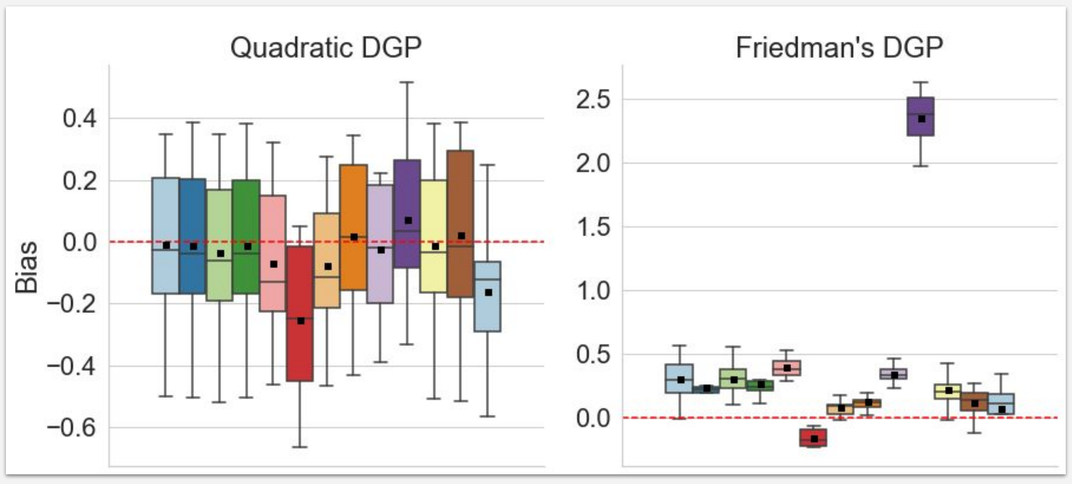
Validating Causal Inference Methods

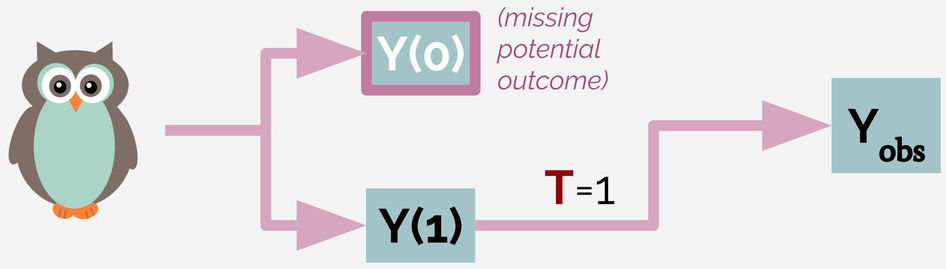
Harsh Parikh[*], Carlos Varjao[^], Louise Xu[^], Eric Tchetgen Tchetgen[#]
[*]Duke University, [^]Amazon.com, [#]University of Pennsylvania

No 'One-Size-Fits-All'

Evaluating performance of methods using True DGP

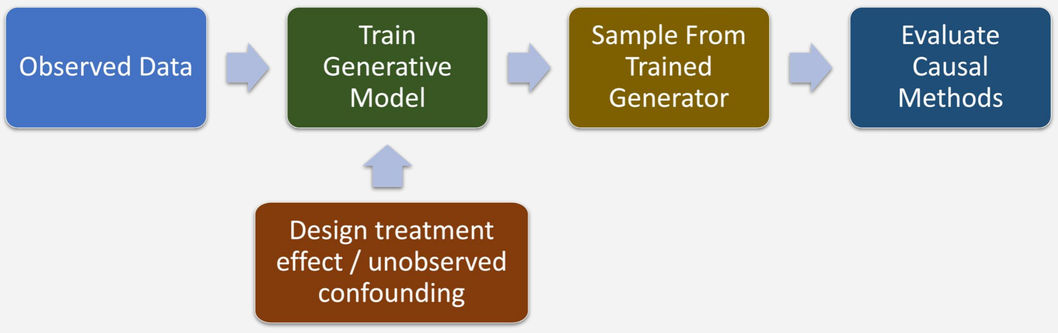


Missing Ground Truth

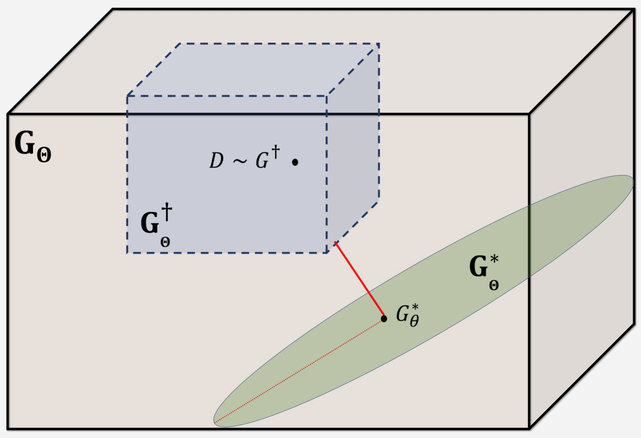


Credence Framework

Pipeline



Schema Space of Generative Models



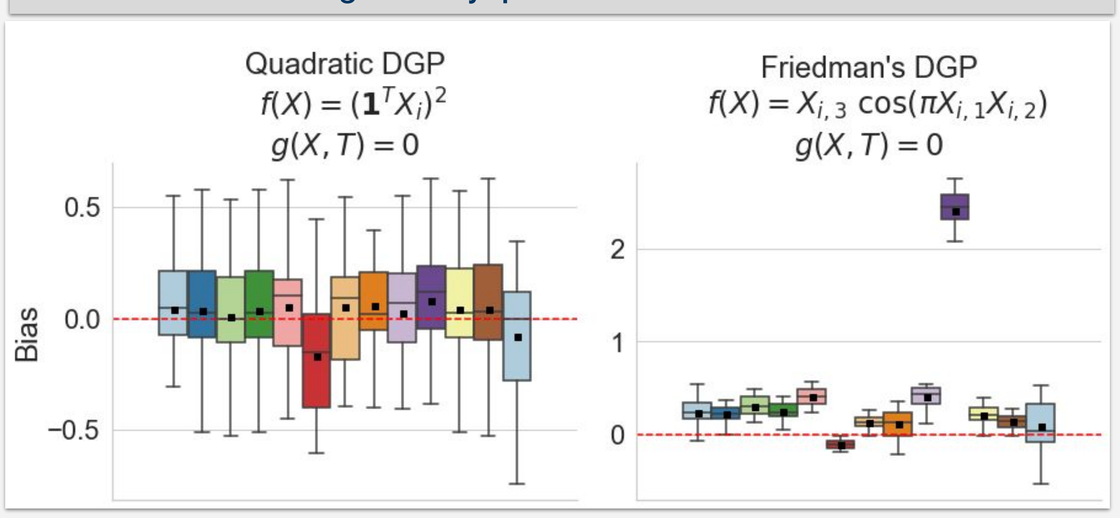
Credence's Objective Function

$$\min_\theta \begin{pmatrix} \mathbf{E}\left[d((X,Y,Z),(X',Y',Z'))\right] \\ +\alpha \left\|\mathbf{E}[Y'(1)-Y'(0)|X'=x'] - f(x')\right\| \\ +\beta \left\|\mathbf{E}[Y'(z')|X'=x',Z'=z'] - \mathbf{E}[Y'(z')|X'=x',Z'=1-z'] - g(x',z')\right\| \end{pmatrix}$$

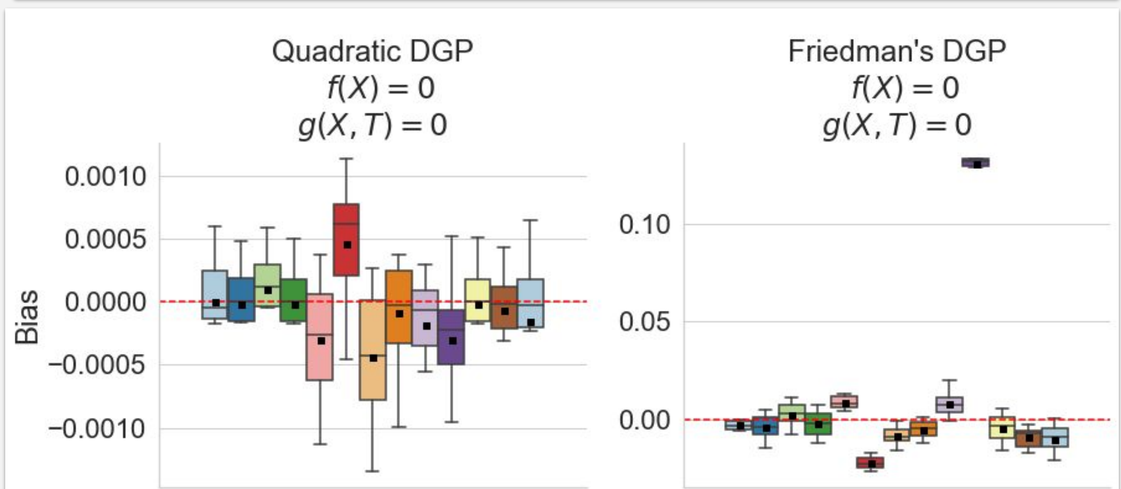
Validating using Credence

Evaluating performance of methods using learned DGP

Credence trained using correctly specified treatment effect and selection bias

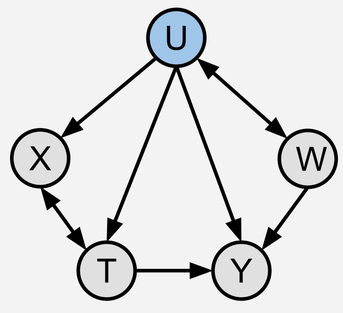


Credence trained assuming *null* treatment effect and selection bias

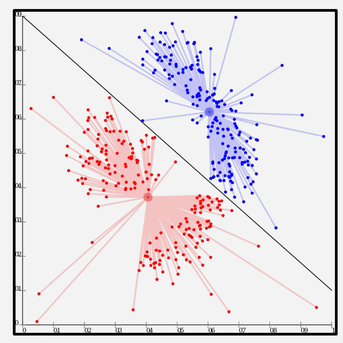


Existing Methods of Validation

Face-Validity Test



Placebo/Negative Control Tests



Handcrafted Synthetic Data Tests

Conclusion and Discussion

Limitation

- Generative models are sensitive to hyper-parameters
- Evaluations as good as the *assumptions* user makes

Future Directions

- Use Credence as a deep-bootstrap for *inference*
- Extension to scenarios with interference/homophily
- Theoretical guarantees on Credence based ranking

Duke UNIVERSITY amazon Wharton University of Pennsylvania