DeepIV: A Flexible Approach for Counterfactual Prediction

Jason Hartford and Kevin Leyton-Brown

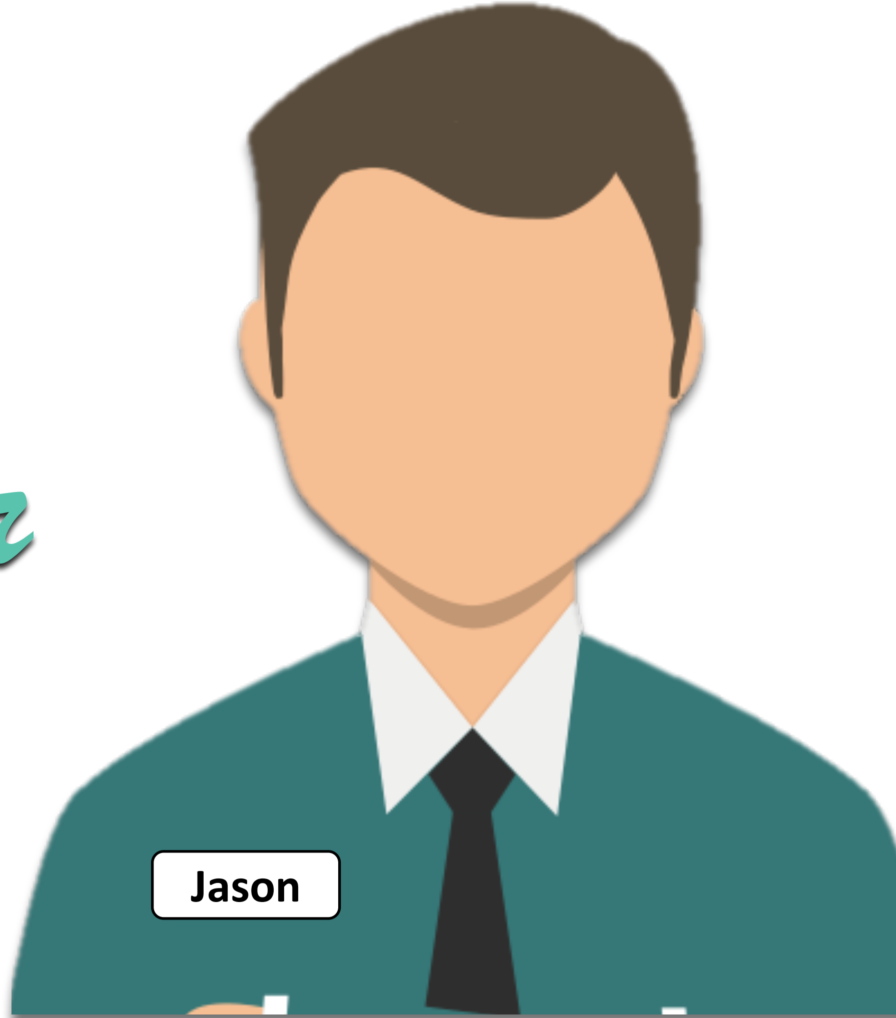
University of British Columbia

Greg Lewis* and Matt Taddy†

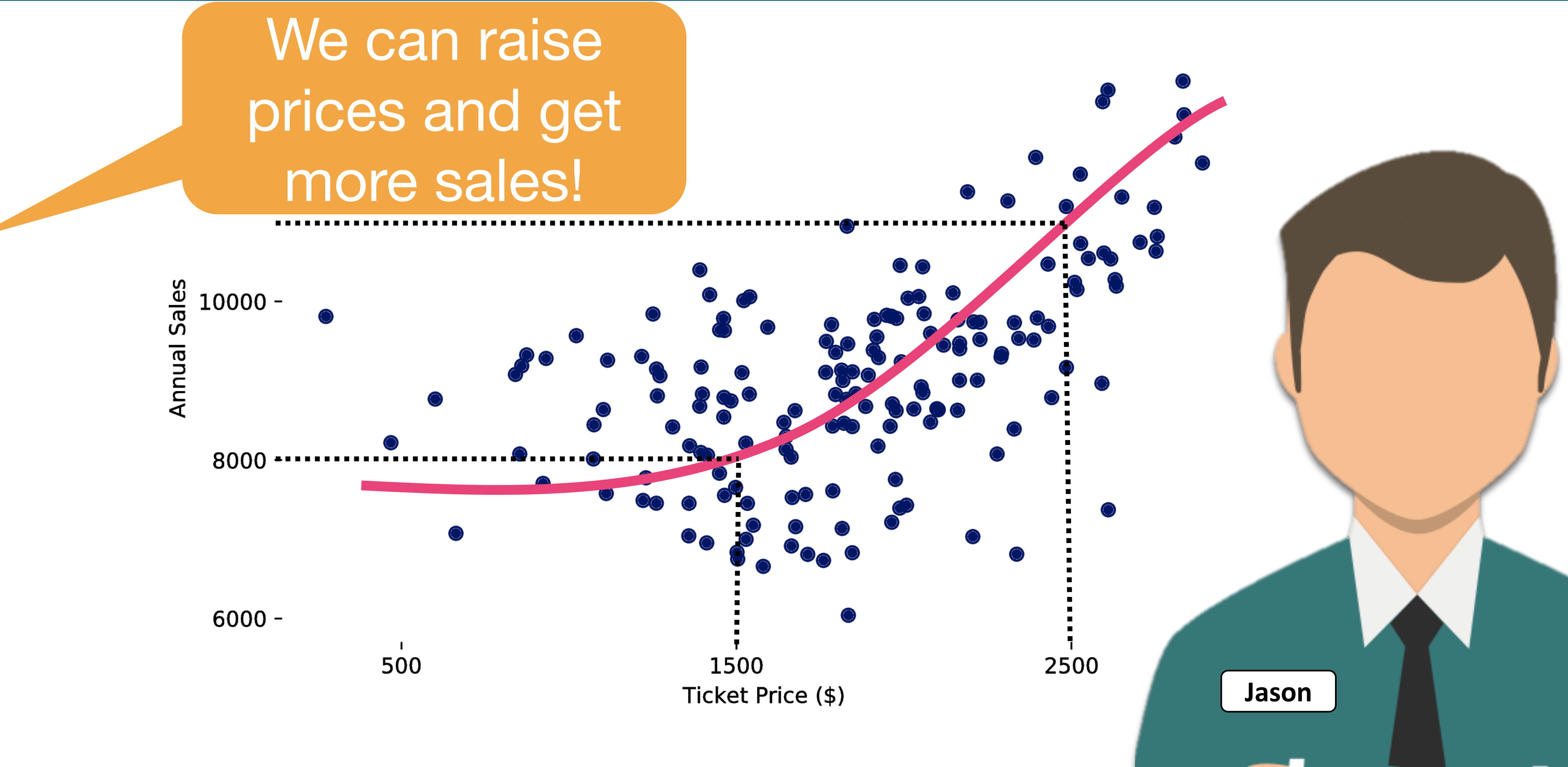
Microsoft Research &
*NBER / † University of Chicago

I need a model that predicts the effect of price on ticket sales

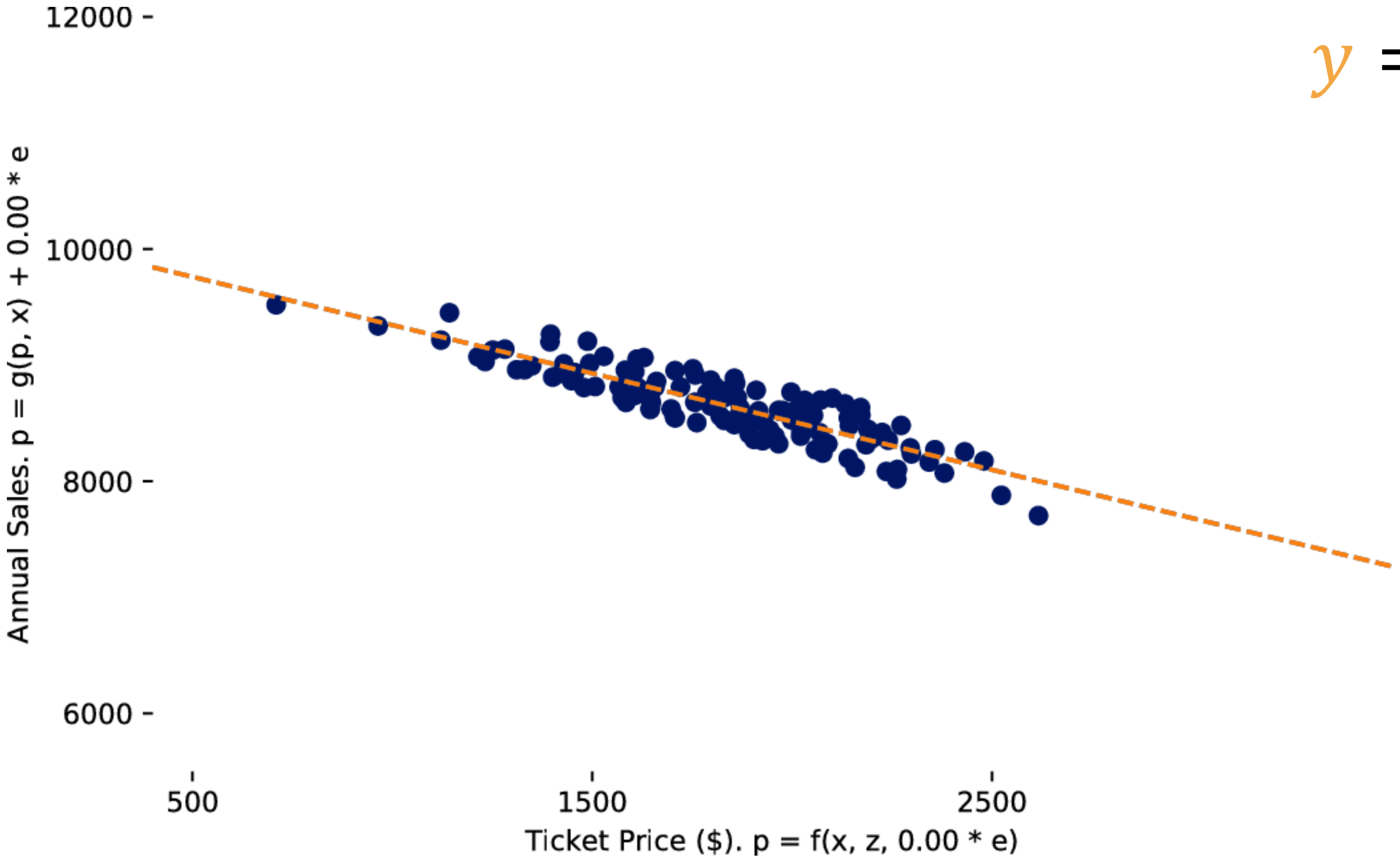
SkyHighAir



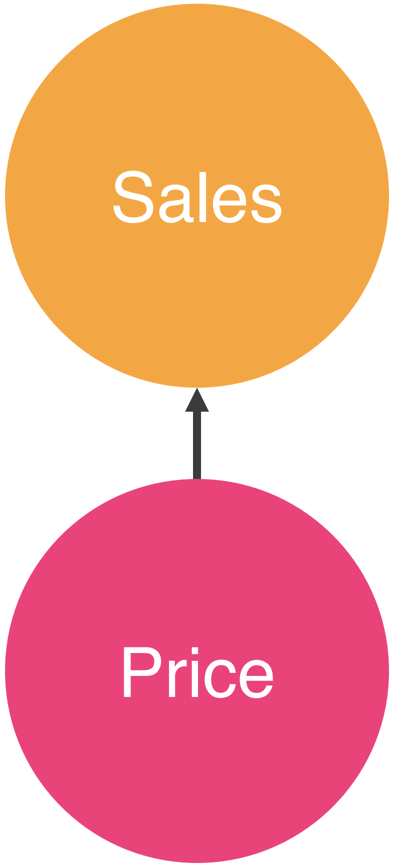
Prediction with confounding effects



Prediction with confounding effects

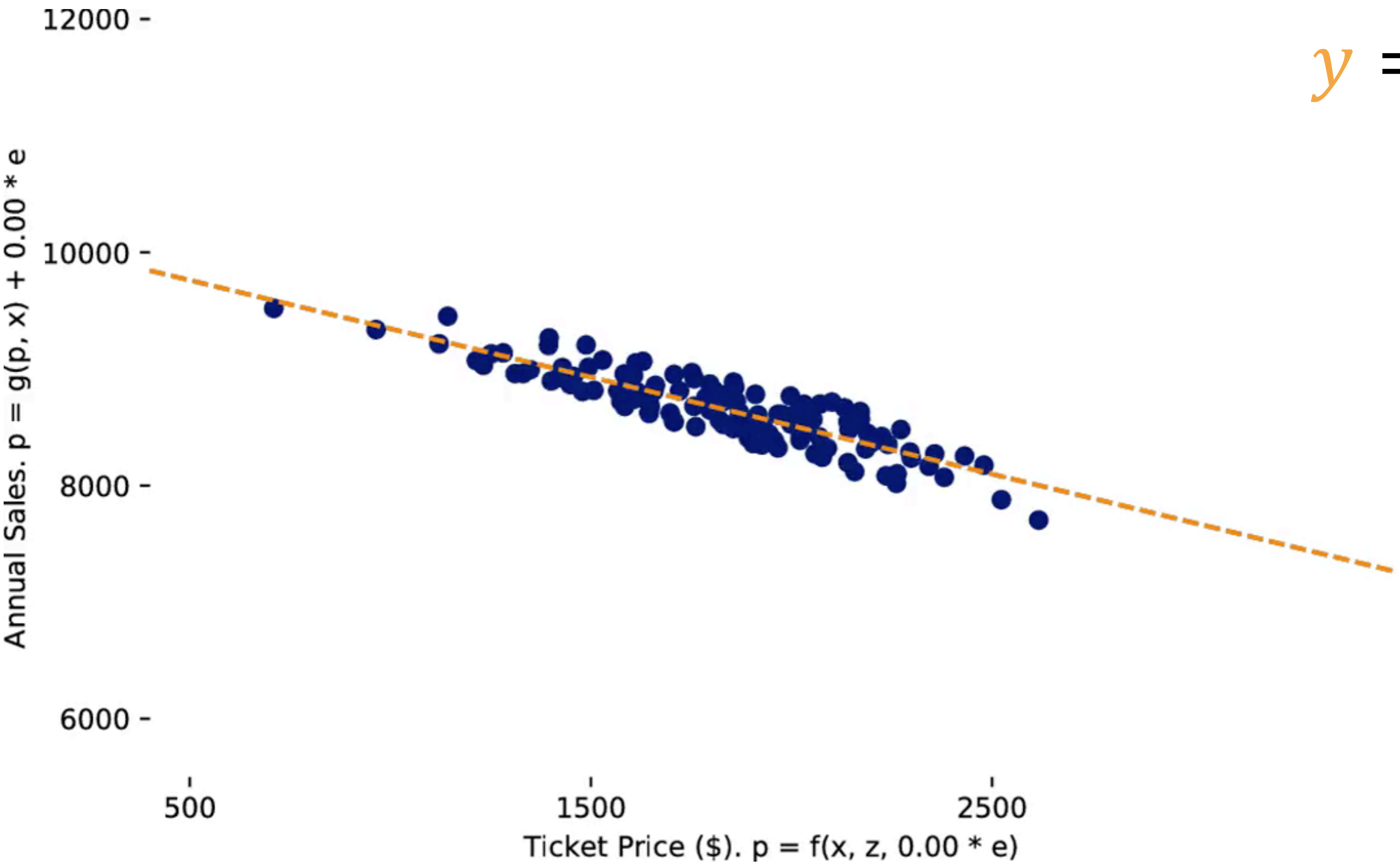


$$y = g(p)$$

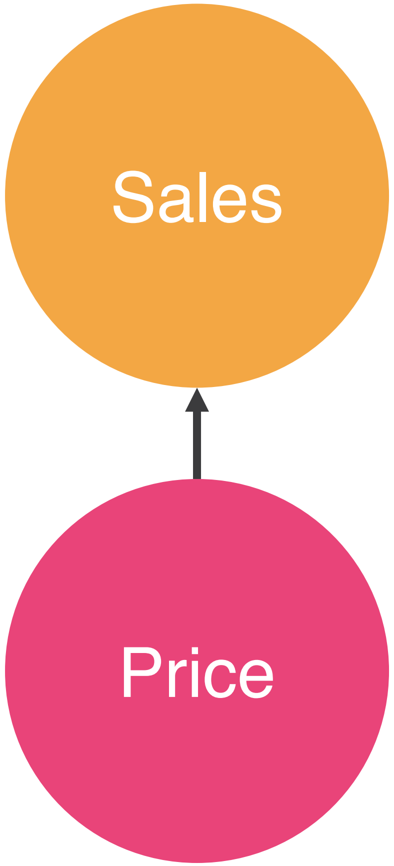


p

Prediction with confounding effects

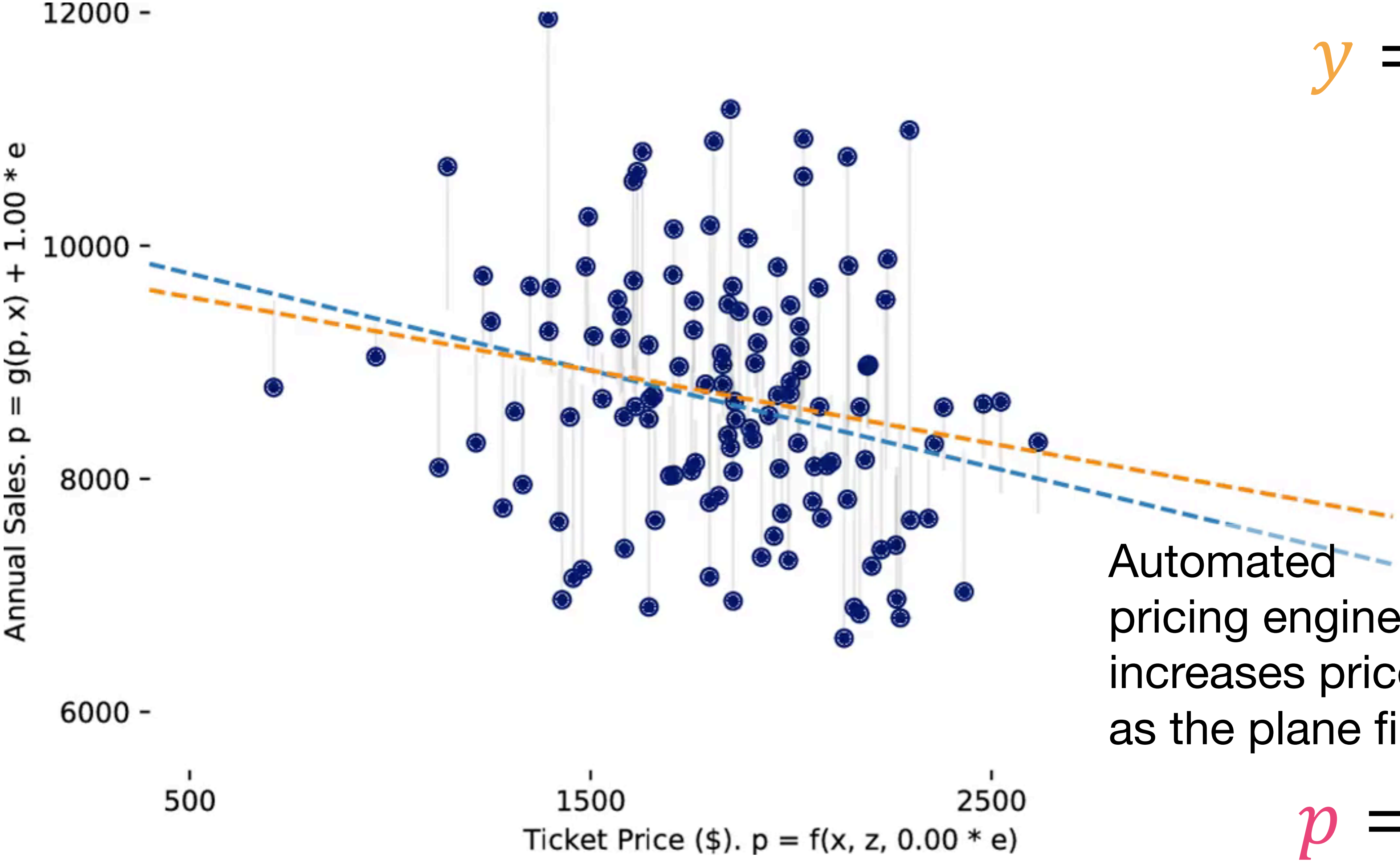


$$y = g(p, e)$$

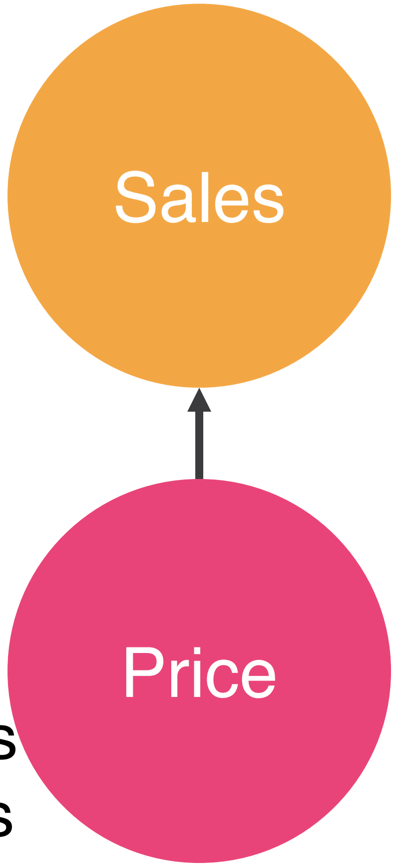


p

Prediction with confounding effects



$$y = g(p, e)$$



Automated pricing engine increases prices as the plane fills

$$p = f(e)$$

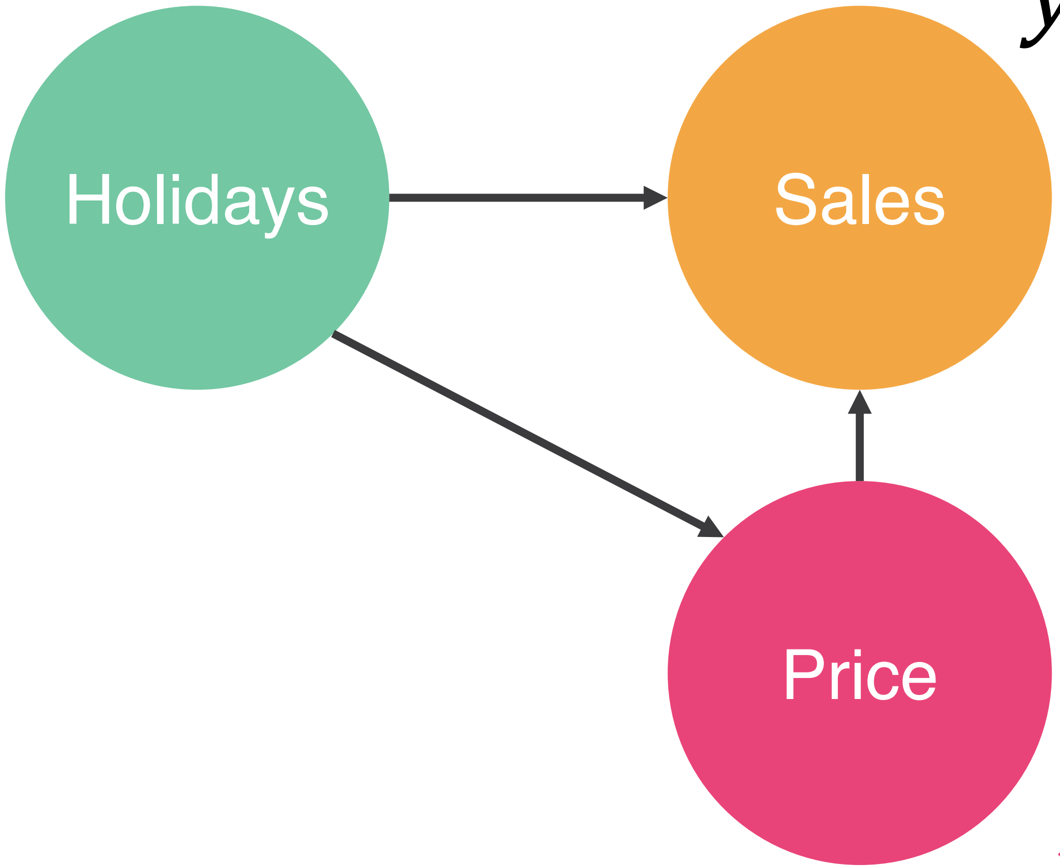
The observational distribution

"features / observed confounders"

x

"response"

$y = g(p, x)$



$p = f(x)$

"policy / treatment"

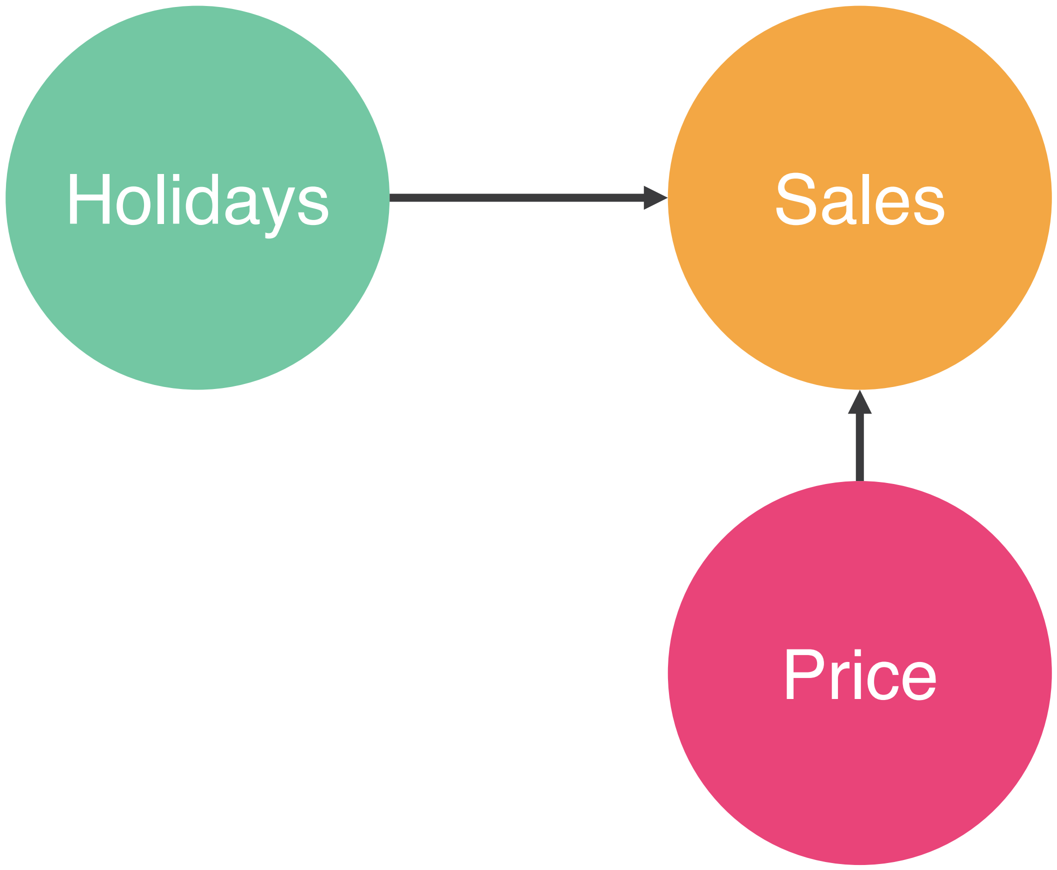
The interventional distribution

"features / observed confounders"

x

"response"

$$P(y|\mathrm{do}(\hat{p}), x)$$



Set $p = \hat{p}$

"policy / treatment"

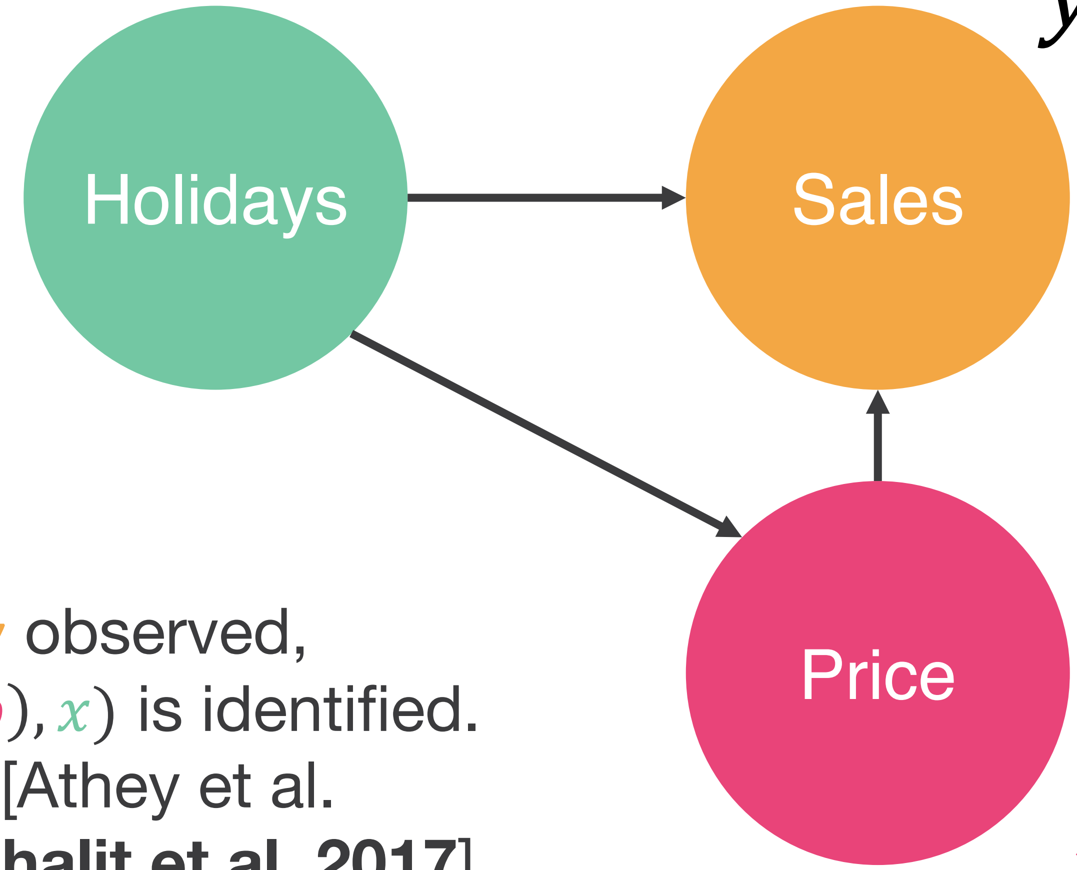
Identification of causal effects

"features / observed confounders"

x

"response"

$y = g(p, x)$

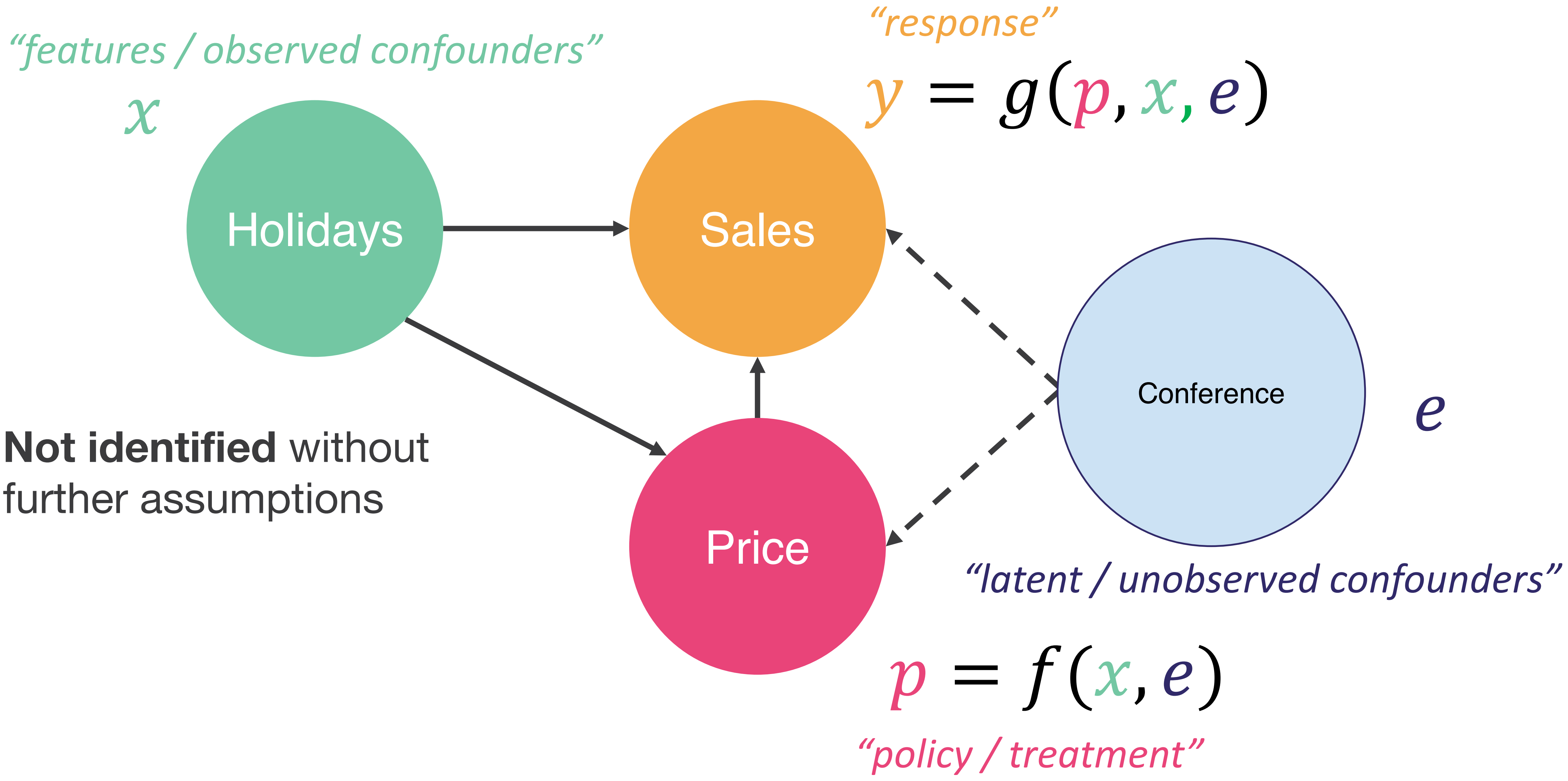


If x, p & y observed, $P(y|\mathrm{do}(p), x)$ is identified. See e.g. [Athey et al. 2016], [**Shalit et al. 2017**]

$p = f(x)$

"policy / treatment"

Identification of causal effects



Not identified without further assumptions

Identification of causal effects

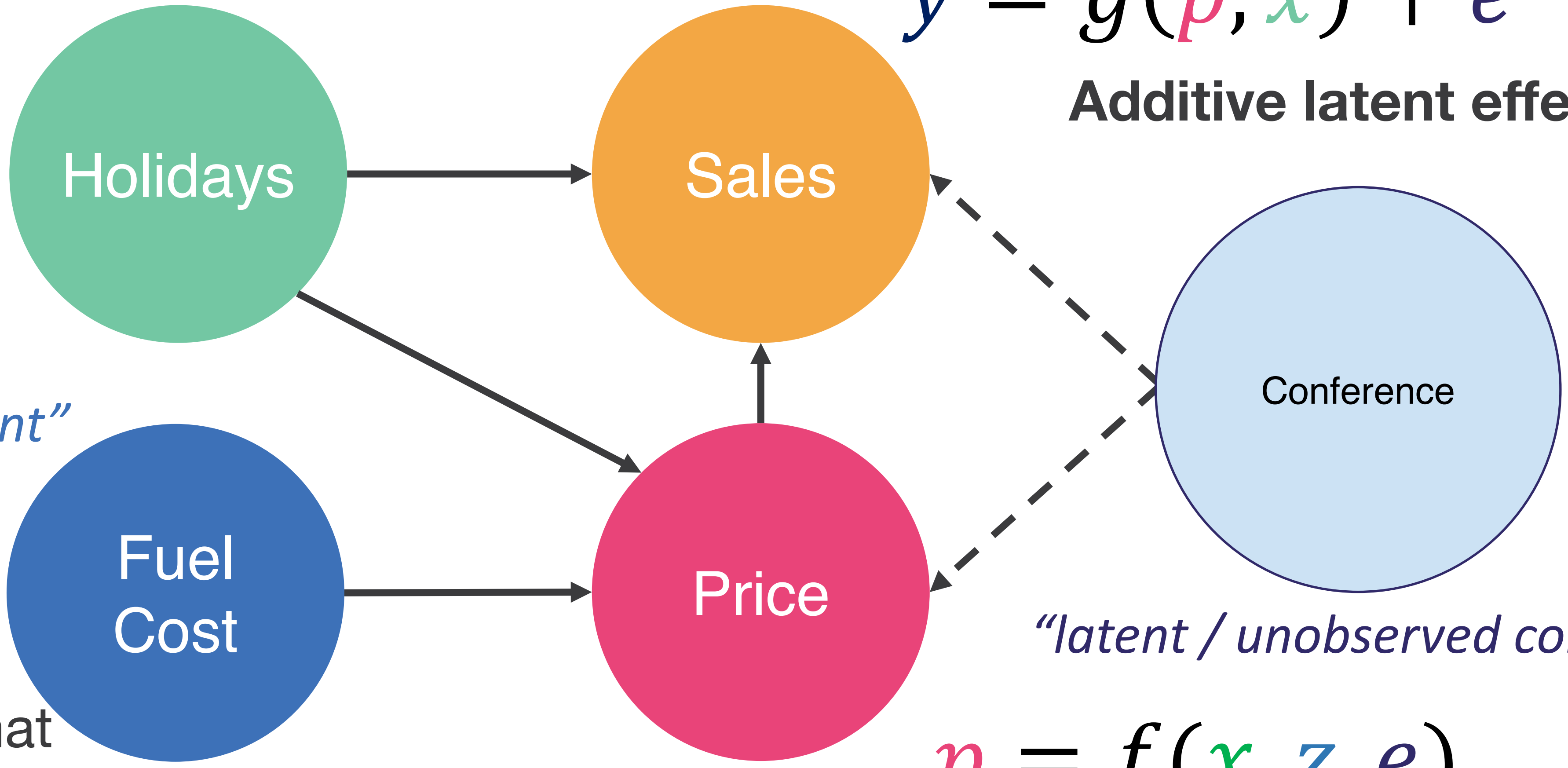
"response"

$$y = g(p, x) + e$$

Additive latent effects

"features / observed confounders"

x



e

"instrument"

z

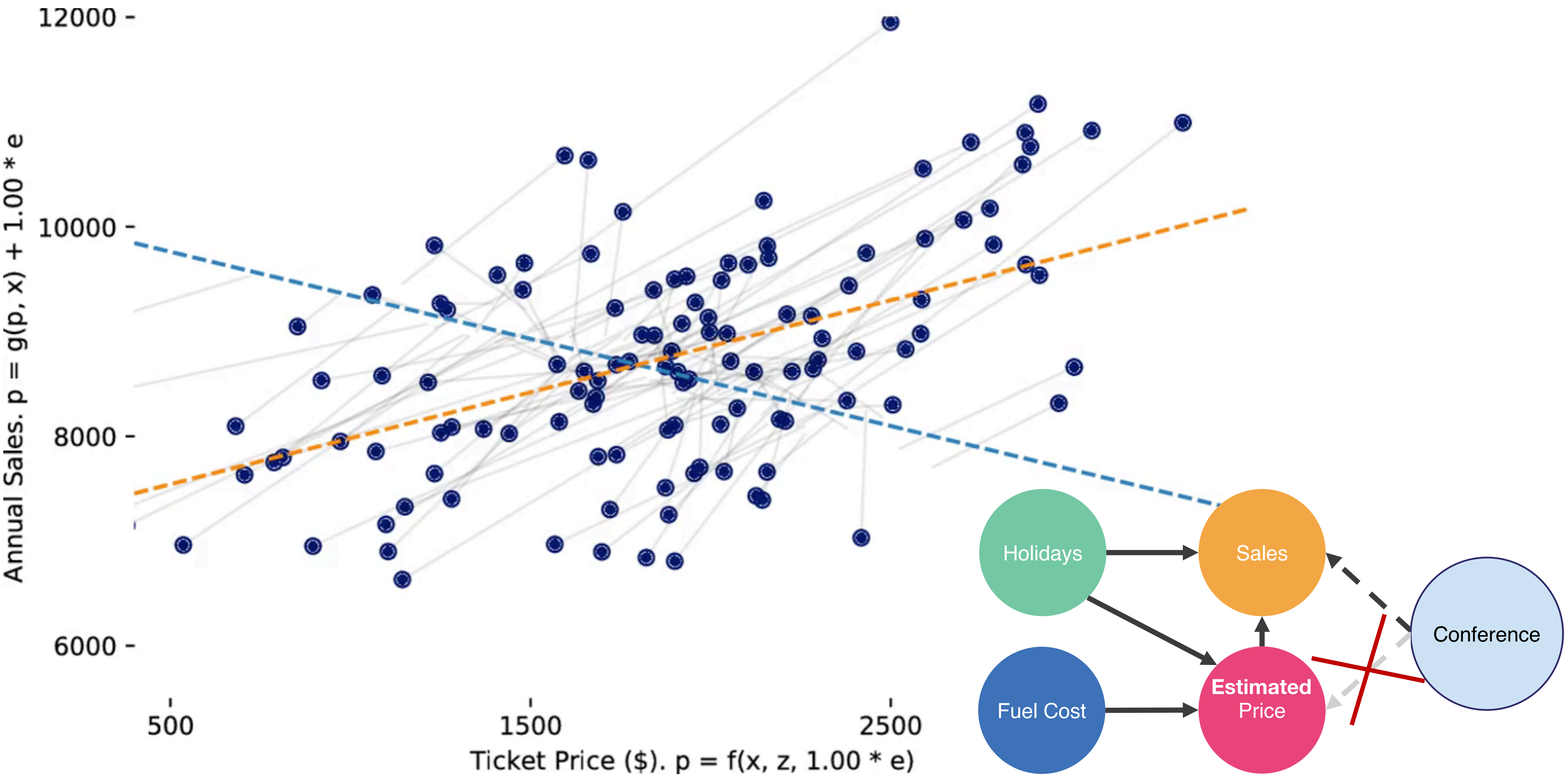
"latent / unobserved confounders"

Variable that only affects the response **indirectly** via its effect on price

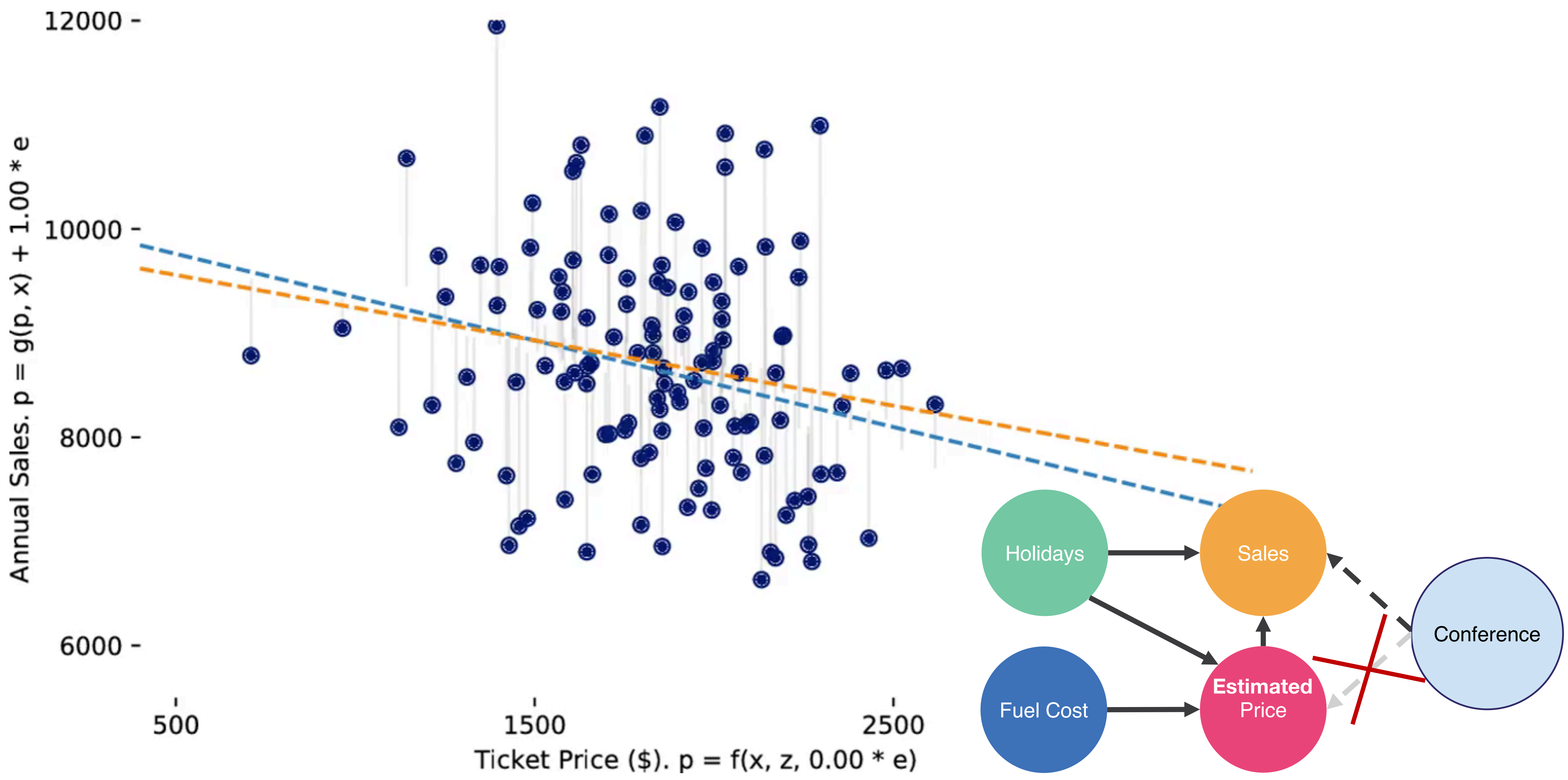
$$p = f(x, z, e)$$

"policy / treatment"

Simulate a world without latent effects on price



Simulate a world without latent effects on price



The learning problem

These assumptions imply the following **identity**[1],

$$\boxed{E[y|x,z]} = E[g(p,x)|x,z] = \int \boxed{g(p,x)}\,\boxed{dF(p|x,z)}$$

So we can recover $g(p,x)$ solve the implied **inversion problem**...

$$\min_{g\in G} \sum_{t=1}^{n} \left(y_t - \int g(p,x_t)dF(p|x,z)\right)^2$$

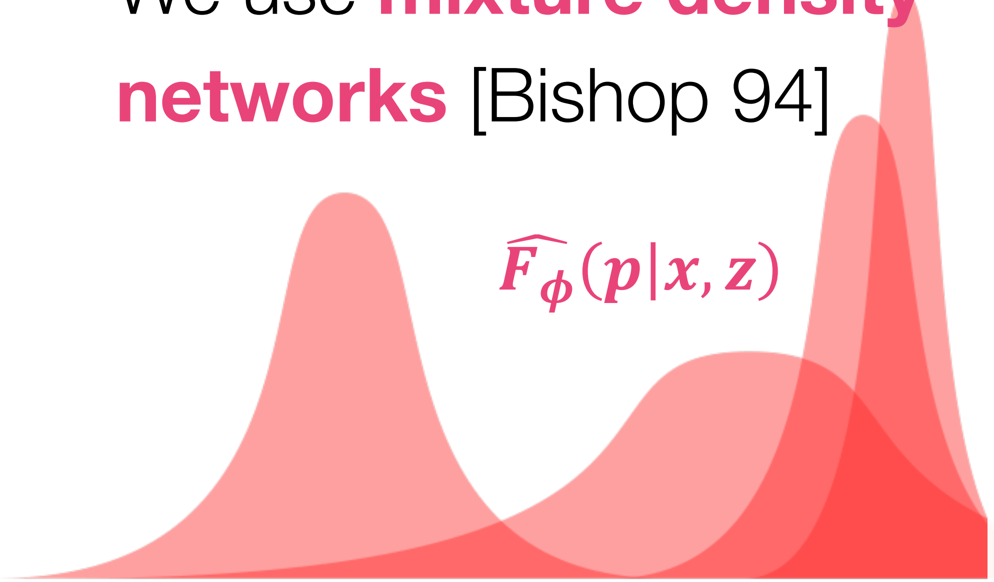
1. This holds if $E[e|x] = 0$. In general we recover $g(p,x)$ up to a constant wrt p – see paper for details.

A two-stage solution

$$\min_{g\in G} \sum_{t=1}^{n} \left(y_t - \int g(p, x_t)\, dF(p|x,z) \right)^2$$

Stage 1: fit $\widehat{F_\phi}(p|x,z)$ using the model of your choice.

We use **mixture density networks** [Bishop 94]

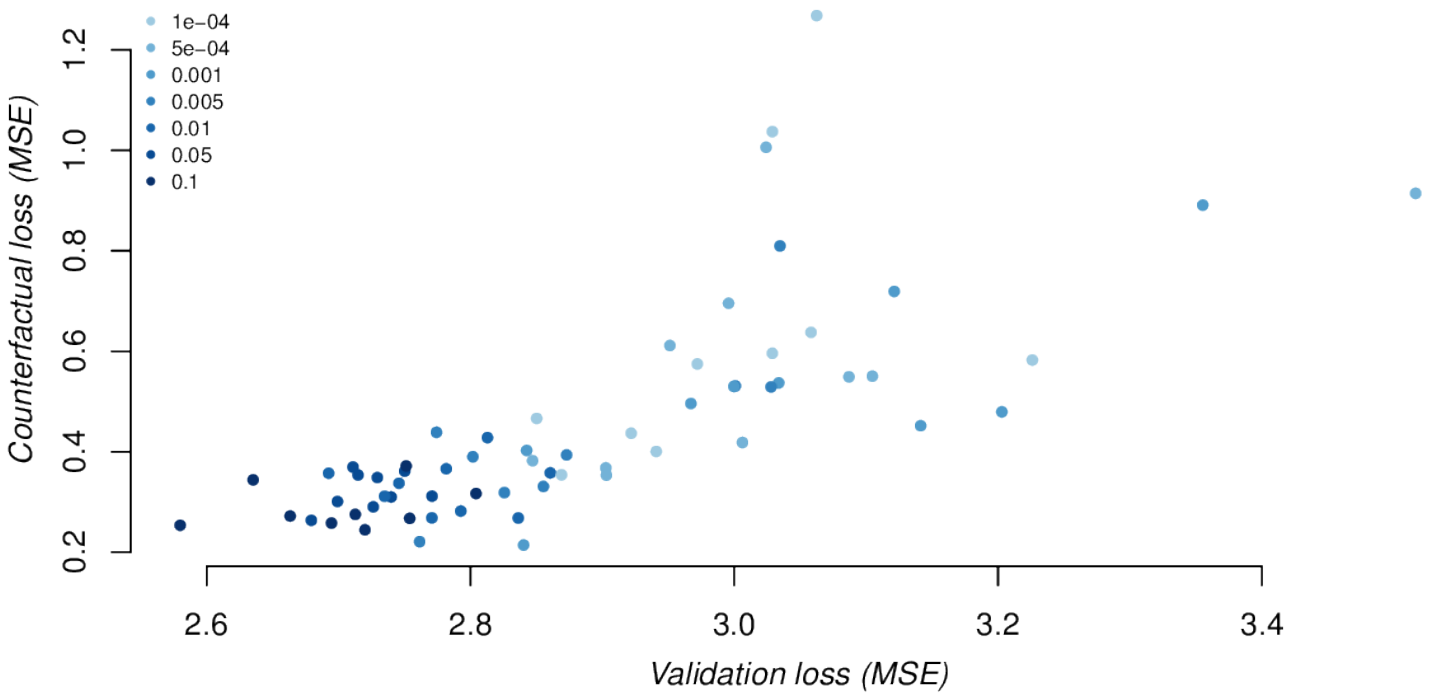


Stage 2: train network $\widehat{g_\theta}$ using **stochastic gradient descent** with **monte-carlo integration**.

$$\nabla L(\theta) = -2 \left(y_t - \frac{1}{|\dot{p}_1|} \sum_{\dot{p}_1 \sim \hat{F}(p|x,z)} \hat{g}(\dot{p}_1, x_t) \right) \times \left(\frac{1}{|\dot{p}_2|} \sum_{\dot{p}_2 \sim \hat{F}(p|x,z)} \nabla_\theta \hat{g}(\dot{p}_2, x_t) \right)$$

Causal Validation

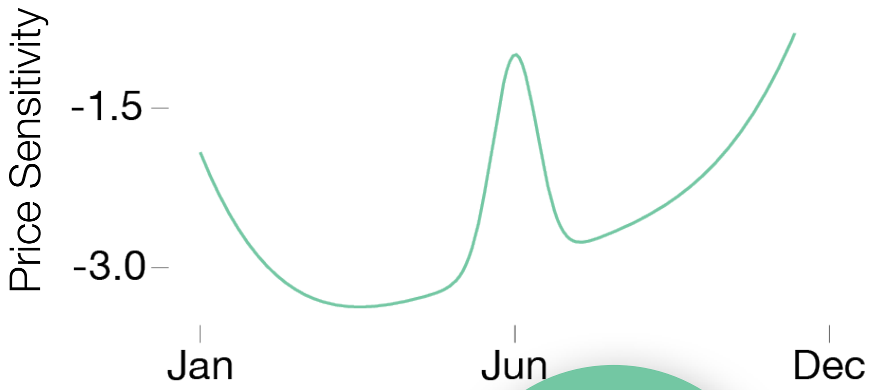
- In general, out-of-sample validation causal models is **challenging** / **impossible**…
- But… both our losses depend only on **observable** quantities ***and*** reflect causal loss, so we can simply use **standard validation sets**.

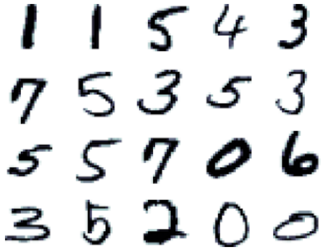


Evaluation

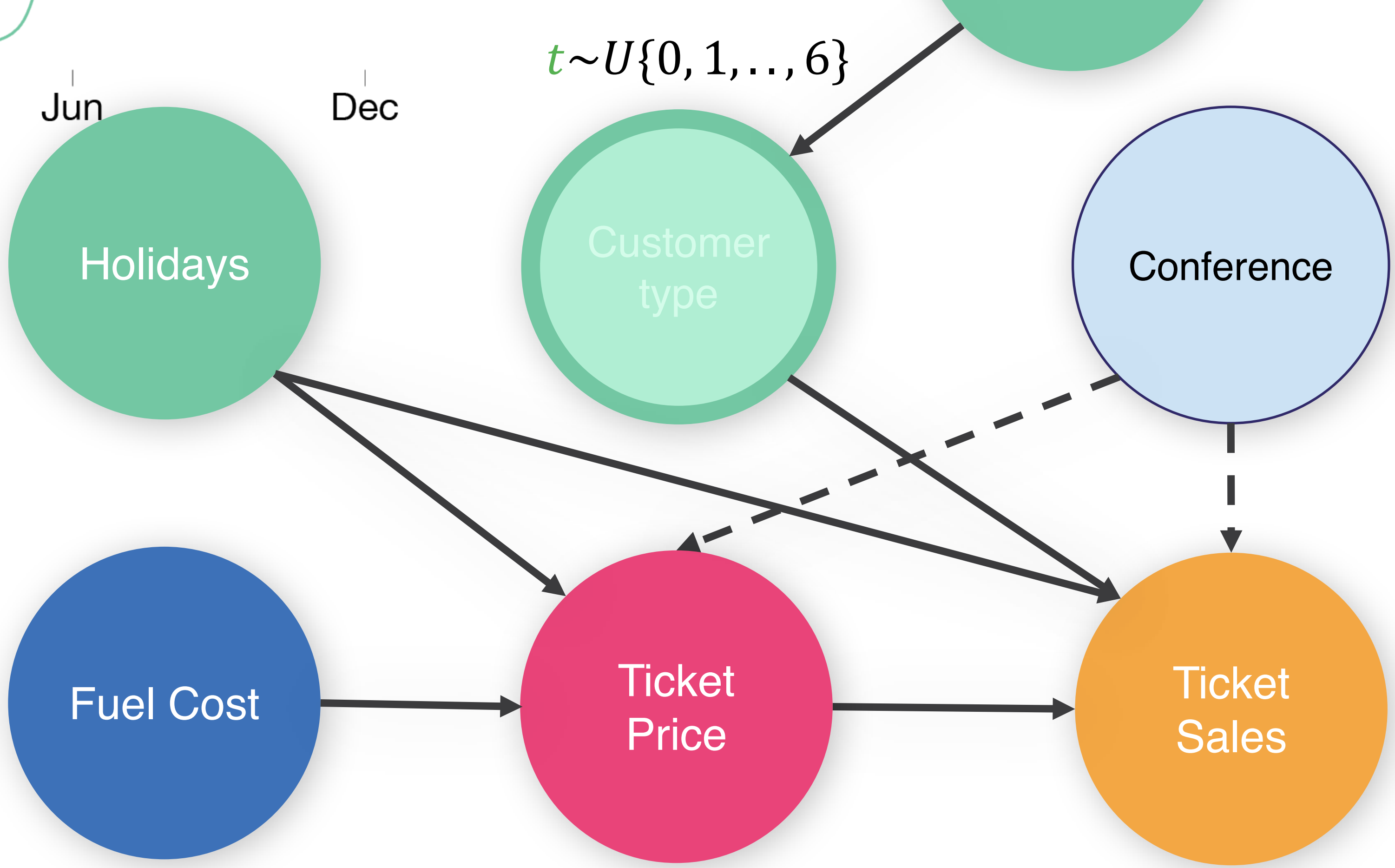
Simulation & Bing Ads Experiments

Simulation Experiments



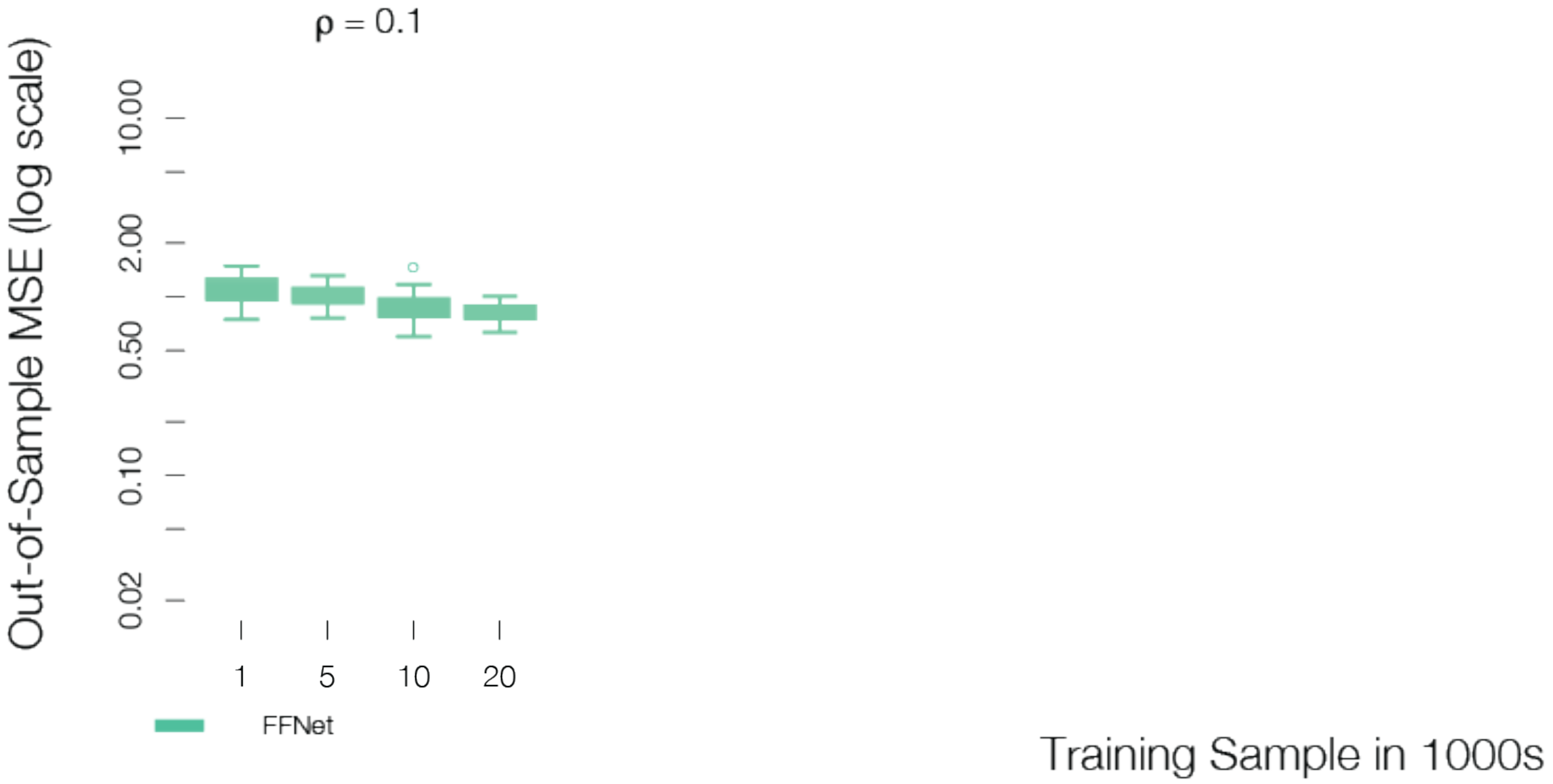


$t \sim U\{0, 1, .., 6\}$

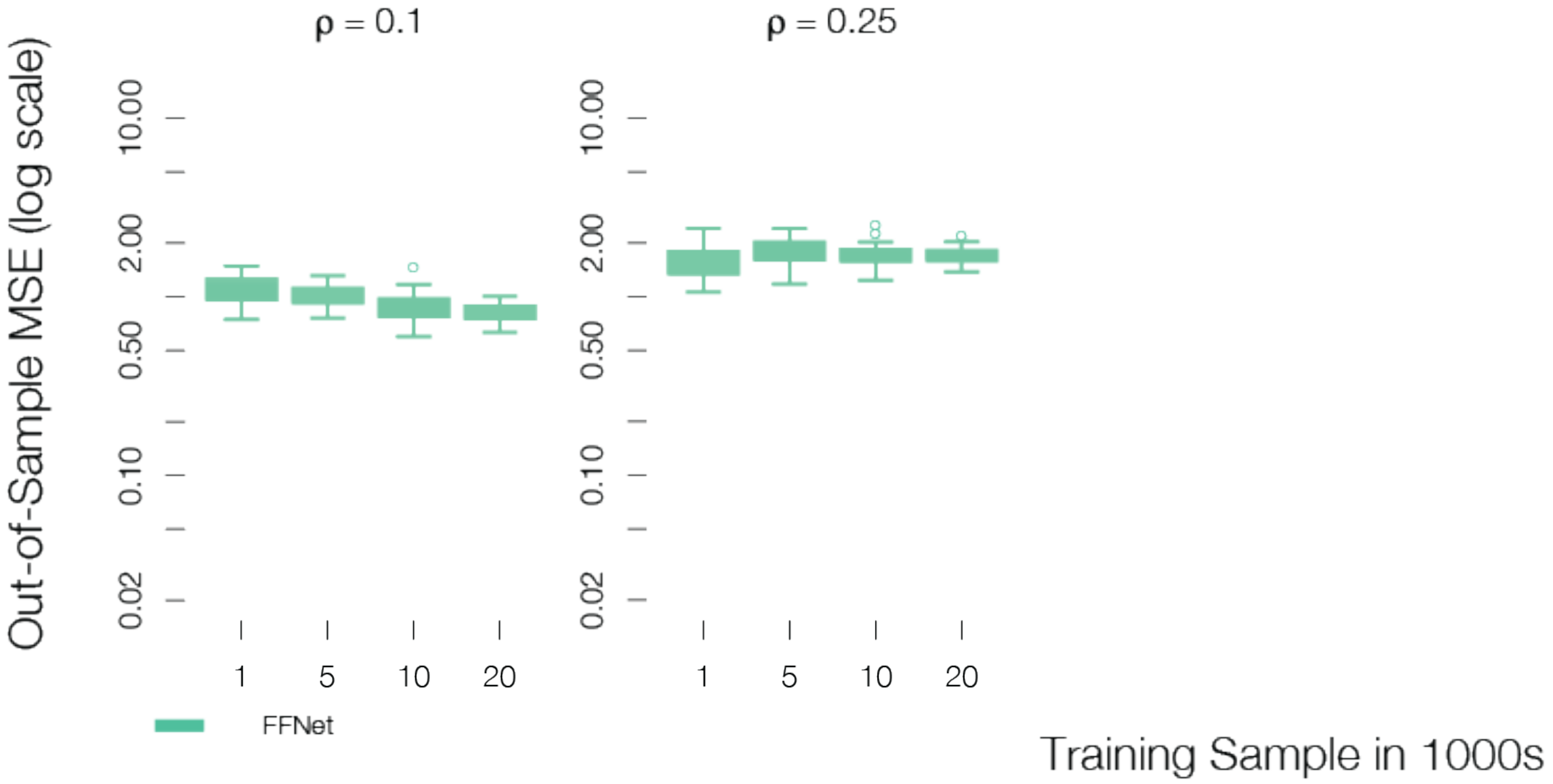


ρ lets us smoothly vary the correlation between sales and price

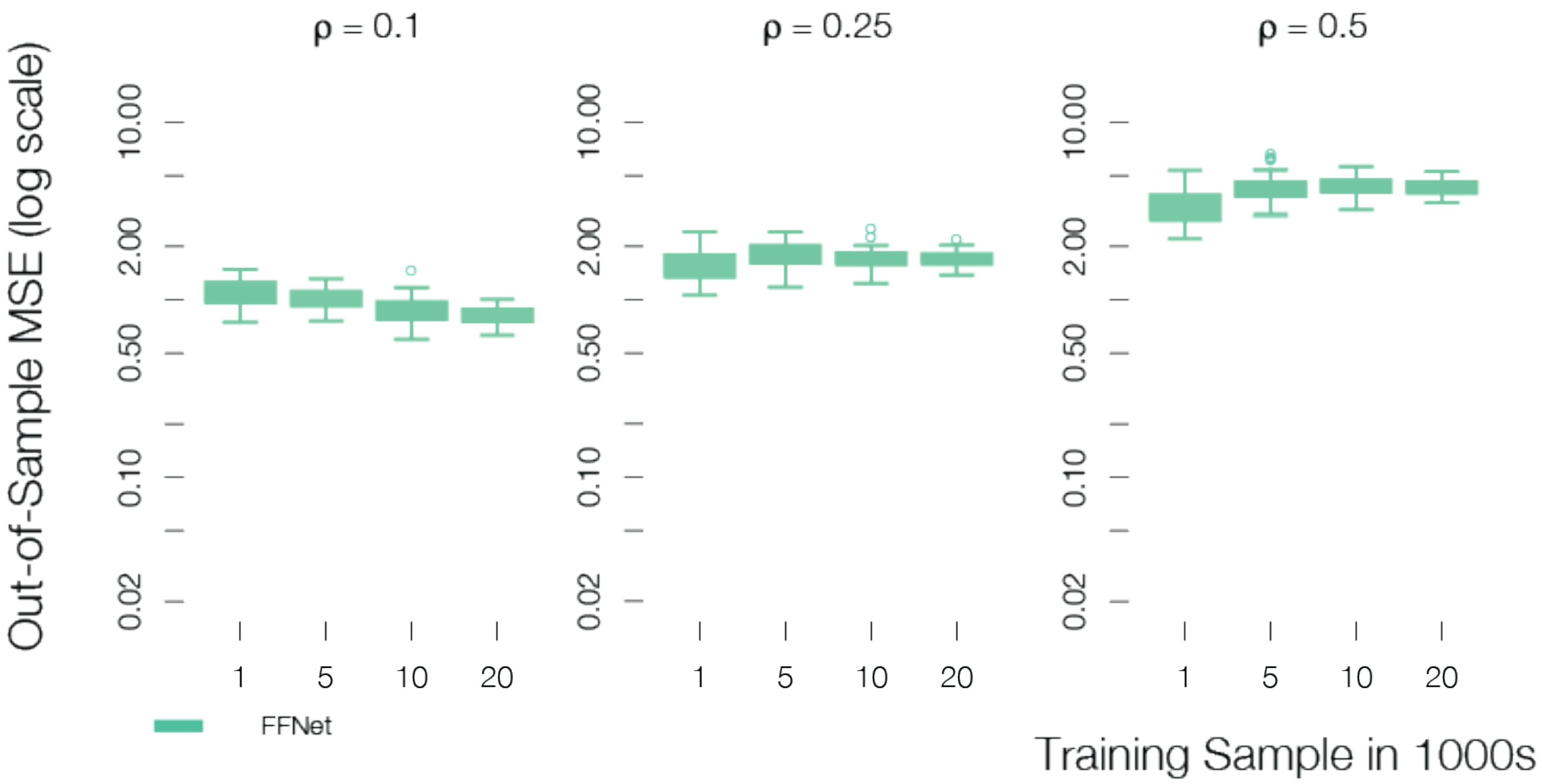
Simulation – low dimensional feature space



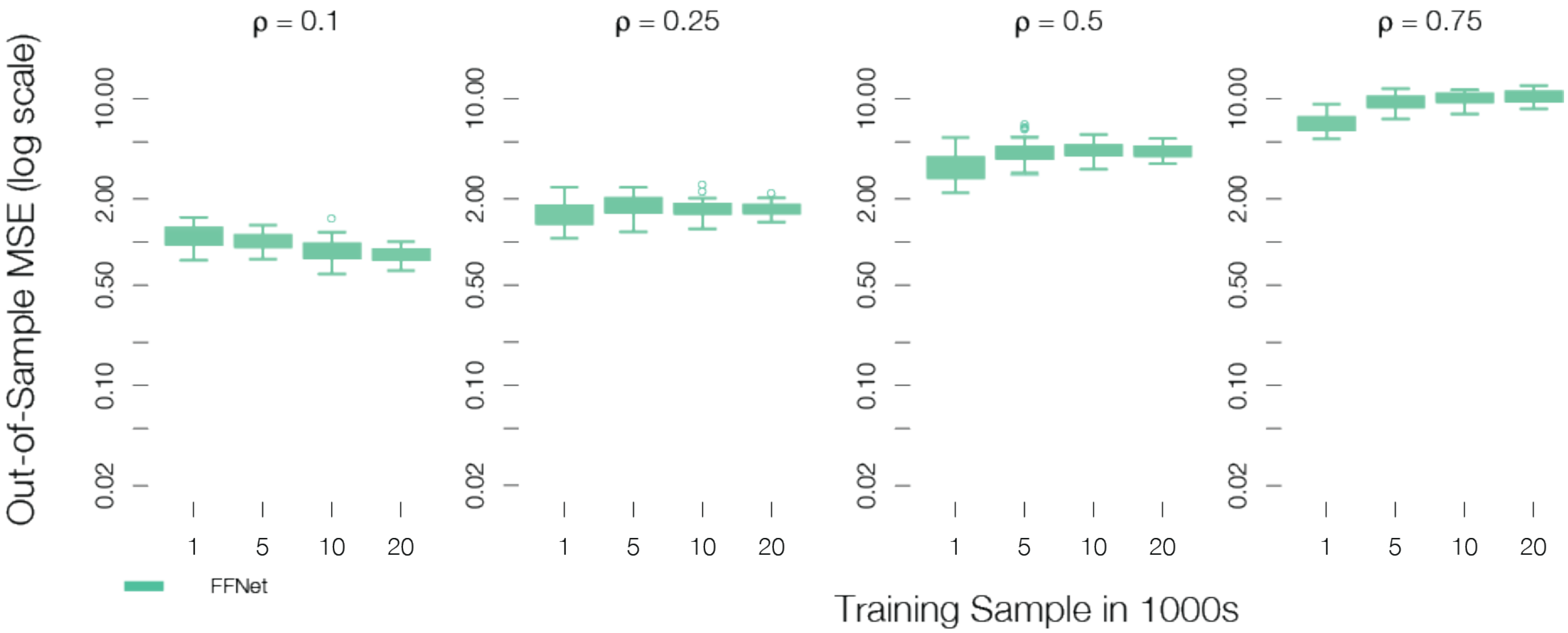
Simulation – low dimensional feature space



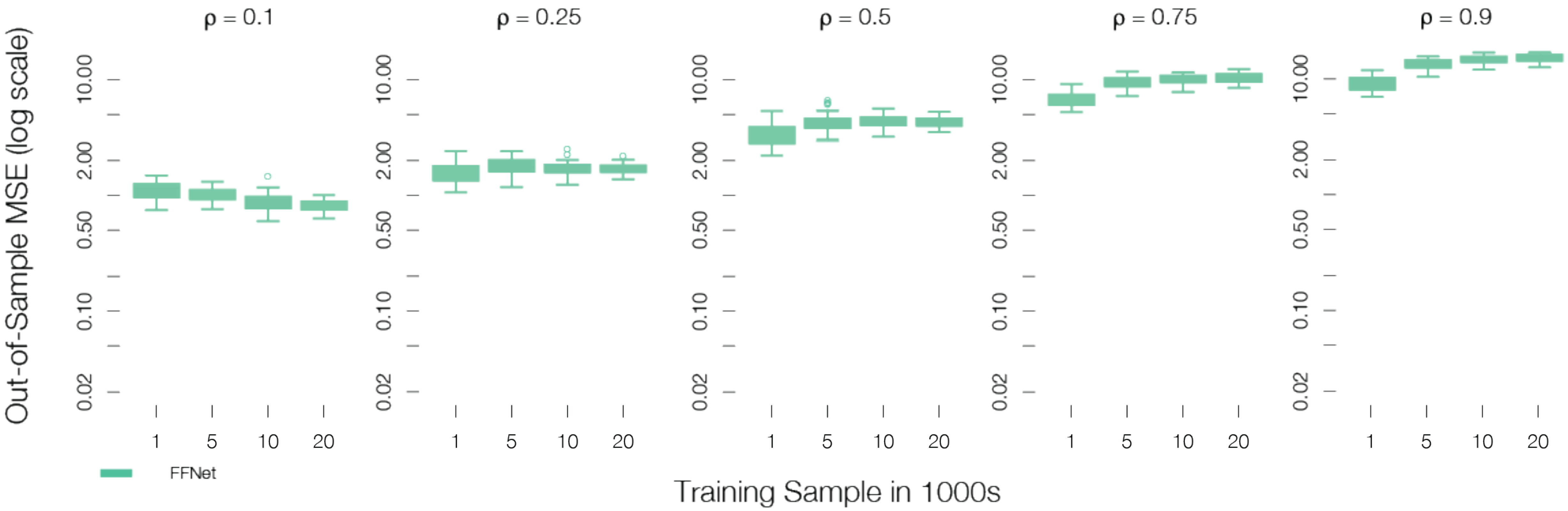
Simulation – low dimensional feature space



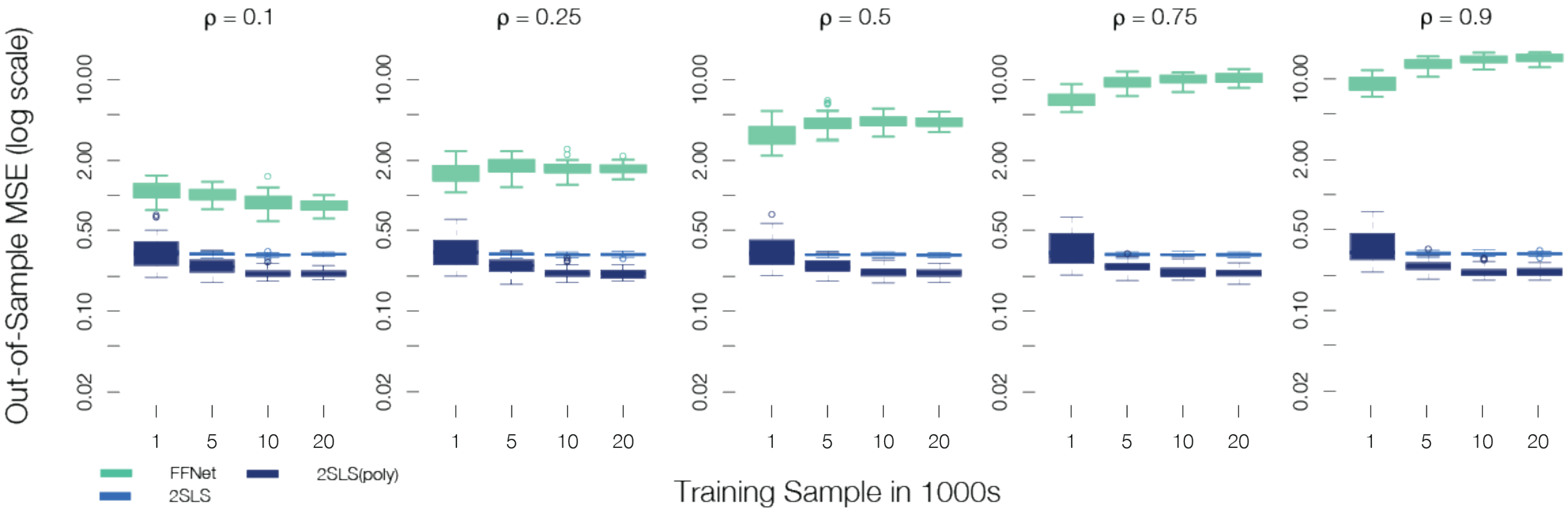
Simulation – low dimensional feature space



Simulation – low dimensional feature space

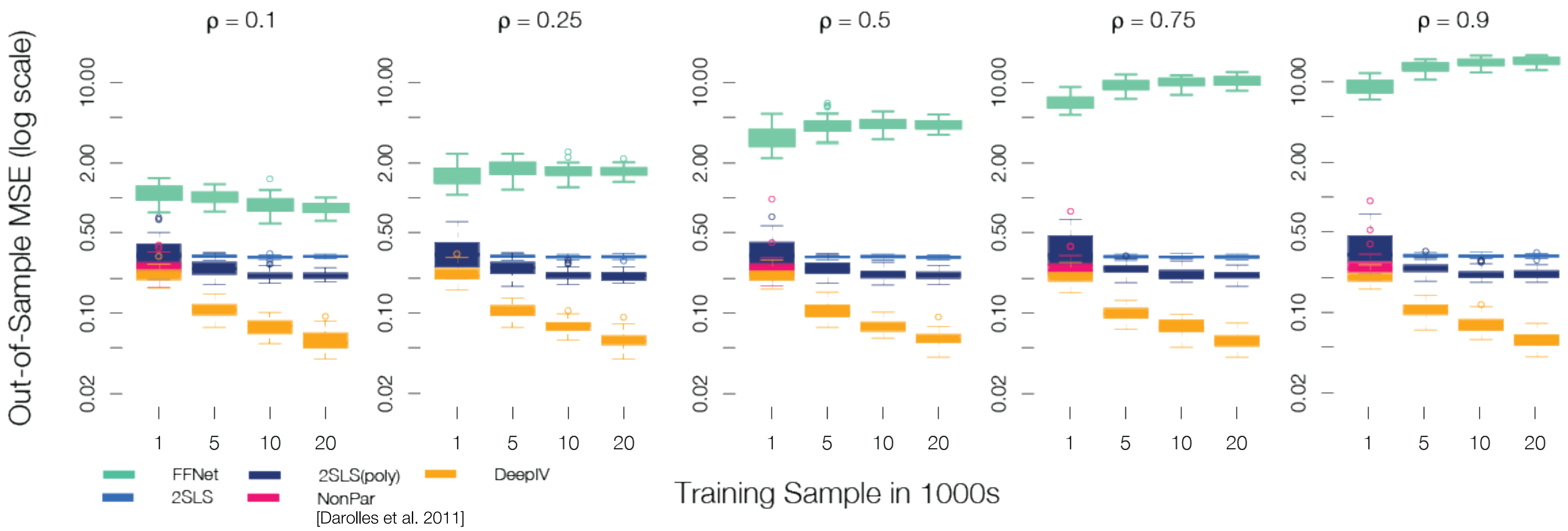


Simulation – low dimensional feature space



Simulation – low dimensional feature space

Simulation – low dimensional feature space



Implications and future directions

- We recover **heterogeneous treatment effects** in settings with **unobserved confounding effects** for both **discrete** and **continuous** variables… and SGD **scales** naturally to **very large datasets.**
- Can leverage the flexibility of deep nets for rich data types. E.g. **raw text** in our Bing ads application experiments / **images** in simulation.

Future work:

- Methods for **uncertainty** estimates over predictions.

Code and paper available at **http://bit.ly/DeepIV** **Poster #127**