcal{M}_1$:** $\pi \equiv (1,1)$, so $G_\Delta(X) \equiv X_{CtMr}$.

	Δ	$G_\Delta(X)$	Cost
π_1	1	X_{Ct}	c_0
π_2	2	X_{CtM}	$\mathrm{c}_0 + c_1$
	∞	X_{CtMr}	$\mathrm{c}_0 + c_1 + c_2$

Problem:

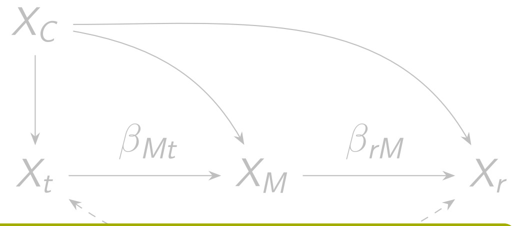
$$\min_{\pi=(\pi_1,\pi_2),\ \mathrm{RAL}\ \hat{\xi}_n} \mathrm{Var}_\infty(\hat{\xi}_n; \pi), \text{ s.t.}$$

$$\mathrm{E}[\mathrm{c}_0 + \pi_1(X_{Ct})\, c_1(X_{Ct}) + \pi_1(X_{Ct})\pi_2(X_{CtM})\, c_2(X_{CtM})] = \mathrm{b}_0.$$

Partial Observation Model $\mathcal{M}_\pi$

Example Dataset:

Δ	X_C	X_t	X_M	X_r
2	0.4	1.3	-0.7	·
∞	1.2	0.8	1.5	-0.3
1	-1.1	0.2	·	·
2	1.4	-0.9	0.6	·



- Mo
  (Δ
- Ful

Two mathematical challenges:

1. **Estimators** $\hat{\xi}_n$: Influence functions φ characterize estimators on $\mathcal{M}_\pi$ and yield $\mathrm{Var}(\hat{\xi}_n;\pi)$.
2. **Optimized design** π^*: Solve the variational problem $\min_\pi \mathrm{Var}_\infty(\hat{\xi}_n;\pi)$.

Problem:

$$\min_{\pi=(\pi_1,\pi_2),\ \mathrm{RAL}\ \hat{\xi}_n} \mathrm{Var}_\infty(\hat{\xi}_n;\pi), \text{ s.t.}$$

$$\mathrm{E}[c_0 + \pi_1(X_{Ct})\, c_1(X_{Ct}) + \pi_1(X_{Ct})\pi_2(X_{CtM})\, c_2(X_{CtM})] = b_0.$$

	Δ	$G_\Delta(X)$	Cost
π_1	1	X_{Ct}	c_0
π_2	2	X_{CtM}	$c_0 + c_1$
	∞	X_{CtMr}	$c_0 + c_1 + c_2$

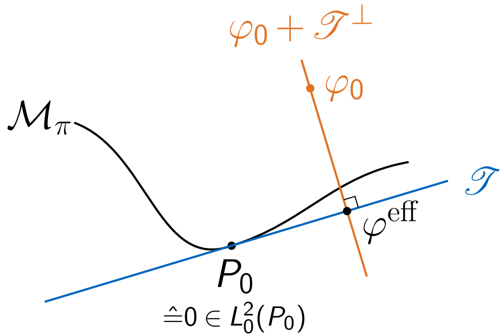
RAL Estimators and Influence Functions

$\hat{\xi}_n$ is regular and asymptotically linear (RAL) for ξ if

$$\sqrt{n}(\hat{\xi}_n - \xi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X^{(i)}) + o_{\mathcal{P}}(1)$$

for some influence function φ, so that

$$\sqrt{n}(\hat{\xi}_n - \xi) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}\left[\varphi\varphi^\top\right]\right).$$



Goal: $\underset{\varphi}{\operatorname{argmin}}\, \mathrm{E}\left[\varphi\varphi^\top\right]$

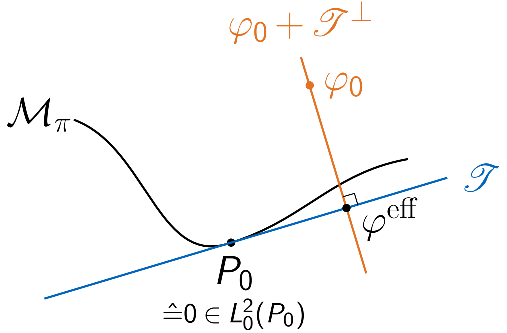
RAL Estimators and Influence Functions

$\hat{\xi}_n$ is regular and asymptotically linear (RAL) for ξ if

$$\sqrt{n}(\hat{\xi}_n - \xi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X^{(i)}) + o_{\mathcal{P}}(1)$$

for some influence function φ, so that

$$\sqrt{n}(\hat{\xi}_n - \xi) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}\left[\varphi\varphi^\top\right]\right).$$



Goal: $\operatorname*{argmin}_{\varphi} \mathrm{E}[\varphi\varphi^\top] = \varphi^{\text{eff}} = \Pi(\varphi_0 | \mathscr{T})$ yields minimal asymptotic variance $\hat{\xi}_n(\varphi^{\text{eff}})$.

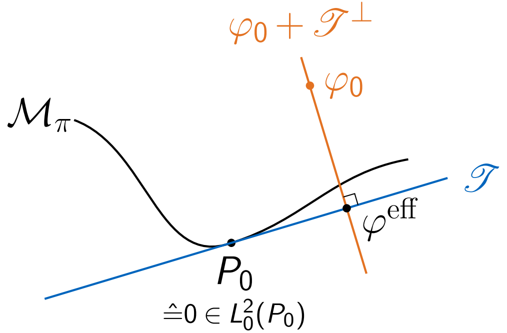
RAL Estimators and Influence Functions

$\hat{\xi}_n$ is regular and asymptotically linear (RAL) for ξ if

$$\sqrt{n}(\hat{\xi}_n - \xi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X^{(i)}) + o_{\mathcal{P}}(1)$$

for some influence function φ, so that

$$\sqrt{n}(\hat{\xi}_n - \xi) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}\left[\varphi\varphi^\top\right]\right).$$



Goal: $\underset{\varphi}{\operatorname{argmin}}\, \mathrm{E}[\varphi\varphi^\top] = \varphi^{\mathrm{eff}} = \Pi(\varphi_0 | \mathscr{T})$ yields minimal asymptotic variance $\hat{\xi}_n(\varphi^{\mathrm{eff}})$.

Insight [Tsiatis, 2006]: Find $\varphi_\xi^{\mathrm{eff}}$ the full-data model $\mathcal{M}_1$; augment to the partial-data models $\mathcal{M}_\pi$.

Estimator $\hat{\xi}_n$ in $\mathcal{M}_\pi$

Theorem 1. Efficient influence function for $\xi = \beta_{rM}\beta_{Mt}$ in $\mathcal{M}_1$:

$$\varphi_\xi^{\mathrm{eff}} = \beta_{rM}\, \varphi_{\beta_{Mt}}^{\mathrm{eff}} + \varphi_{\beta_{rM}}^{\mathrm{eff}}\, \beta_{Mt},$$

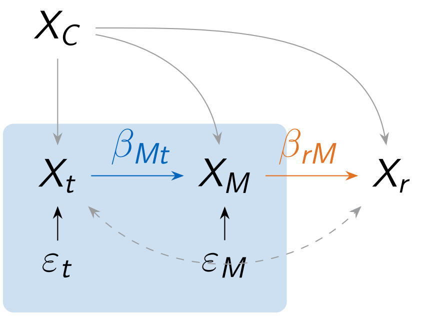
$$\varphi_{\beta_{Mt}}^{\mathrm{eff}} = \varepsilon_M\, \varepsilon_t\, \mathrm{Var}(\varepsilon_t)^{-1}, \qquad \varphi_{\beta_{rM}}^{\mathrm{eff}} = \varepsilon_r^{\perp}\, \varepsilon_M^{\top}\, \mathrm{Var}(\varepsilon_M)^{-1}.$$

Estimator $\hat{\xi}_n$ in $\mathcal{M}_\pi$

Theorem 1. Efficient influence function for $\xi = \beta_{rM}\beta_{Mt}$ in $\mathcal{M}_1$:

$$\varphi_\xi^{\mathrm{eff}} = \beta_{rM}\,\varphi_{\beta_{Mt}}^{\mathrm{eff}} + \varphi_{\beta_{rM}}^{\mathrm{eff}}\,\beta_{Mt},$$

$$\varphi_{\beta_{Mt}}^{\mathrm{eff}} = \varepsilon_M\,\varepsilon_t\,\mathrm{Var}(\varepsilon_t)^{-1}, \qquad \varphi_{\beta_{rM}}^{\mathrm{eff}} = \varepsilon_r^{\perp}\,\varepsilon_M^{\top}\,\mathrm{Var}(\varepsilon_M)^{-1}.$$



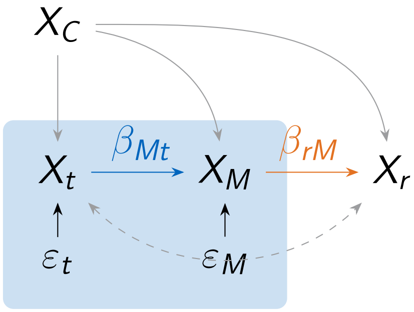
Estimator: $\hat{\beta}_{Mt,n}$ s.t. $\mathrm{E}[\hat{\varepsilon}_t\hat{\varepsilon}_M] \overset{!}{=} 0$

Estimator $\hat{\xi}_n$ in $\mathcal{M}_\pi$

Theorem 1. Efficient influence function for $\xi = \beta_{rM}\beta_{Mt}$ in $\mathcal{M}_1$:

$$\varphi_\xi^{\text{eff}} = \beta_{rM}\,\varphi_{\beta_{Mt}}^{\text{eff}} + \varphi_{\beta_{rM}}^{\text{eff}}\,\beta_{Mt},$$

$$\varphi_{\beta_{Mt}}^{\text{eff}} = \varepsilon_M\,\varepsilon_t\,\mathrm{Var}(\varepsilon_t)^{-1}, \qquad \varphi_{\beta_{rM}}^{\text{eff}} = \varepsilon_r^{\perp}\,\varepsilon_M^{\top}\,\mathrm{Var}(\varepsilon_M)^{-1}.$$



Estimator: $\hat{\beta}_{Mt,n}$ s.t. $\mathrm{E}[\hat{\varepsilon}_t\hat{\varepsilon}_M] \stackrel{!}{=} 0$

Optimal augmentation of φ_ξ^{eff} in $\mathcal{M}_\pi$ [Tsiatis, 2006]:

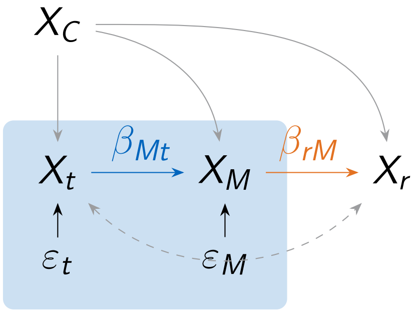
$$\varphi_\xi^{\text{opt},\pi} = \frac{I(\Delta \geq 2)}{\pi_1(X_{Ct})}\,\beta_{rM}\,\varphi_{\beta_{Mt}}^{\text{eff}} + \frac{I(\Delta = \infty)}{\pi_1(X_{Ct})\pi_2(X_{CtM})}\,\varphi_{\beta_{rM}}^{\text{eff}}\,\beta_{Mt}.$$

Estimator $\hat{\xi}_n$ in $\mathcal{M}_\pi$

Theorem 1. Efficient influence function for $\xi = \beta_{rM}\beta_{Mt}$ in $\mathcal{M}_1$:

$$\varphi_\xi^{\text{eff}} = \beta_{rM}\,\varphi_{\beta_{Mt}}^{\text{eff}} + \varphi_{\beta_{rM}}^{\text{eff}}\,\beta_{Mt},$$

$$\varphi_{\beta_{Mt}}^{\text{eff}} = \varepsilon_M\,\varepsilon_t\,\mathrm{Var}(\varepsilon_t)^{-1}, \qquad \varphi_{\beta_{rM}}^{\text{eff}} = \varepsilon_r^{\perp}\,\varepsilon_M^{\top}\,\mathrm{Var}(\varepsilon_M)^{-1}.$$



Estimator: $\hat{\beta}_{Mt,n}$ s.t. $\mathrm{E}[\hat{\varepsilon}_t\hat{\varepsilon}_M] \overset{!}{=} 0$

Optimal augmentation of φ_ξ^{eff} in $\mathcal{M}_\pi$ [Tsiatis, 2006]:

$$\varphi_\xi^{\text{opt},\pi} = \frac{I(\Delta \geq 2)}{\pi_1(X_{Ct})}\,\beta_{rM}\,\varphi_{\beta_{Mt}}^{\text{eff}} + \frac{I(\Delta = \infty)}{\pi_1(X_{Ct})\pi_2(X_{CtM})}\,\varphi_{\beta_{rM}}^{\text{eff}}\,\beta_{Mt}.$$

Asymptotic Variance:

$$\mathrm{Var}_\infty(\hat{\xi}_n;\pi) = \mathrm{E}\left[\frac{g_1(X_{Ct})}{\pi_1(X_{Ct})}\right] + \mathrm{E}\left[\frac{g_2(X_{CtM})}{\pi_1(X_{Ct})\pi_2(X_{CtM})}\right],$$

$$\begin{pmatrix} g_1(X_{Ct}) := \beta_{rM}\mathrm{Var}(\varepsilon_M)\,\beta_{rM}^{\top}\,\frac{\varepsilon_t^2}{\mathrm{Var}(\varepsilon_t)^2} \\ g_2(X_{CtM}) := \mathrm{Var}(\varepsilon_r \mid \varepsilon_t)\,\left(\varepsilon_M^{\top}\mathrm{Var}(\varepsilon_M)^{-1}\,\beta_{Mt}\right)^2 \end{pmatrix}$$

Optimized Design π^*

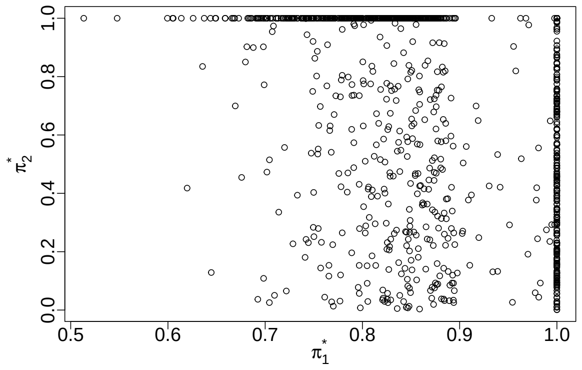
Theorem 2. $\hat{\xi}_n^{\text{opt}}$ in $\mathcal{M}_{\pi^*}$ minimizes estimation variance[1] of ξ under budget constraint ($\lambda > 0$ multiplier):

$$\pi_1^* = \min\left(1, \max\left(\sqrt{\frac{g_1}{\lambda c_1}}, \sqrt{\frac{g_1 + \mathrm{E}\left[g_2 \mid \varepsilon_{C,t}\right]}{\lambda\left(c_1 + \mathrm{E}\left[c_2 \mid \varepsilon_{C,t}\right]\right)}}\right)\right), \qquad \pi_2^* = \min\left(1, \sqrt{\frac{g_2\, c_1}{g_1\, c_2}}\right).$$

Intuition: Measure further proportionally to the information-to-cost ratio $\sqrt{g_j/c_j}$:

$$g_1(X_{Ct}) = \underbrace{\frac{\beta_{rM} \mathrm{Var}(\varepsilon_M)\, \beta_{rM}^\top}{\mathrm{Var}(\varepsilon_t)^2}}_{\text{constant}} \cdot \underbrace{\varepsilon_t^2}_{\text{leverage for } \hat{\beta}_{Mt}} .$$

Optimal Propensity π^* Distribution



[1] Over augmentations of $\varphi^{F,\text{eff}}$; see arxiv.org/pdf/2603.22024 for details.

Applications and Code

Application	n	Rel. efficiency
Sachs	853	76.6%
Sachs	853	68.0%
MIMIC-III (ICU)	999	90.0%
causalAssembly	3,000	94.7%

Rel. efficiency $= \mathrm{Var}_\infty(\hat{\xi}_n; \pi^*)/\mathrm{Var}_\infty(\hat{\xi}_n; 1)$

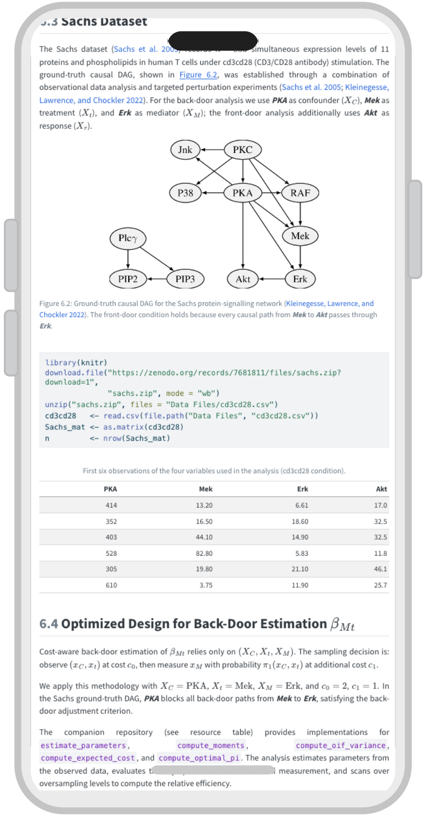
- Optimized design π^* **dominates** full measurements ($\pi \equiv 1$)
- Gains persist under mild departures from linearity

Applications and Code

Application	n	Rel. efficiency
Sachs	853	76.6%
Sachs	853	68.0%
MIMIC-III (ICU)	999	90.0%
causalAssembly	3,000	94.7%

Rel. efficiency $= \mathrm{Var}_\infty(\hat{\xi}_n; \pi^*)/\mathrm{Var}_\infty(\hat{\xi}_n; 1)$

- Optimized design π^* **dominates** full measurements ($\pi \equiv 1$)
- Gains persist under mild departures from linearity



6.3 Sachs Dataset

The Sachs dataset (Sachs et al. 200[redacted] simultaneous expression levels of 11 proteins and phospholipids in human T cells under cd3cd28 (CD3/CD28 antibody) stimulation. The ground-truth causal DAG, shown in Figure 6.2, was established through a combination of observational data analysis and targeted perturbation experiments (Sachs et al. 2005; Kleinegesse, Lawrence, and Chockler 2022). For the back-door analysis we use ***PKA*** as confounder (X_C), ***Mek*** as treatment (X_t), and ***Erk*** as mediator (X_M); the front-door analysis additionally uses ***Akt*** as response (X_r).

Figure 6.2: Ground-truth causal DAG for the Sachs protein-signalling network (Kleinegesse, Lawrence, and Chockler 2022). The front-door condition holds because every causal path from ***Mek*** to ***Akt*** passes through ***Erk***.

```
library(knitr)
download.file("https://zenodo.org/records/7681811/files/sachs.zip?download=1",
              "sachs.zip", mode = "wb")
unzip("sachs.zip", files = "Data Files/cd3cd28.csv")
cd3cd28   <- read.csv(file.path("Data Files", "cd3cd28.csv"))
Sachs_mat <- as.matrix(cd3cd28)
n         <- nrow(Sachs_mat)
```

First six observations of the four variables used in the analysis (cd3cd28 condition).

PKA	Mek	Erk	Akt
414	13.20	6.61	17.0
352	16.50	18.60	32.5
403	44.10	14.90	32.5
528	82.80	5.83	11.8
305	19.80	21.10	46.1
610	3.75	11.90	25.7

6.4 Optimized Design for Back-Door Estimation β_{Mt}

Cost-aware back-door estimation of β_{Mt} relies only on (X_C, X_t, X_M). The sampling decision is: observe (x_C, x_t) at cost c_0, then measure x_M with probability $\pi_1(x_C, x_t)$ at additional cost c_1.

We apply this methodology with $X_C = \mathrm{PKA}$, $X_t = \mathrm{Mek}$, $X_M = \mathrm{Erk}$, and $c_0 = 2$, $c_1 = 1$. In the Sachs ground-truth DAG, ***PKA*** blocks all back-door paths from ***Mek*** to ***Erk***, satisfying the back-door adjustment criterion.

The companion repository (see resource table) provides implementations for `estimate_parameters`, `compute_moments`, `compute_oif_variance`, `compute_expected_cost`, and `compute_optimal_pi`. The analysis estimates parameters from the observed data, evaluates t[illegible] measurement, and scans over oversampling levels to compute the relative efficiency.

https://st-mardi.quarto.pub/gmci/

Conclusion

- Efficient influence function for the front-door effect $\xi = \beta_{rM}\beta_{Mt}$ in $\mathcal{M}_1$, and its optimal augmentation.
- Closed-form sequential design π^* based on unit-level information-to-cost ratios g_j/c_j.
- Optimized **back-door** experimental design follows as a corollary.
- 5–32% variance reduction over full measurement in biological, medical, and industrial datasets.

Outlook: Efficient estimation in general ADMGs. Sampling designs that depart from the causal ordering.

A. Tsiatis.
Semiparametric Theory and Missing Data.
Springer Series in Statistics. Springer, New York, 2006.

L. Mareis and M. Drton.
Cost-Aware Optimized Front-Door Experimental Design.
To appear: *Proceedings of Machine Learning Research*, PMLR, 2026. Preprint: arXiv:2603.22024.

Personal Website

LinkedIn