# Robust Latent Data Representations

Larry Samuelson[1] Jakub Steiner[2,3]

[1]Yale
[2]Zurich U
[3]Cerge-Ei

London, September 2026

# The Question

Reasoning about data-generating process vs about data

- when data refute one's statistical theory

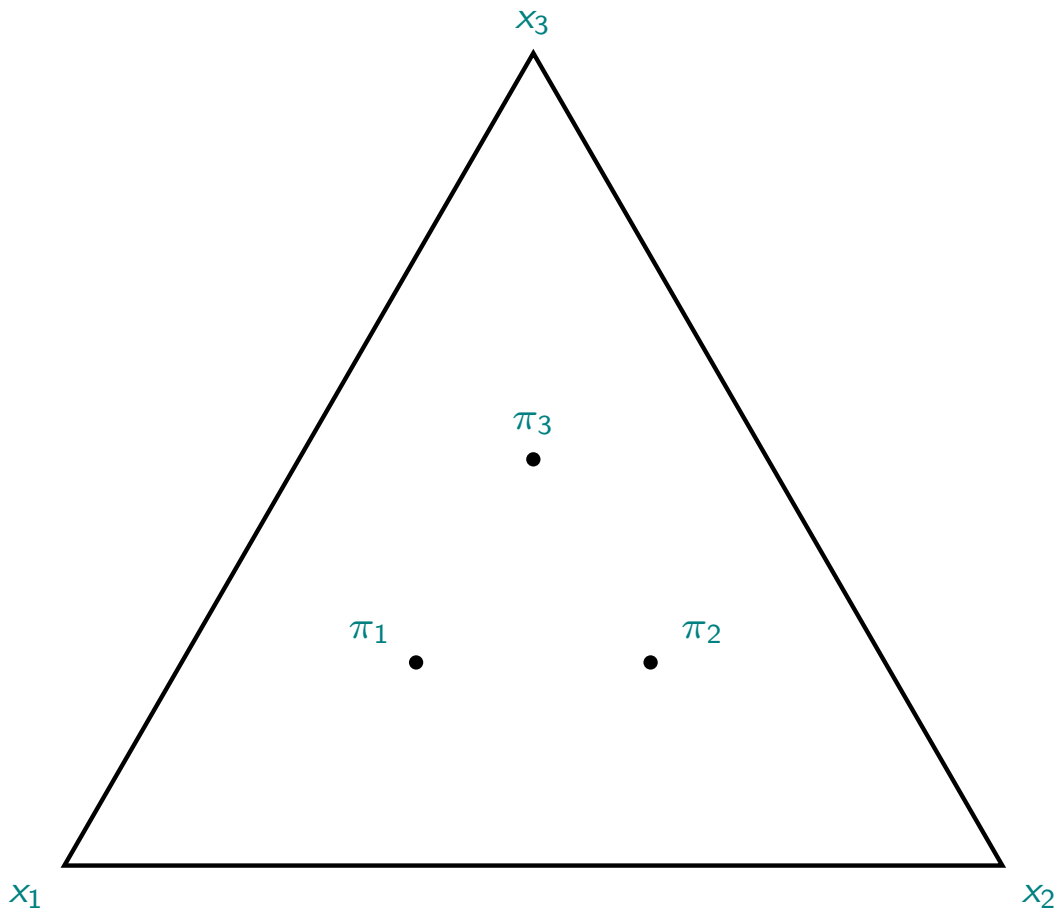
E.g. principal observes outputs of her workforce

- no combination of hidden efforts explains the output frequencies
- beliefs about individual and aggregate effort?

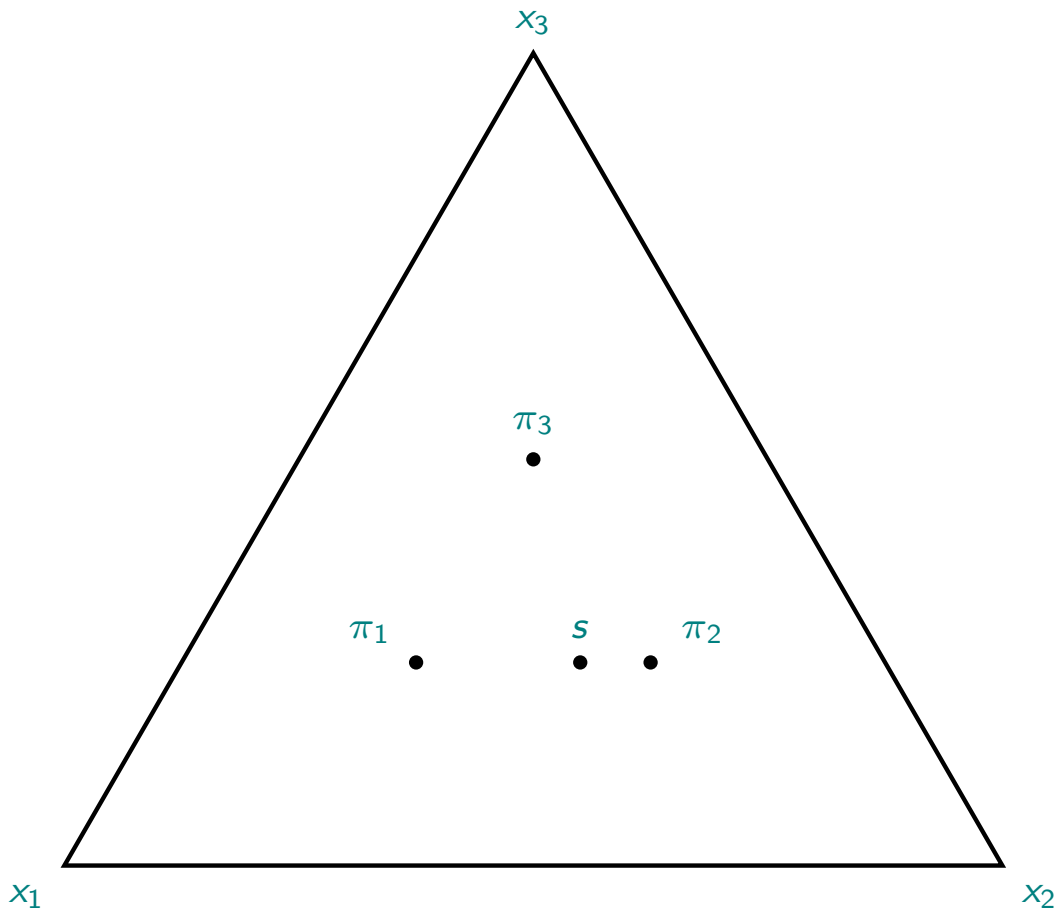
Complementary to Berk-Nash:

- our analyst engages with the wedge between predictions and data

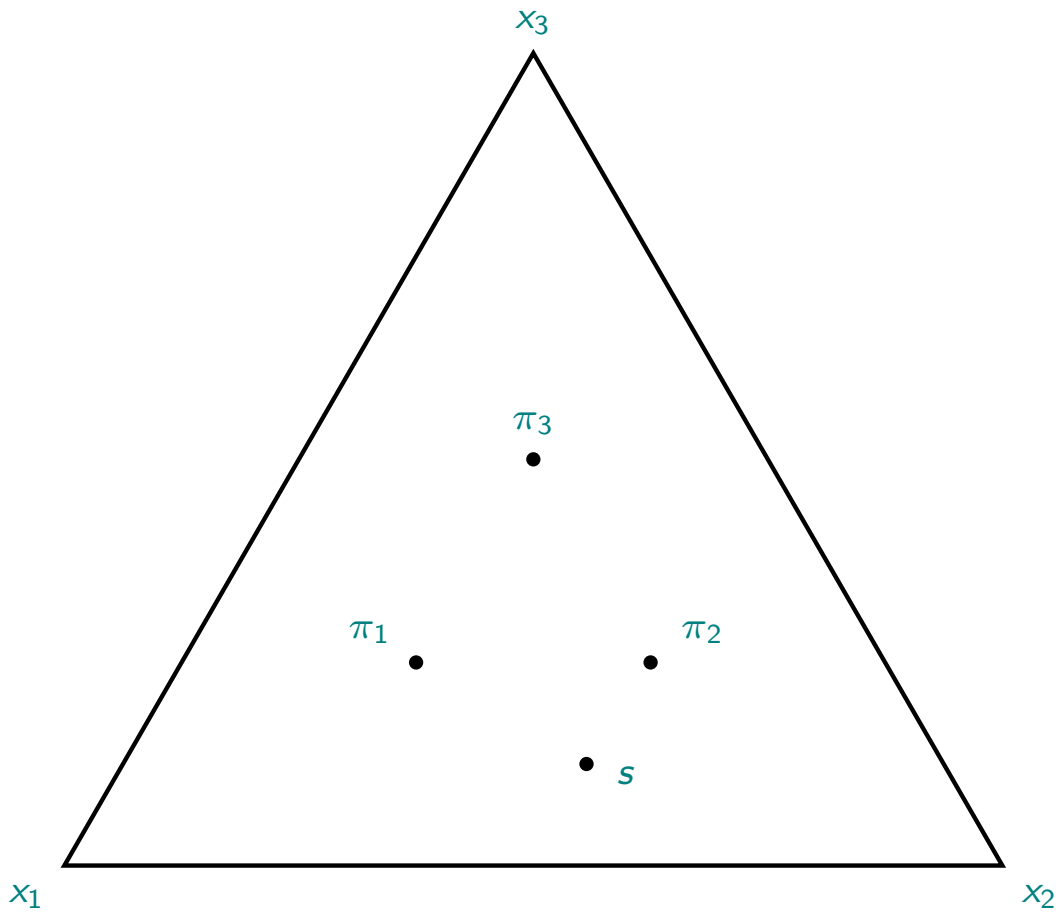
# Sample and Primitive Types



# Sample and Primitive Types



# Sample and Primitive Types



# Sample and Primitive Types

observations $x_1, \ldots, x_n$ from finite $\mathcal{X}$

sample $s(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x^i = x}, \qquad x \in \mathcal{X}$

- extension to $s \in \operatorname{int} \Delta(\mathcal{X})$

primitive theory $\mathcal{T} = \{\pi_z\}_{z \in \mathcal{Z}}$

- $\pi_z$ are primitive types, $z$ from finite $\mathcal{Z}$
- **affine independence**

# Example

the analyst is a principal

she observes the outputs $x^i \in \mathcal{X}$ of her workers $i = 1, \ldots, n$

her theory $\mathcal{T} = \{\pi_\ell, \pi_h\}$ specifies stochastic output for low and high effort

how does she form beliefs about

- effort shares?
- effort $z^i$ of a worker $i$, given his output $x^i$?

# Sample Model

belief about the representative sample element:

$$q^s_{XZ} \in \Delta(\mathcal{X} \times \mathcal{Z})$$

- it is the probability the analyst assigns to

$$(x^i, z^i) = (x, z), \qquad \text{for uniformly random } i$$

derived distributions:

- sample types $q^s_{X|Z=z}$
- sample prior $q^s_Z$

# Realism

**A1**

$q_X^s = s$.

belief about output of representative worker matches observed outputs

# Diagnostic Conservatism

**A2**

The analyst preserves the diagnostic likelihood ratio:

$$\frac{q^s_{X|Z=z}(x)}{q^s_{X|Z=z'}(x)} = \frac{\pi_z(x)}{\pi_{z'}(x)} \qquad \text{for all } z, z' \in \operatorname{supp} q^s_Z \text{ and all } x \in \mathcal{X}.$$

principal has a fixed view of how each $x$ is diagnostic of high vs low effort

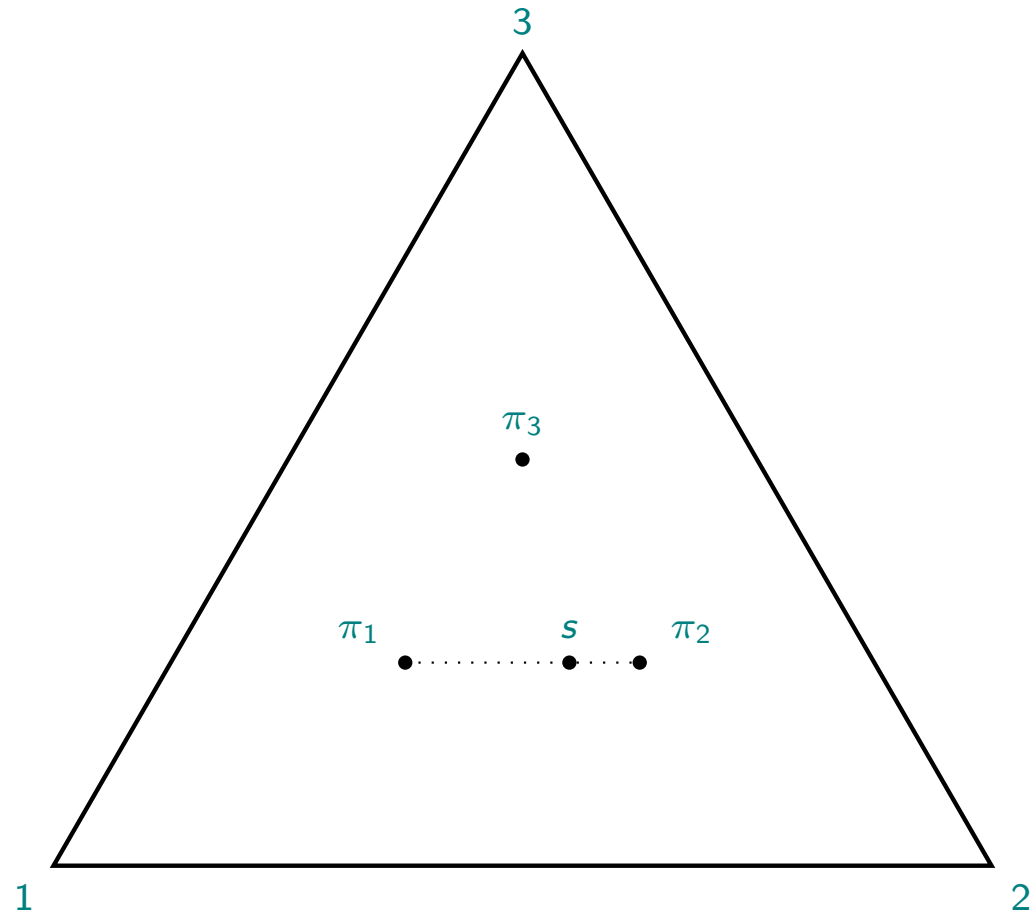
**motivation**:

- realism may force type adjustments $q^s_{X|Z=z}(x) = \pi_z(x) d_z(x)$.
- since $z$ isn't observed, the analyst chooses $d(x)$ independent of $z$

# Perfect Fit

**A3**

If $s \in \mathrm{co}(\mathcal{T})$, then $q^s_{X|Z=z} = \pi_z$ for all $z \in \operatorname{supp} q^s_Z$.



# Continuity

**A4**

The map $s \mapsto q_Z^s$ is continuous on $\operatorname{int} \Delta(\mathcal{X})$.

The analyst extends the perfect-fit continuously.

# MLE-Bayes

**stage 1: maximum-likelihood estimation.**

generative prior:

$$p_Z^s := \arg\max_{p_Z \in \Delta(\mathcal{Z})} \sum_{x \in \mathcal{X}} s(x) \ln p_X(x) \qquad \text{where } p_X := \sum_{z \in \mathcal{Z}} p_Z(z) \pi_z$$

generative model:

$$p_{XZ}^s(x, z) := p_Z^s(z) \pi_z(x)$$

**stage 2: Bayes' updating.**

MLE-Bayes sample model:

$$q_{XZ}^{\text{MLE}, s}(x, z) = s(x) p_{Z|X=x}^s(z)$$

# Representation Theorem

**Theorem**

*$s \mapsto q^s_{XZ}$ satisfies A1–A4 if and only if $q^s_{XZ} = q^{\mathrm{MLE},\, s}_{XZ}$.*

# Stereotypes

- observations $x = (x_1, x_2)$
- $x_1 \in \{L, H\}$ is labor output
- $x_2 \in \{0, 1\}$ is demographic group
- latent type $z \in \{\ell, h\}$ is talent

primitive theory:

| | $(L,0)$ | $(H,0)$ | $(L,1)$ | $(H,1)$ |
|---|---|---|---|---|
| $\pi_\ell$ | $3/5$ | $1/5$ | $1/10$ | $1/10$ |
| $\pi_h$ | $3/10$ | $1/10$ | $1/10$ | $1/2$ |

# Stereotypical Diagnostic Content

let $\lambda(x) := \frac{\pi_h(x)}{\pi_\ell(x)}$

group 0: $\lambda(L,0) = \lambda(H,0) = \frac{1}{2}$

- high output is interpreted as luck, not talent

group 1: $1 = \lambda(L,1) < \lambda(H,1) = 5$

- output is strongly diagnostic of talent

# What Evidence Corrects

sample:

$$s = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

MLE-Bayes sample types:

| | $(L,0)$ | $(H,0)$ | $(L,1)$ | $(H,1)$ |
|---|---|---|---|---|
| $q^s_{X\|Z=\ell}$ | $1/3$ | $1/3$ | $1/4$ | $1/12$ |
| $q^s_{X\|Z=h}$ | $1/6$ | $1/6$ | $1/4$ | $5/12$ |

predictive meanings are adjusted

# What Evidence Does Not Correct

diagnostic likelihood ratios are preserved:

$$\frac{q^s(H, 0 \mid h)}{q^s(H, 0 \mid \ell)} = \frac{q^s(L, 0 \mid h)}{q^s(L, 0 \mid \ell)} = \frac{1}{2}$$

$$1 = \frac{q^s(L, 1 \mid h)}{q^s(L, 1 \mid \ell)} < \frac{q^s(H, 1 \mid h)}{q^s(H, 1 \mid \ell)} = 5$$

# Setting

continuum of risk-neutral workers

latent efforts $z^i \in \mathcal{Z} = \{\ell, h\}$, $c_\ell = 0 < c_h$

labor outcomes $x^i \in \mathcal{X} = \{L, H, N\}$

objective technology $\rho_z \in \Delta(\mathcal{X})$; known by the workers

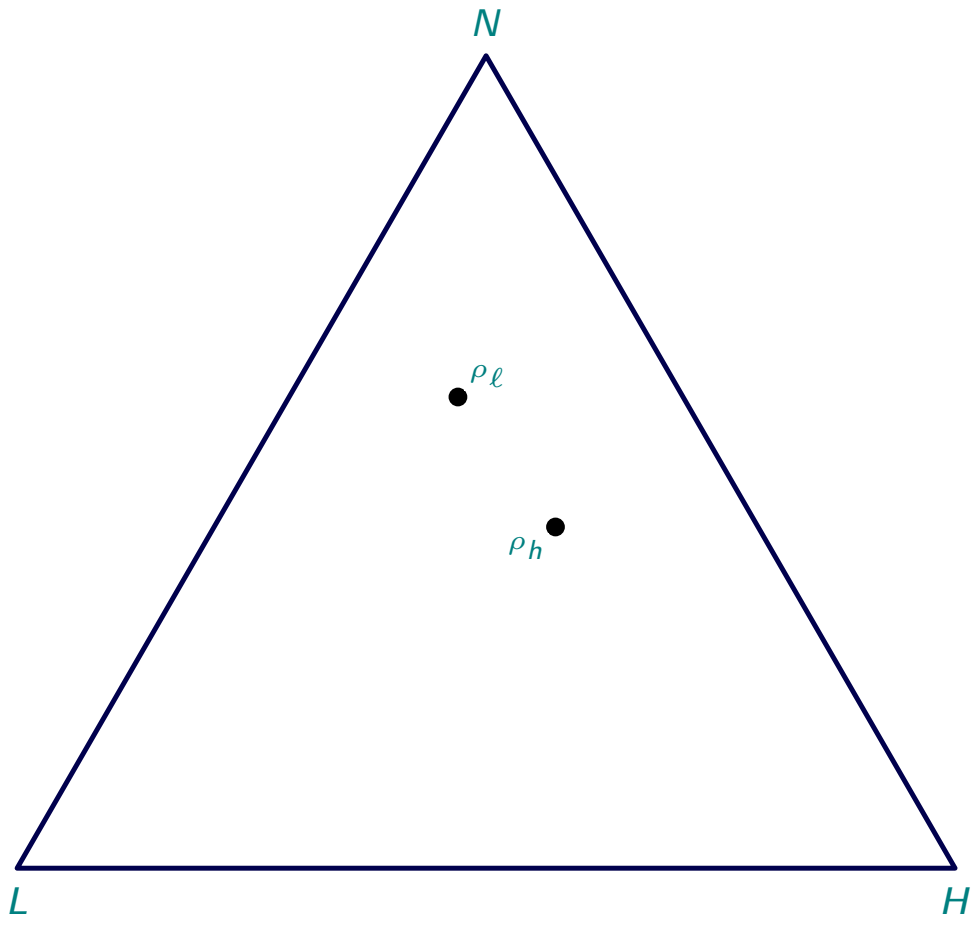
principal has revenue $r(x)$ from outcome $x$ and commits to wages $w(x)$

two versions of the principal:

1. well-specified
2. misspecified, using sample model

# Well-specified Principal

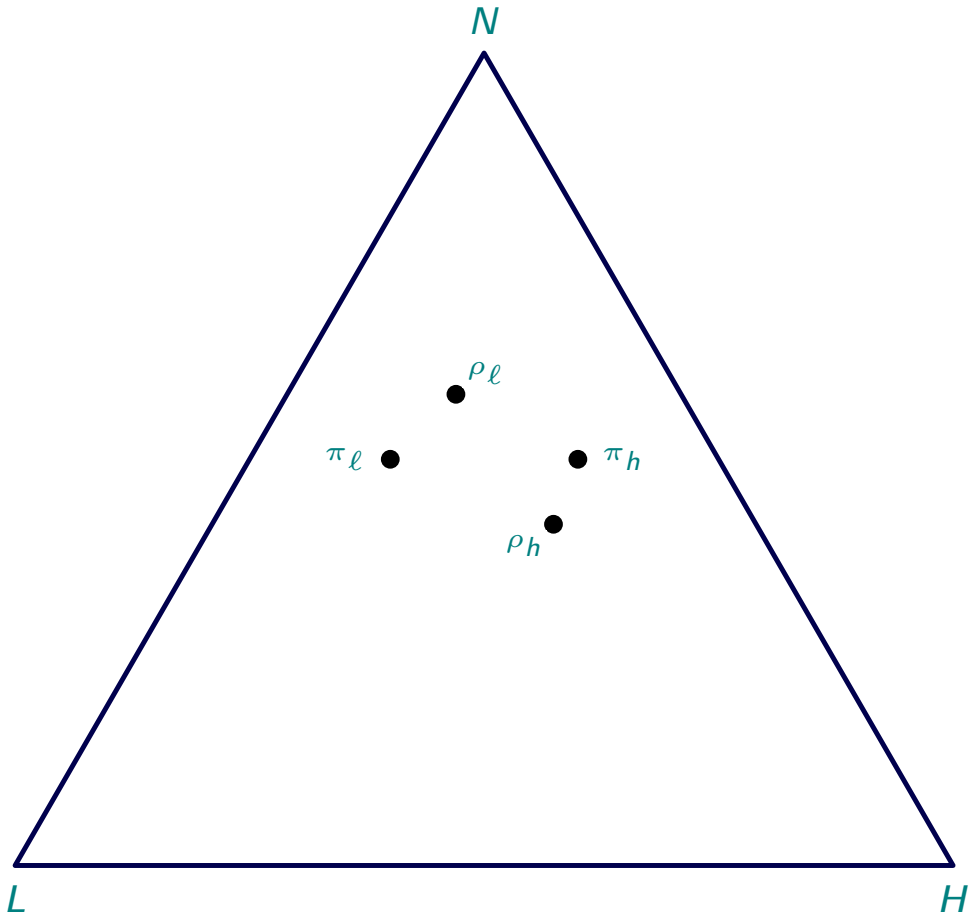
$$r(L) = 0, \qquad r(N) = 1, \qquad r(H) = 2, \qquad c_h = 0.0184.$$



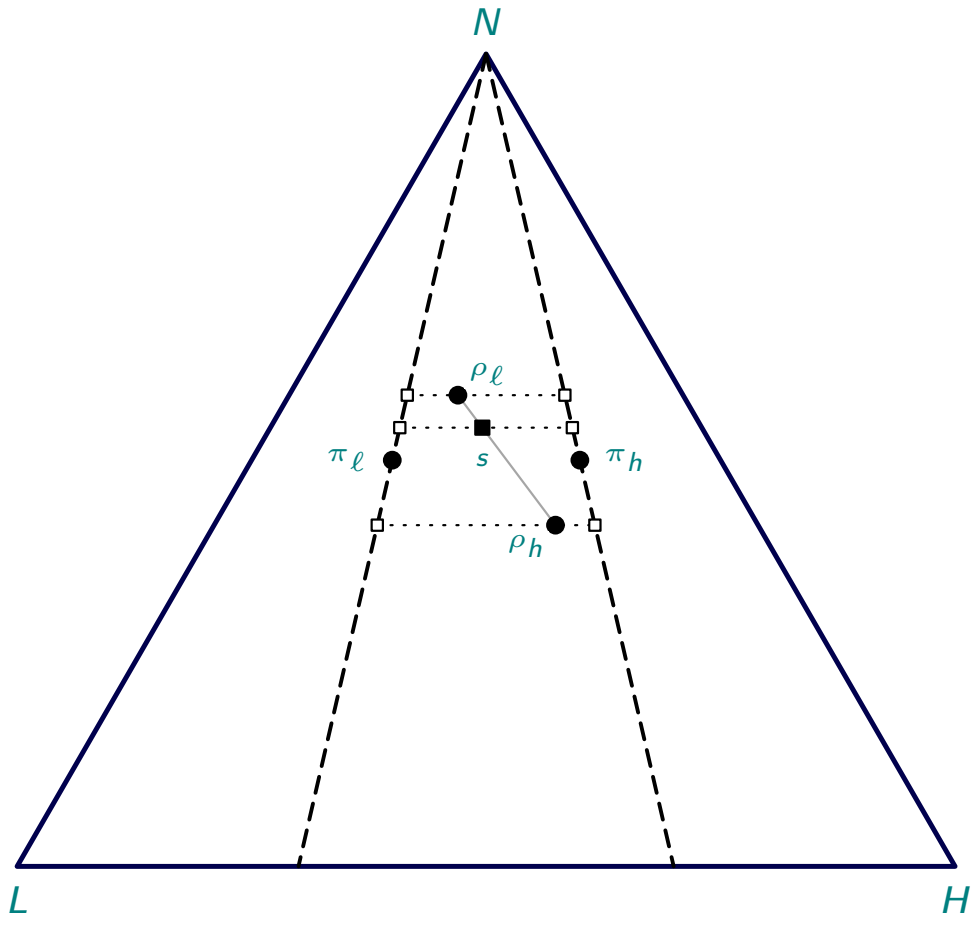

$$w(H) = 0.1 \qquad w(L) = w(N) = 0$$

# Misspecified Principal

misspecified primitive types $\pi_z \neq \rho_z$



# Misspecified Principal: Sample Model



# Sample Equilibrium

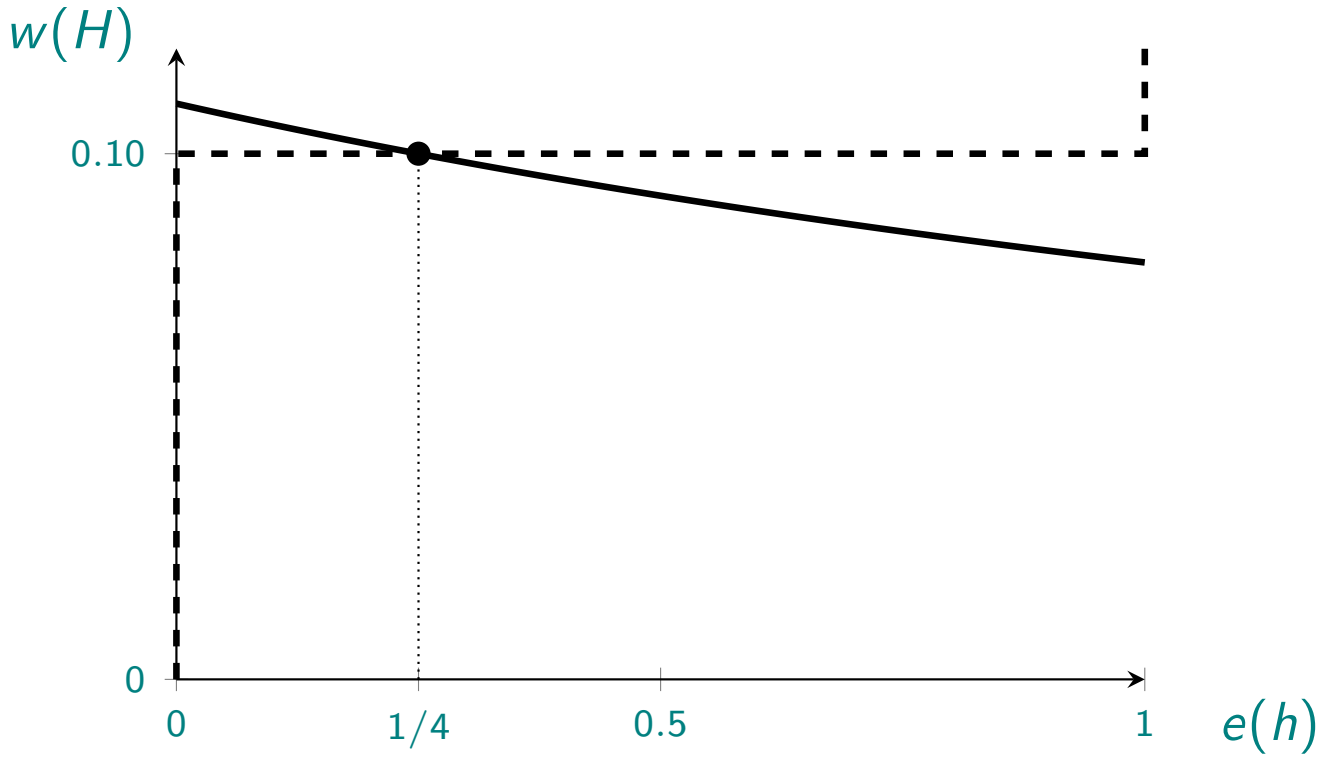
a fixed point $(e^*, s^*, w^*)$:

1. given wage profile $w^*$, workers choose an optimal mixed effort $e^*$,
2. sample $s^*$ is generated by $e^*$ according to the objective technology,
3. given subjective sample types, principal optimizes wage profile $w^*$.

for the considered parametrization:

- unique mixed equilibrium
- the principal chooses the well-specified contract
- achieves only interior effort

# Sample Equilibrium



# Misspecified Principal: Generative Model

Given $s$, the principal estimates generative model $p^s_{XZ}(x, z)$.

By construction, she keeps the primitive types: $p^s_{X|Z=z} = \pi_z$.

She chooses

$$w(H) = 0.092 \qquad w(L) = w(N) = 0$$

and thus fails to induce high effort.

# Partial Representation

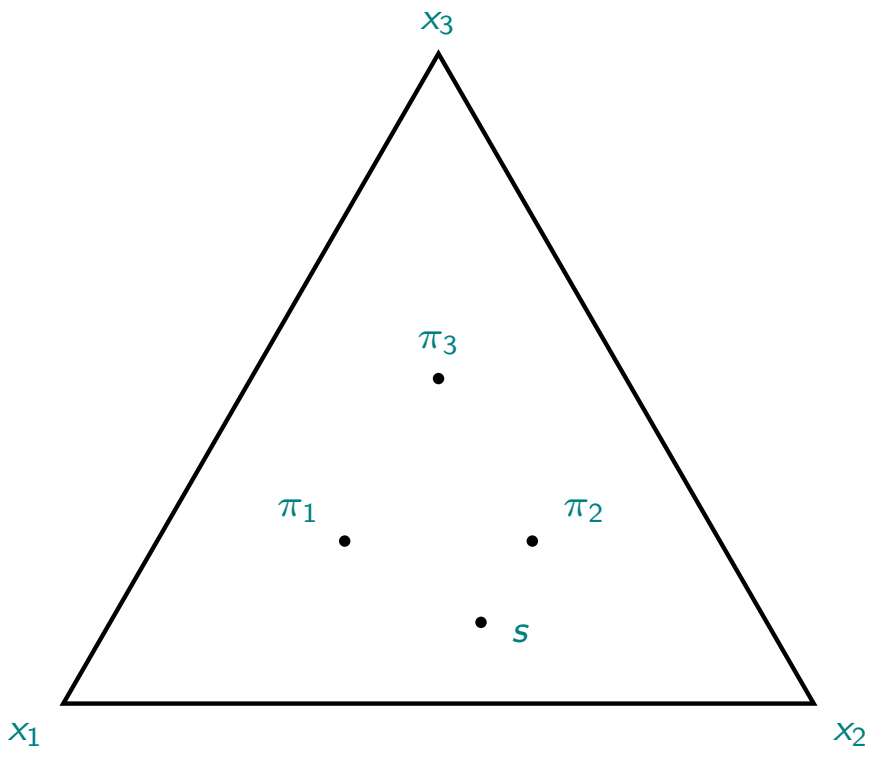
$q_{XZ}^{\mathrm{MLE},\, s,\mathcal{E}}$ – MLE-Bayes sample model restricted to types in $\mathcal{E} \subseteq \mathcal{Z}$

**fixed-support representation**

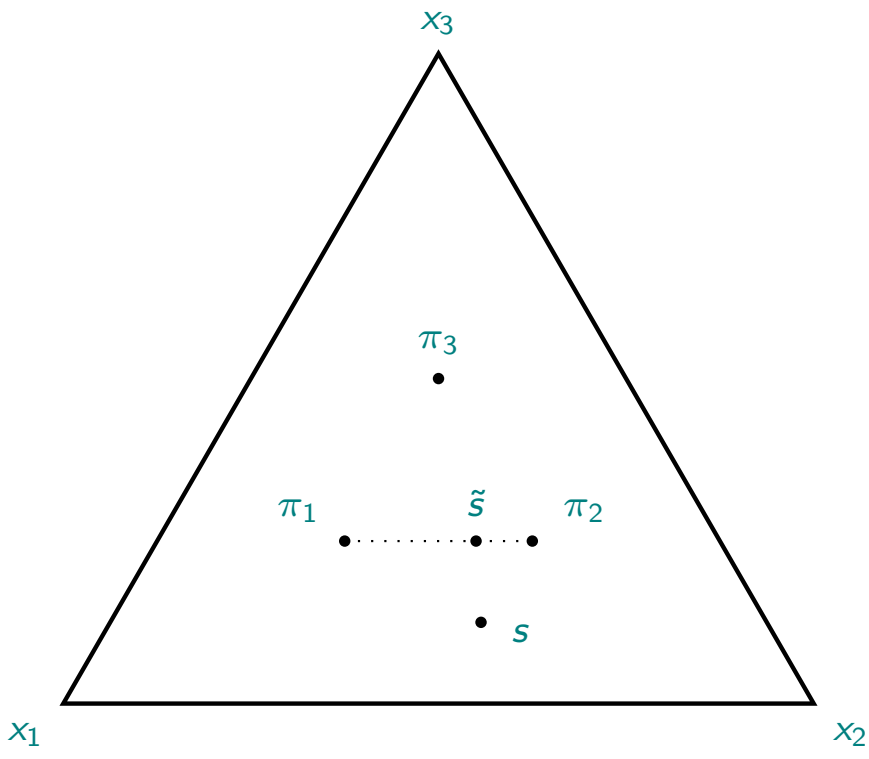
Fix $\mathcal{E} \subseteq \mathcal{Z}$. There exists at most one sample model $q_{XZ}^{s}$ satisfying A1 and A2 with $\operatorname{supp} q_{Z}^{s} = \mathcal{E}$. If it exists, then

$$q_{XZ}^{s} = q_{XZ}^{\mathrm{MLE},\, s,\mathcal{E}}.$$

# Continuous Expansion

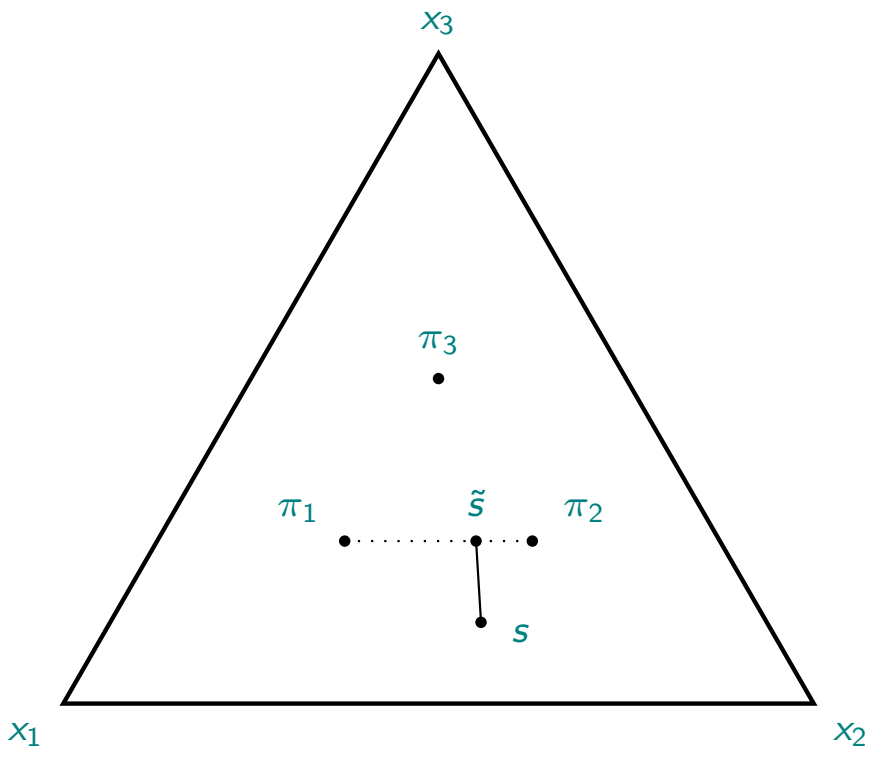


# Continuous Expansion




anchor sample $\tilde{s} := p_X^s \in \mathrm{co}(\mathcal{T})$

# Continuous Expansion




constant MLE estimate $p_Z^{s_t}$ for all $s_t$ on the line

$q^{s_t} = p_Z^{s_t}$ for all $s_t$ on the line

- this holds for $\tilde{s}$
- continuous expansion preserves the set of employed types

# Literature

**reconciling misspecification**: Ortoleva'12, Cho&Kasa'15, Ba'26, Gagnon-Bartsch et al.'18, Karni and Vierø'13, Lanzani'25

**variational Bayes**: Dempster et al.'77, Jordan et al.'99, Blei et al.'17, Kingma&Welling'13, Aridor et al.'20,'25

**mixture estimates**: Lazarsfeld&Henry'68, Kiefer&Wolfowitz'56, Lindsay'83, Heckman&Singer'84, Bonhomme et al.'22

**formal analogy to rational inattention**: Matějka&McKay'15, Caplin&Dean'15

# Summary

Reasoning about observations that refute one's own statistical model

- adjustments of some aspects of types
- rigidity in other aspects