# The Limits of Experimental Design: Covariate Balance Beyond Low Dimension

Max Cytrynbaum

Yale University

November 2026

# Introduction

In experiments, **balancing covariates** improves the efficiency of causal parameter estimates.

**Matched Pairs.** Let $D \in \{0, 1\}$ and $\Pr(D = 1) = 1/2$. Match units into pairs $X_i \approx X_j$, assign $D_i = 1$ and $D_j = 0$ or vice versa.

Such designs make unadjusted estimation automatically semiparametrically efficient (Bai et al., 2022, 2026).

Let $V^*$ be the Hahn (1998) efficiency bound for estimating $\theta_0 =$ ATE. Then difference in means $\widehat{\theta} = \bar{Y}_1 - \bar{Y}_0$ has

$$\sqrt{n}(\widehat{\theta} - \theta_0) \Rightarrow N(0, V^*).$$

# Tight Matching in Practice

Efficiency follows by showing optimally matched pairs are asymptotically tight, with $\frac{1}{n}\sum_{p=1}^{n/2} \|X_{i_p} - X_{j_p}\|_2^2 \to_p 0$.

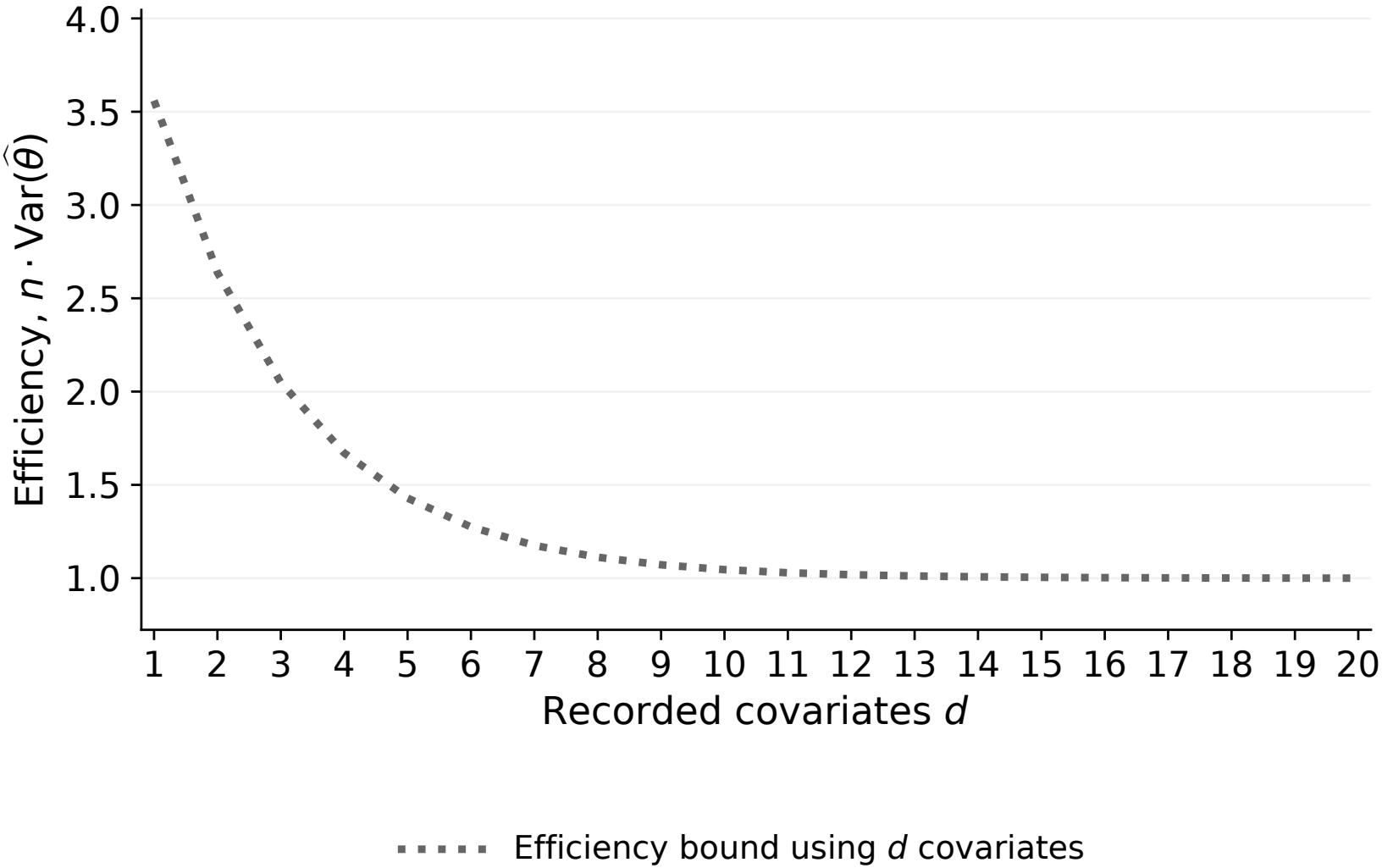
For $d = \dim(X)$, nearest-neighbor distances have scale $n^{-1/d}$.

For $d = 6$, halving matching objective requires increasing sample size $n \to 8n$ in the best case. For general $d$, $n \to 2^{d/2} \cdot n$.

**Example.** OpenResearch funded \$40 million study on UBI. Experiment design matched 3,000 participants into triples using $d =$ several dozen covariates (Broockman et al., 2024).

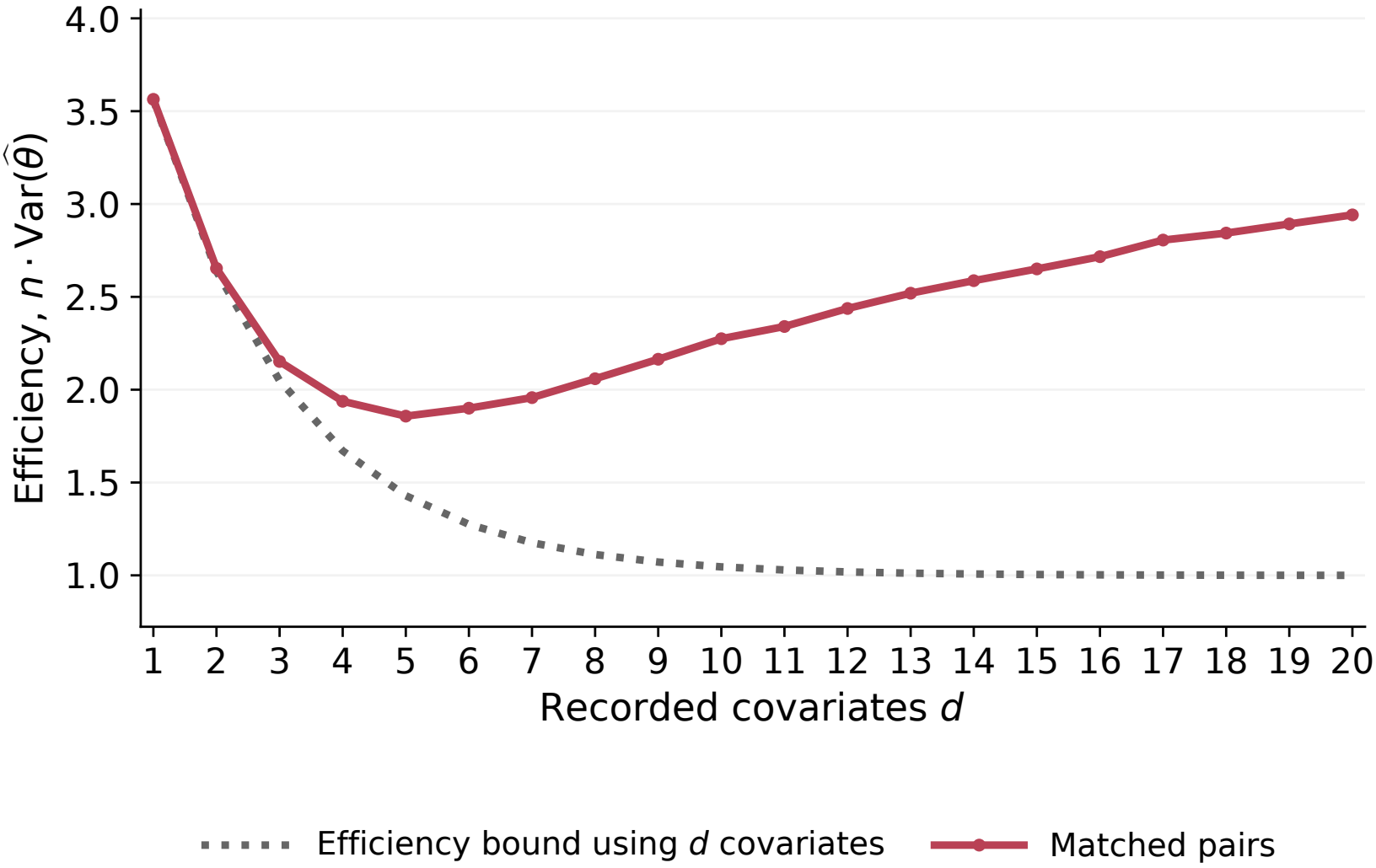
Is matched pairs actually fully efficient in this regime?

# Simulation



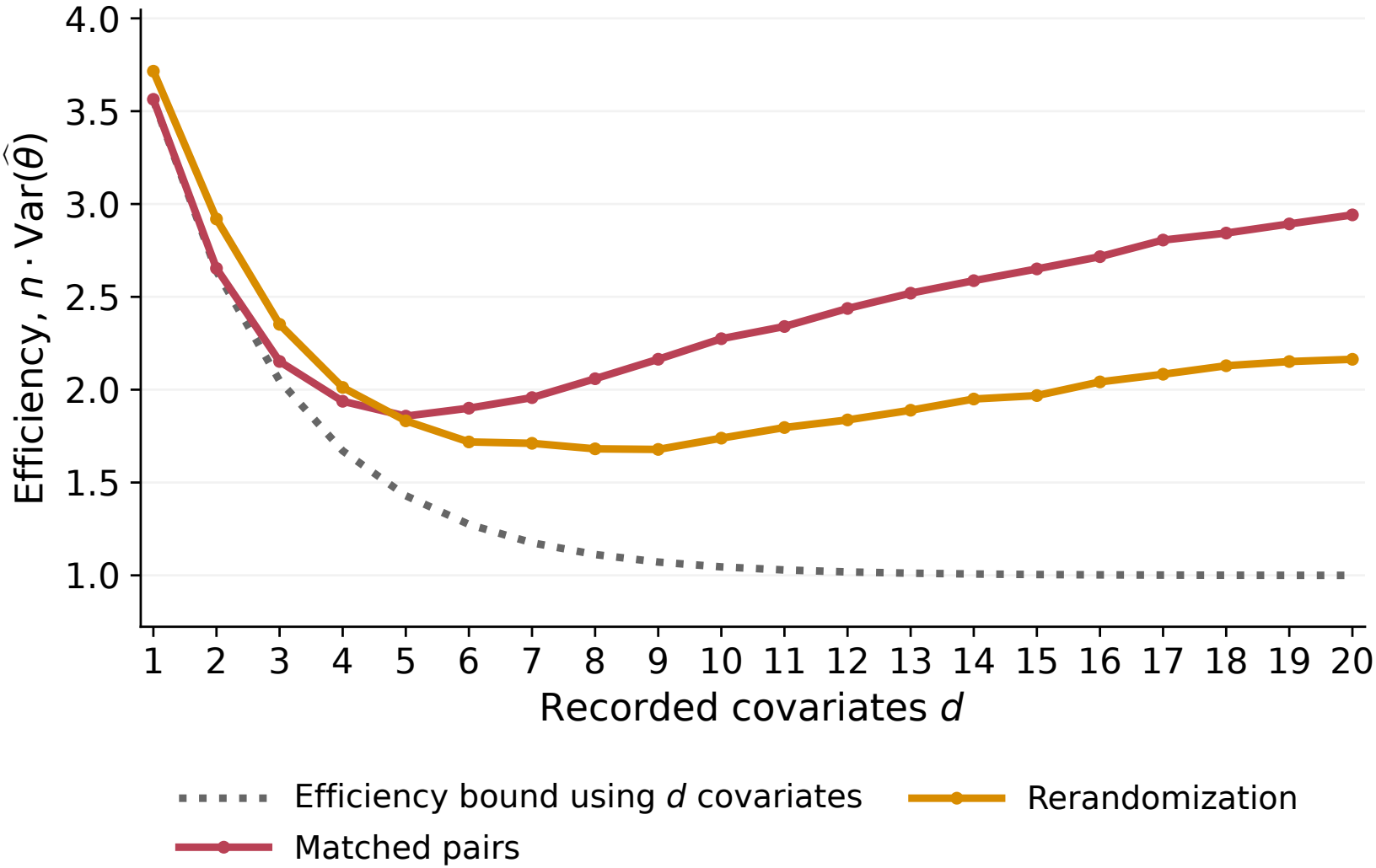

Sample size $n = 240$. Model $E[Y(a)|X] = c + L'X + X'QX$.

# Simulation



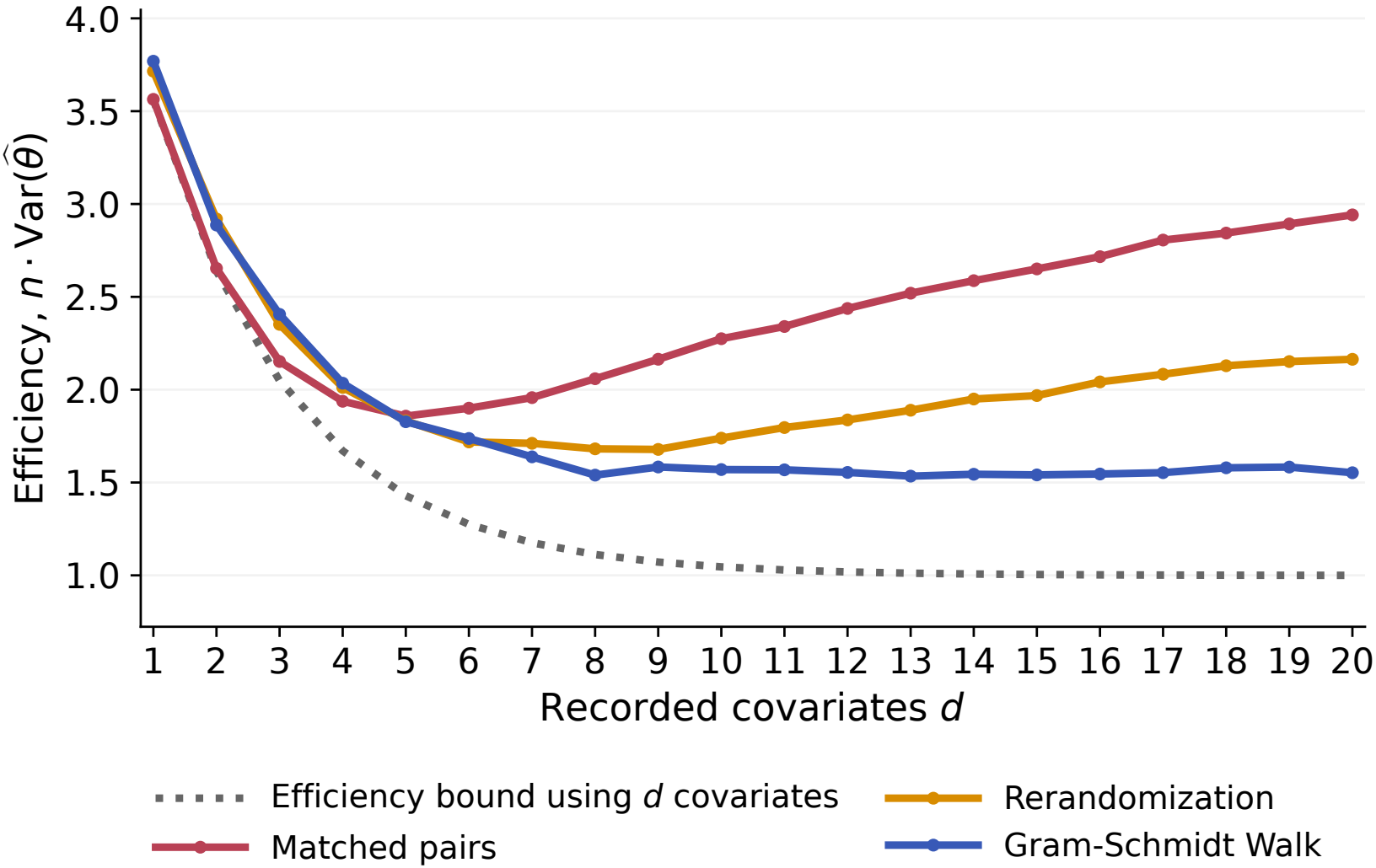

Sample size $n = 240$. Model $E[Y(a)|X] = c + L'X + X'QX$.

# Simulation




Sample size $n = 240$. Model $E[Y(a)|X] = c + L'X + X'QX$.

# Simulation




Sample size $n = 240$. Model $E[Y(a)|X] = c + L'X + X'QX$.

# Setup

Asymptotic efficiency guarantees can be a very poor guide to finite-sample performance.

We introduce a **finite-sample efficiency theory** to guide the comparison and development of new experimental designs.

Let $(X_i, Y_i(0), Y_i(1))$ be iid from $P$, with $X_i \in [0,1]^d$.

Given $X_{1:n}$, require design $\sigma$ to draw $D_{1:n}$ independently of all potential outcomes, with $\Pr_\sigma(D_i = 1 | X_{1:n}) = 1/2$ for each $i$.

Observe $Y_i = Y_i(D_i)$, and let $Z_i = 2D_i - 1 \in \{\pm 1\}$. The Horvitz-Thompson estimator can be written

$$\widehat{\theta} = \frac{2}{n} \sum_{i=1}^{n} Z_i Y_i.$$

# Efficiency Theory

For a design $\sigma$ and DGP $P$, define the **excess variance:**

$$\mathcal{V}_n(\sigma, P) \equiv n \operatorname{Var}_\sigma(\widehat{\theta}) - V^*(P).$$

Define the **outcome model** $m_P(X) = E_P[(Y(1) + Y(0))/2|X]$.
Calculation shows excess variance is due to covariate imbalance:

$$\mathcal{V}_n(\sigma, P) = 4n \cdot E_\sigma\left[\left(n^{-1}\sum_{i=1}^{n} Z_i m_P(X_i)\right)^2\right].$$

Classical pointwise, asymptotic efficiency requires $\mathcal{V}_n(\sigma_n, P) \to 0$.

Instead, here we study finite-sample minimax bounds over $P \in \mathcal{P}$ defined by regularity of outcome model $m_P(X)$. Results of form

$$\inf_{\sigma \in \mathcal{D}_n} \sup_{P \in \mathcal{P}} \mathcal{V}_n(\sigma, P) \asymp r_n$$

# Main Insights

Matched pairs is asymptotically inefficient unless $d = o(\log n)$, even if $m_P(X)$ is linear. Failure to exploit smoothness or additivity.

Motivated by this, develop new nonparametric Gram-Schmidt walk (GSW) designs targeting balance over smooth functions.

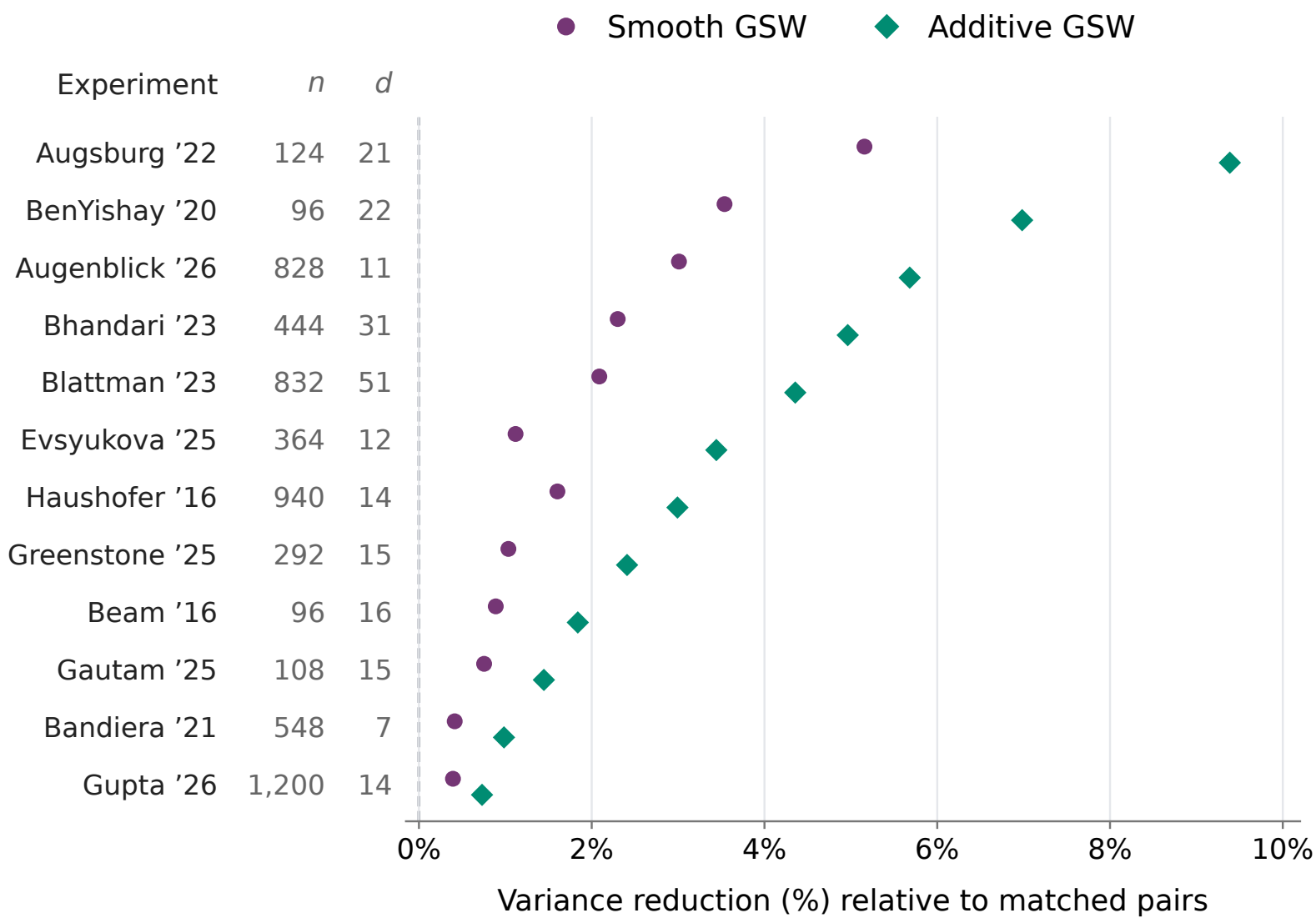
Such designs achieve much faster excess variance rates, adapting without prior knowledge to smoothness $s > 0$ of $m_P(X)$.

**Impossibility Theorem.** No design is uniformly asymptotically efficient over smooth outcome models unless $d = o(\log n)$.

This suggests a restricted efficiency target. Only attempt to balance additive functions $\mathcal{G}_{add}$ with $g(X) = \sum_{j=1}^{d} g_j(X_j)$.

Attains restricted efficiency with $d = o(n)$, up to logs.

# Empirical Preview



# Related Literature

*Efficiency of Fine Stratification.* Bai (2022), Bai et al. (2022, 2026), Cytrynbaum (2024, 2026a,b).

*Gram-Schmidt Walk.* Bansal et al. (2019), Harshaw (2021), Harshaw et al. (2024).

*Minimax Analysis.* Stone (1982), van der Vaart and van Zanten (2011), Fischer and Steinwart (2020), Kennedy et al. (2024).

# The Limits of Matched Pairs Designs

# Lower Bounds for Matched Pairs

Require covariates $X$ to have density $0 < \underline{p} \leq p(x) \leq \overline{p} < \infty$ on $[0,1]^d$, and $E_P[Y(a)^2] < \infty$. Define the linear model

$$\mathcal{P}_{\text{Lin}} = \{P : m_P(X) = \alpha + \gamma' X, \|\gamma\|_2 \leq 1\}.$$

**Theorem.** Let $\mathcal{M}_n$ be matched-pairs designs. For $n \geq 2$, $d \geq 1$

$$\inf_{\sigma \in \mathcal{M}_n} \sup_{P \in \mathcal{P}_{\text{Lin}}} \mathcal{V}_n(\sigma, P) \geq \frac{1}{2\pi e} \cdot n^{-2/d}$$

Note $n^{-2/d} = \exp((-2/d) \cdot \log n) \to 0$ if and only if $d = o(\log n)$.

**Corollary.** (Asymptotic Inefficiency) If $d \geq c \log n$ for $c > 0$, then

$$\liminf_{n \to \infty} \inf_{\sigma \in \mathcal{M}_n} \sup_{P \in \mathcal{P}_{\text{Lin}}} \mathcal{V}_n(\sigma, P) > 0.$$

# Matching Targets Rough Models

Let $\mathcal{P}_\beta$ contain laws with $\beta$-Hölder outcome models ($0 < \beta \leq 1$) satisfying $|m_P(x) - m_P(y)| \leq \|x - y\|_2^\beta$.

For a symmetric matching cost $c(x, y)$ and some $q > 0$, assume

$$a\|x - y\|_2^q \leq c(x, y) \leq A\|x - y\|_2^q.$$

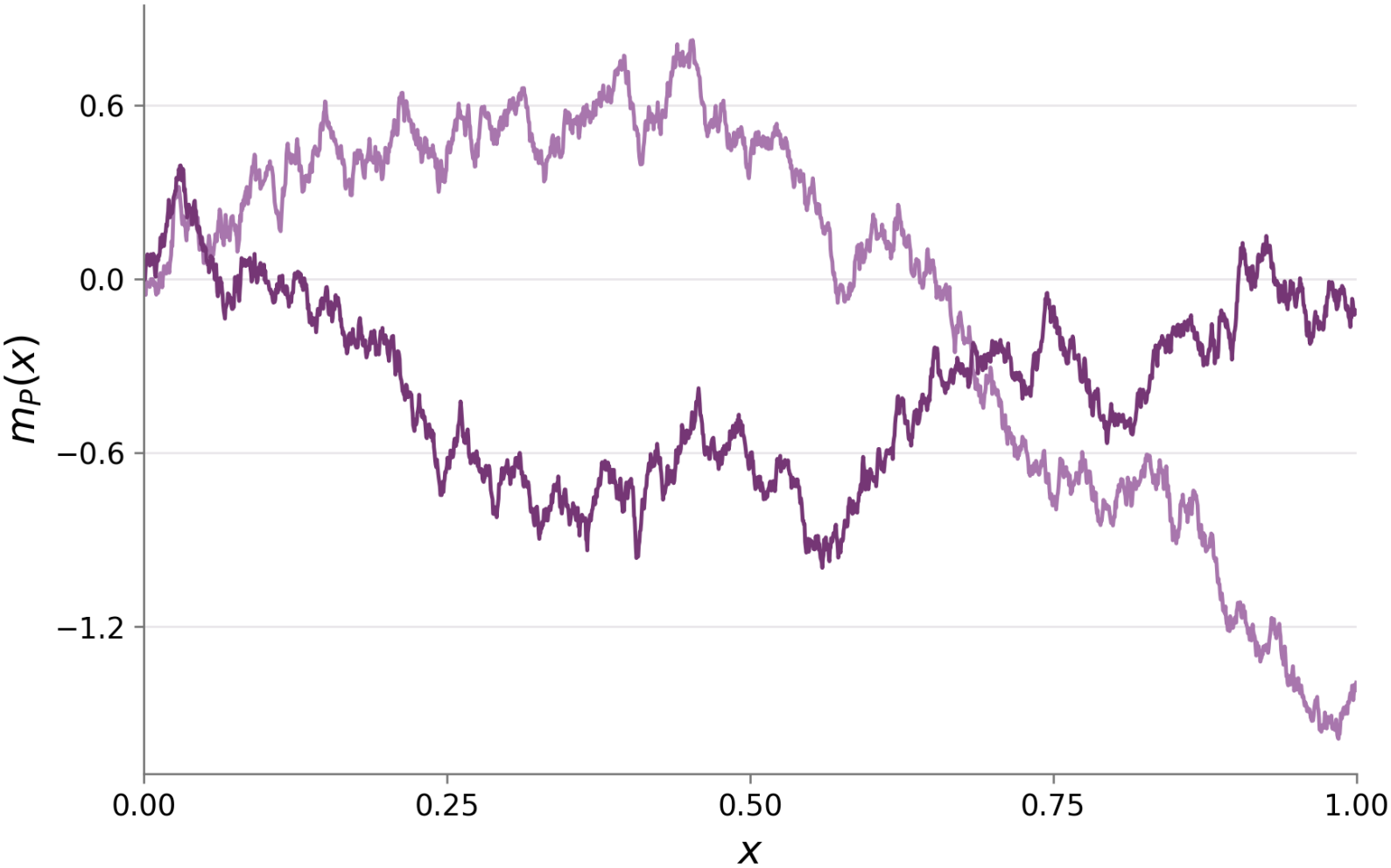
Optimal pairs for $c(x, y)$ take $O(n^3)$ time (Derigs, 1988).

Weaker condition is **stability:** no rematching of two pairs can reduce the total matching cost $\sum_{p=1}^{n/2} c(X_{i_p}, X_{j_p})$.

**Theorem.** Fix $d \geq 3$. Stable matched-pairs designs are minimax rate-optimal, adaptively over $\beta \in (0, 1]$,

$$\inf_\sigma \sup_{P \in \mathcal{P}_\beta} \mathcal{V}_n(\sigma, P) \asymp n^{-2\beta/d}.$$

# Rough Outcome Models, $\beta = 1/2$



# Beyond Matched Pairs

Consider the Lipschitz endpoint $\beta = 1$, with $\mathcal{P}_{\text{Lip}} = \mathcal{P}_1$. We have

$$\inf_{\sigma \in \mathcal{M}_n} \sup_{P \in \mathcal{P}_{\text{Lip}}} \mathcal{V}_n(\sigma, P) \asymp n^{-2/d}.$$

The much smaller linear class gives the same rate:

$$\inf_{\sigma \in \mathcal{M}_n} \sup_{P \in \mathcal{P}_{\text{Lin}}} \mathcal{V}_n(\sigma, P) \asymp n^{-2/d}.$$

Note that $\mathcal{P}_{\text{Lin}} \ll \mathcal{P}_{\text{Lip}}$. Matching **fails to exploit** any of the strong structure of linear models (smoothness, additivity).

This motivates new designs that can exploit smoothness and other potential low-dimensional structures of $m_P(X)$.

# Balancing Smooth Functions

# Spaces of Smooth Functions

To parameterize smoothness beyond Lipschitz order, we use the **Sobolev spaces** $H^s(\mathbb{R}^d)$, defined for $s > 0$.

For $k = 1, 2, \ldots$, a function $g$ belongs to $H^k(\mathbb{R}^d)$ if

$$\|g\|_{H^k}^2 \asymp \sum_{|\alpha| \leq k} \int_{\mathbb{R}^d} |D^\alpha g(x)|^2 \, dx < \infty.$$

Definition extends to real smoothness orders $s > 0$ using Fourier transforms. Writing $\widehat{g}(\omega) = (2\pi)^{-d/2} \int e^{-i\omega' x} g(x) \, dx$, define

$$\|g\|_{H^s}^2 \equiv \int |\widehat{g}(\omega)|^2 (1 + d^{-1} \|\omega\|_2^2)^s \, d\omega.$$

We introduce new experimental designs to exploit smoothness.

# The Matérn Kernel

Do this by **kernelizing** Gram-Schmidt walk (GSW) algorithm.

Define Matérn kernel $W(x,y) \propto \|x-y\|_2^{\nu} K_\nu(\|x-y\|_2)$, where $K_\nu$ is a modified Bessel function and $\nu > 0$ is a smoothness parameter.

The **RKHS** $\mathcal{H}_W$ is a Sobolev space: $\mathcal{H}_W = H^{d/2+\nu}(\mathbb{R}^d)$.

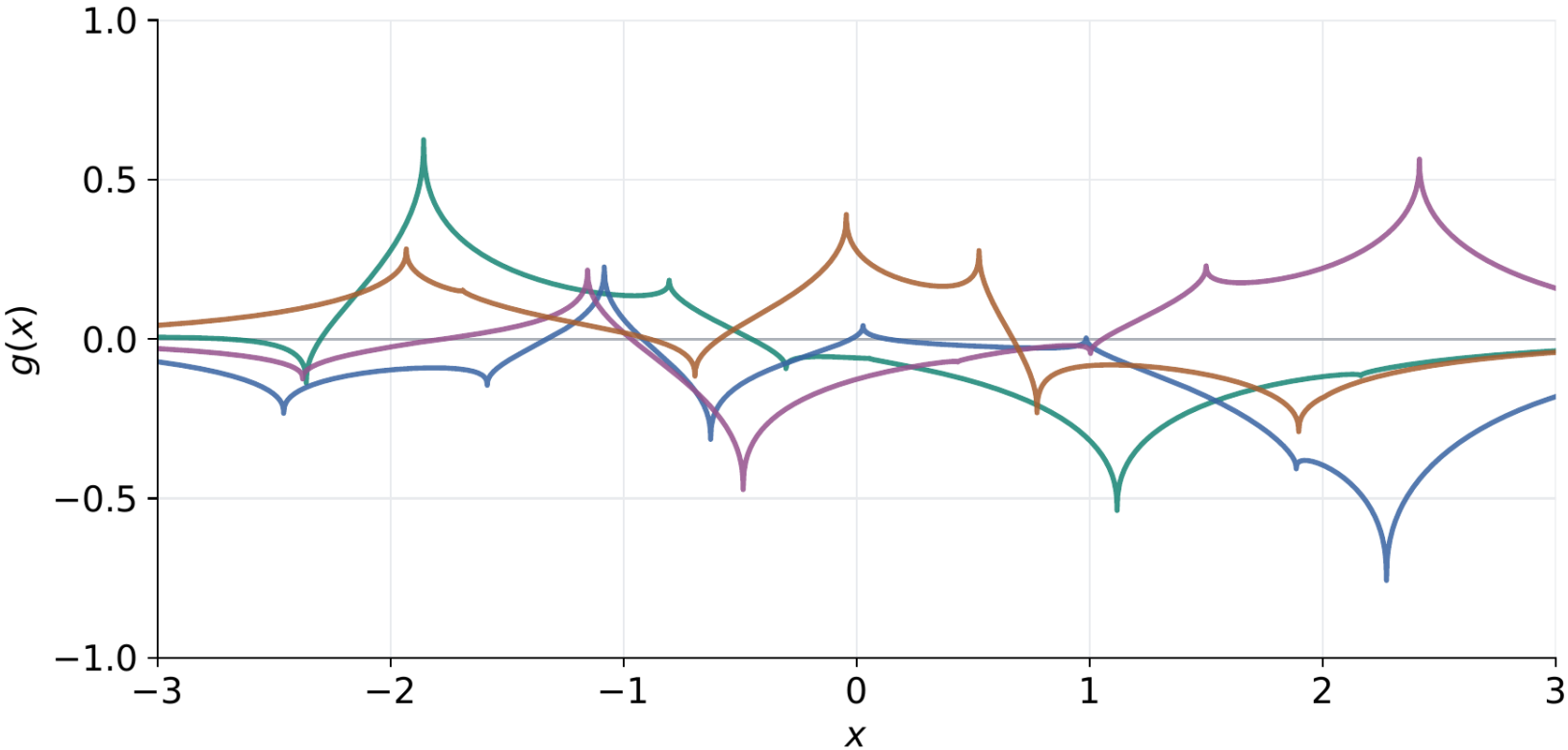
What if we try to uniformly balance all $g$ with $\|g\|_{\mathcal{H}_W} \leq 1$. Let $Z \in \{\pm 1\}^n$ and define kernel matrix $W_n = [W(X_i, X_j)]_{i,j=1}^n$. Then

$$\sup_{\|g\|_{\mathcal{H}_W} \leq 1} \left( \frac{1}{n} \sum_{i=1}^n Z_i g(X_i) \right)^2 = \frac{1}{n^2} Z' W_n Z.$$

Design should randomize $Z \in \{\pm 1\}^n$ so $Z' W_n Z$ is small.

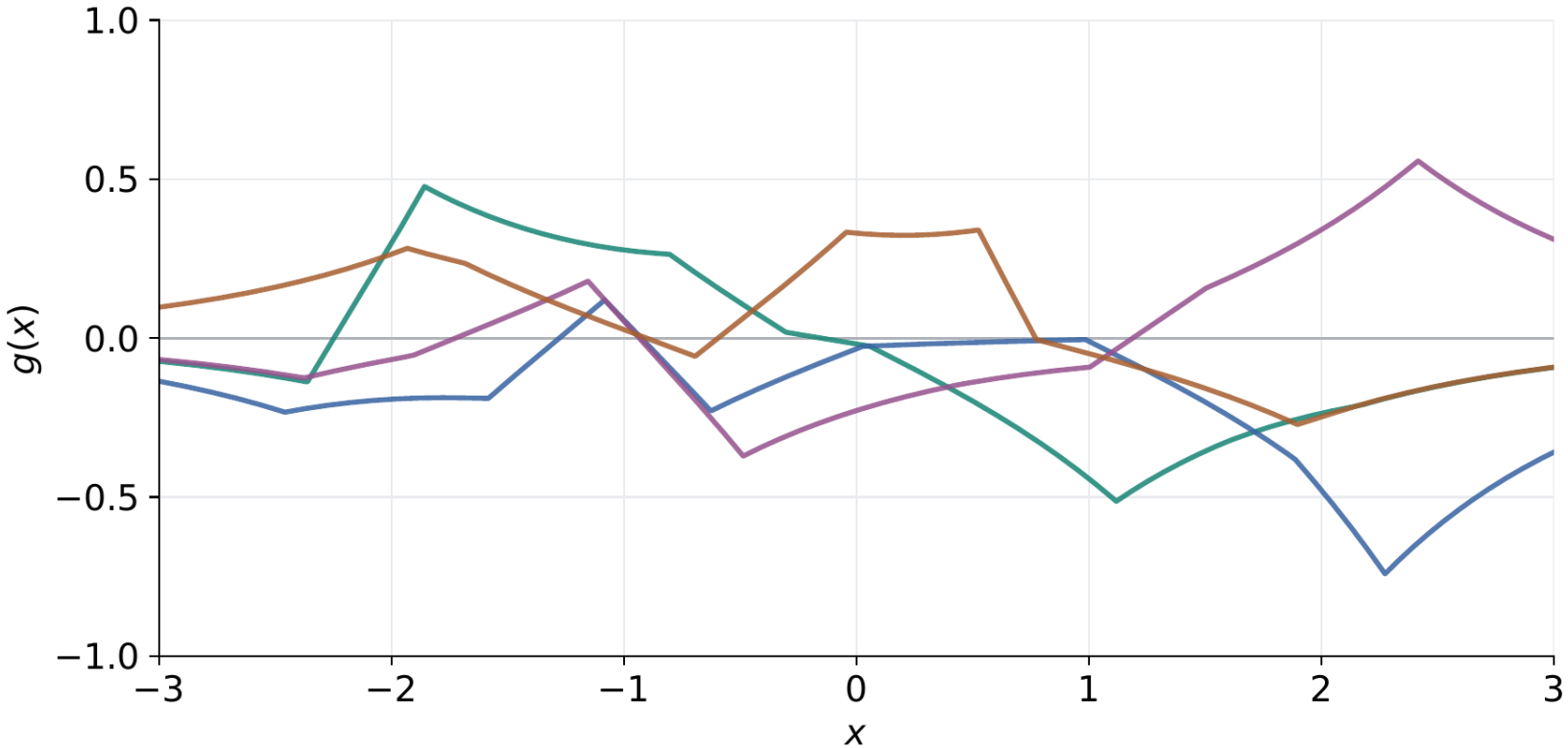
# Matérn RKHS

Functions from the RKHS unit ball $\|g\|_{\mathcal{H}_W} \leq 1$ with $\nu = 0.2$.



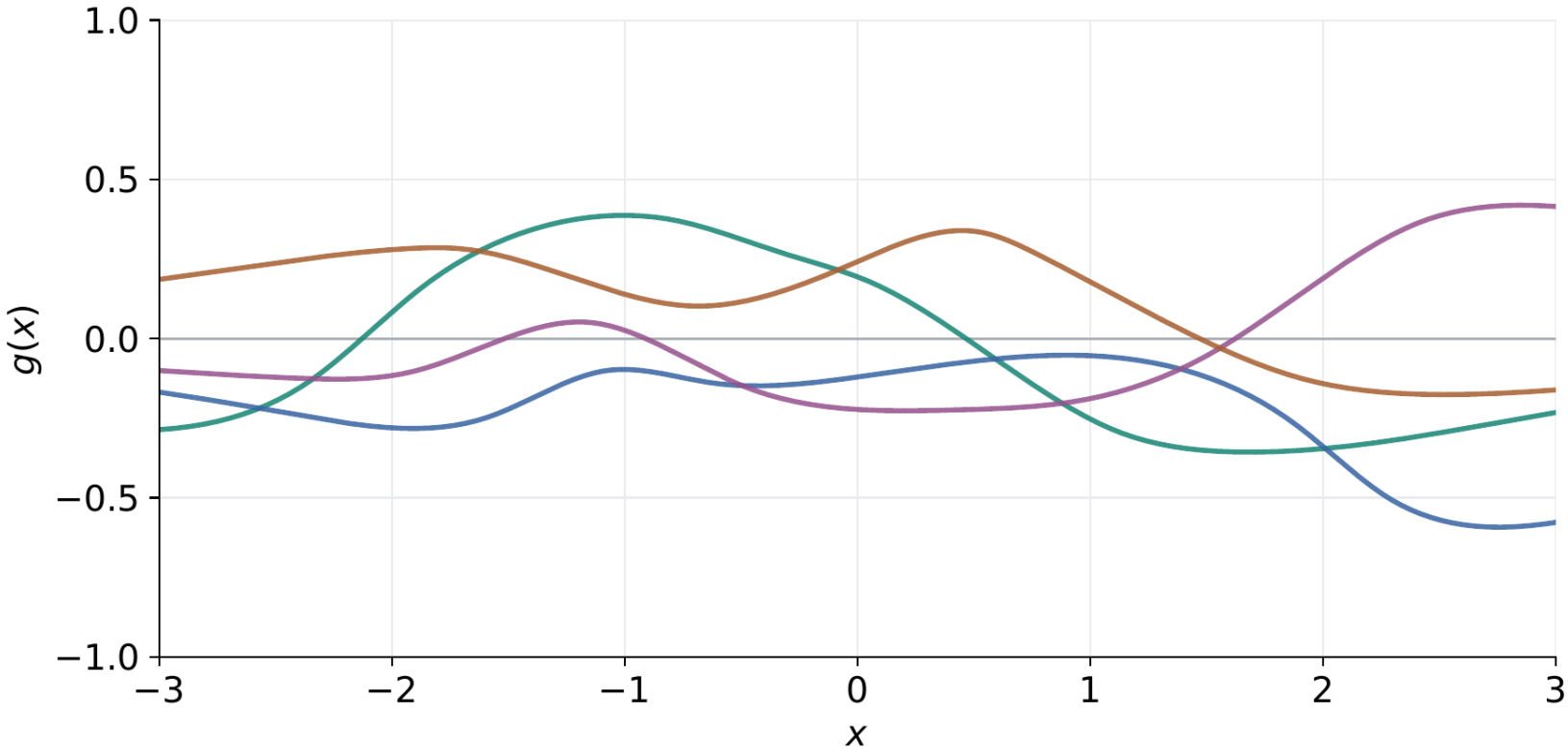
# Matérn RKHS

Functions from the RKHS unit ball $\|g\|_{\mathcal{H}_W} \leq 1$ with $\nu = 0.5$.



# Matérn RKHS

Functions from the RKHS unit ball $\|g\|_{\mathcal{H}_W} \leq 1$ with $\nu = 1.5$.



# The Gram-Schmidt Walk

Randomize to control: $Z' W_n Z \quad Z \in \{\pm 1\}^n$.

Use Gram-Schmidt walk (GSW) of Bansal et al. (2019). First applied to experiments by Harshaw et al. (2024), Harshaw (2021).

GSW is an iterative rounding scheme. Let $Z^{(0)} = 0$. Each step $Z^{(t)} \to Z^{(t+1)}$ randomly fixes at least one assignment at $\pm 1$.

Randomly draw pivot index $p \in [n]$ among units with fractional assignments $Z_i^{(t)} \notin \{\pm 1\}$ at $t$. Update $Z^{(t+1)} = Z^{(t)} + \delta u$:

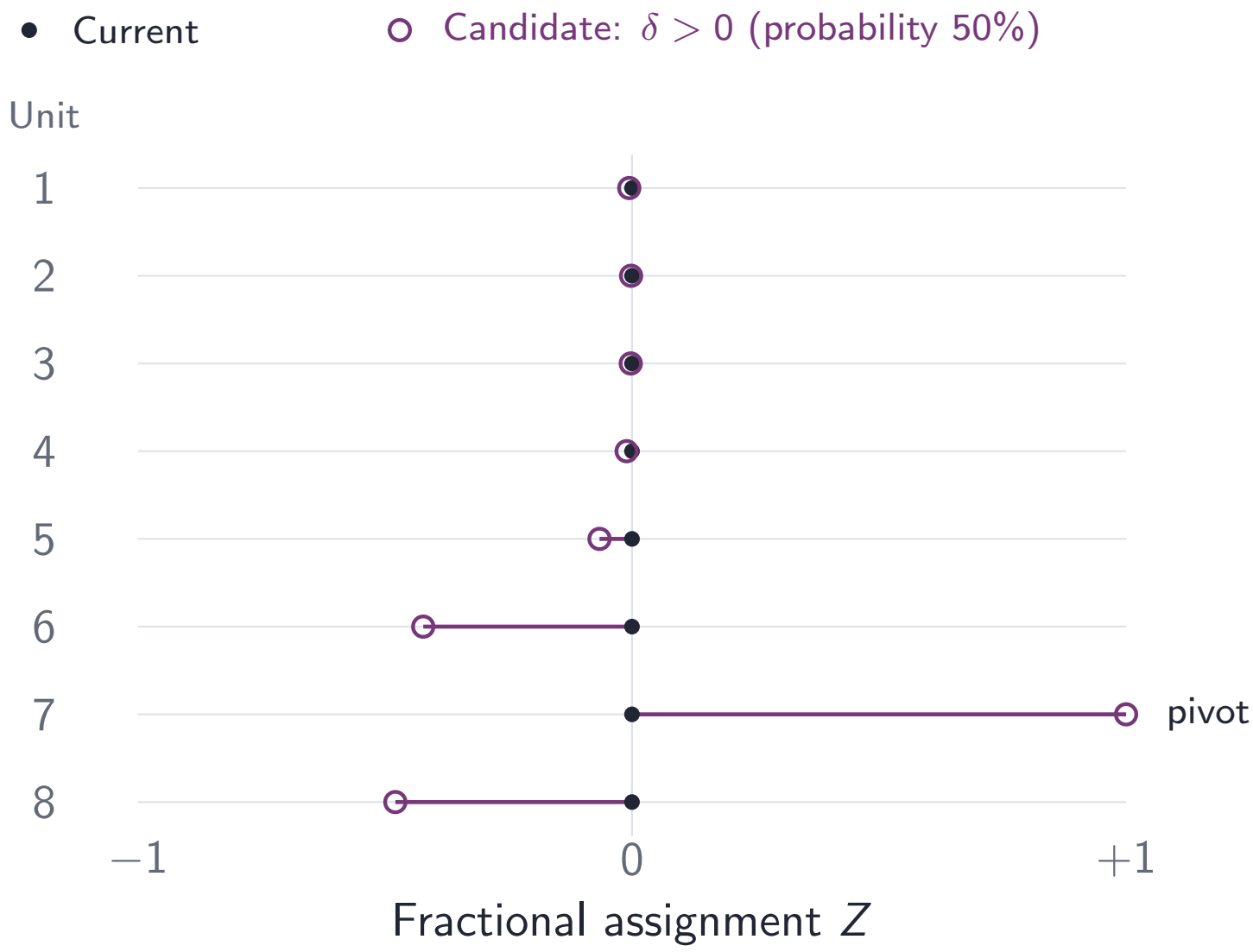
$$\min_{u \in \mathbb{R}^n} u' W_n u \quad \text{subject to} \quad u_p = 1, \quad u_i = 0 \ \text{ if } \ |Z_i^{(t)}| = 1.$$

Increase or decrease $\delta$ from zero until another assignment hits $\pm 1$. Randomize $\delta$ from those two values with conditional mean zero.

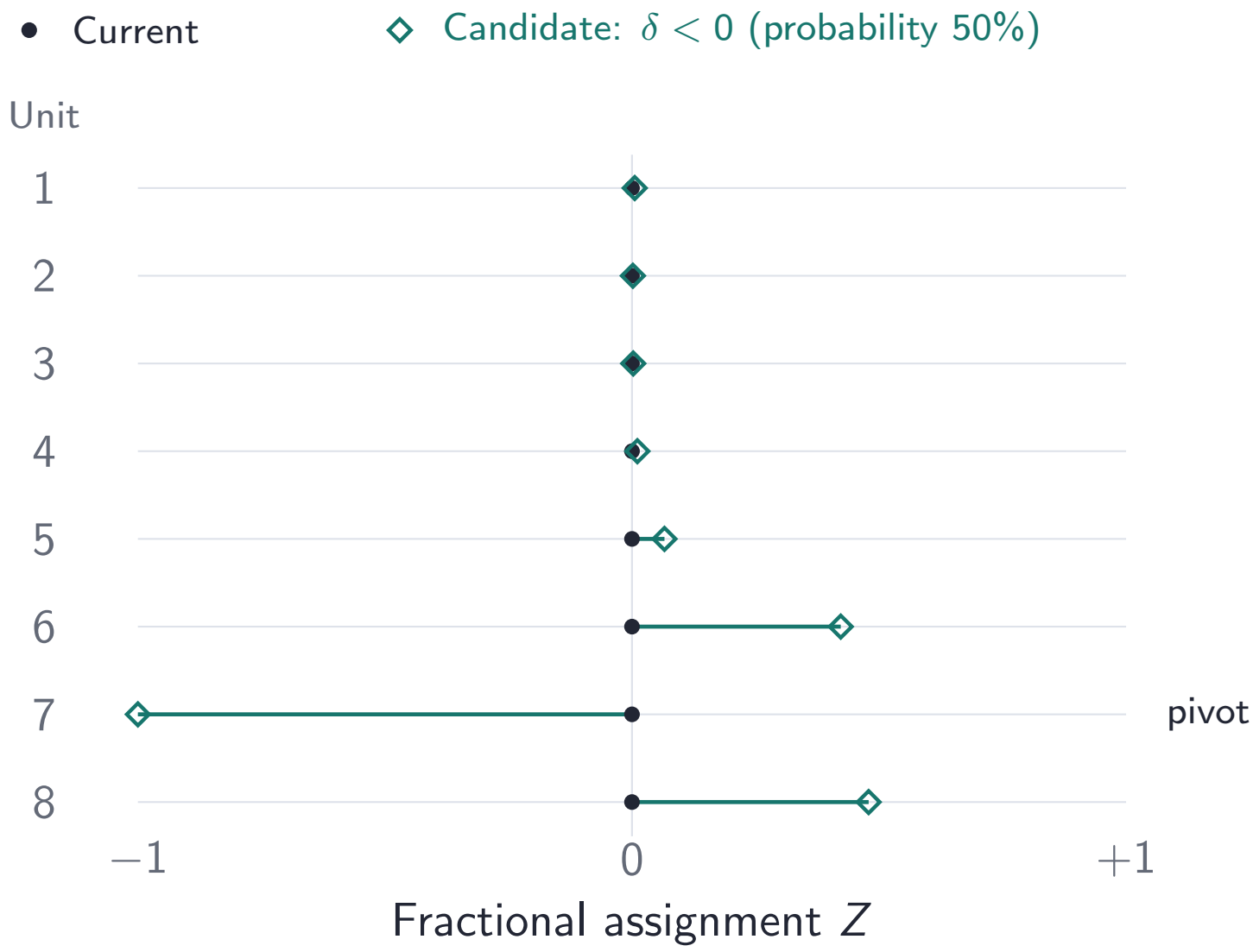
# Two steps of the Gram-Schmidt walk



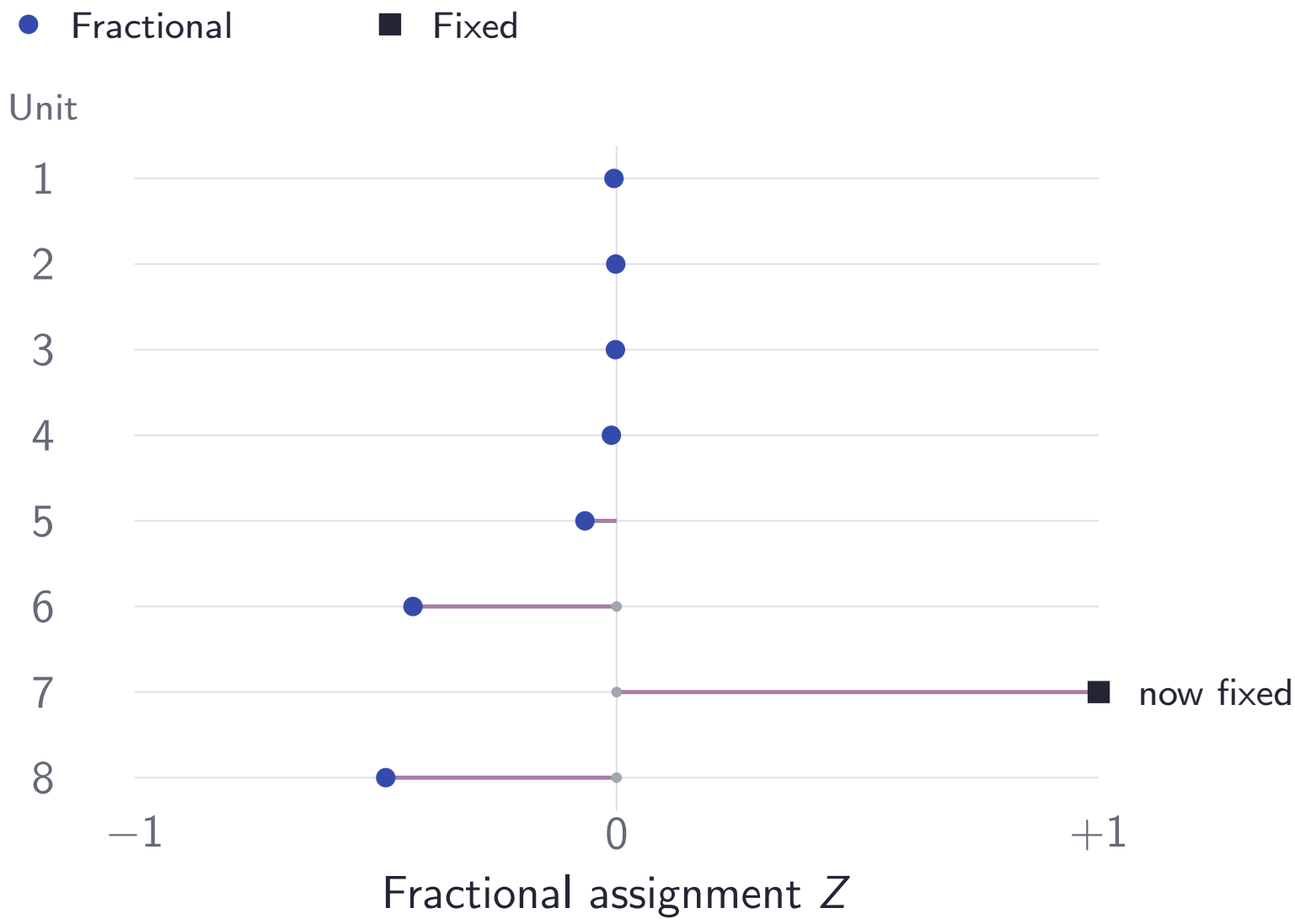
# Two steps of the Gram-Schmidt walk



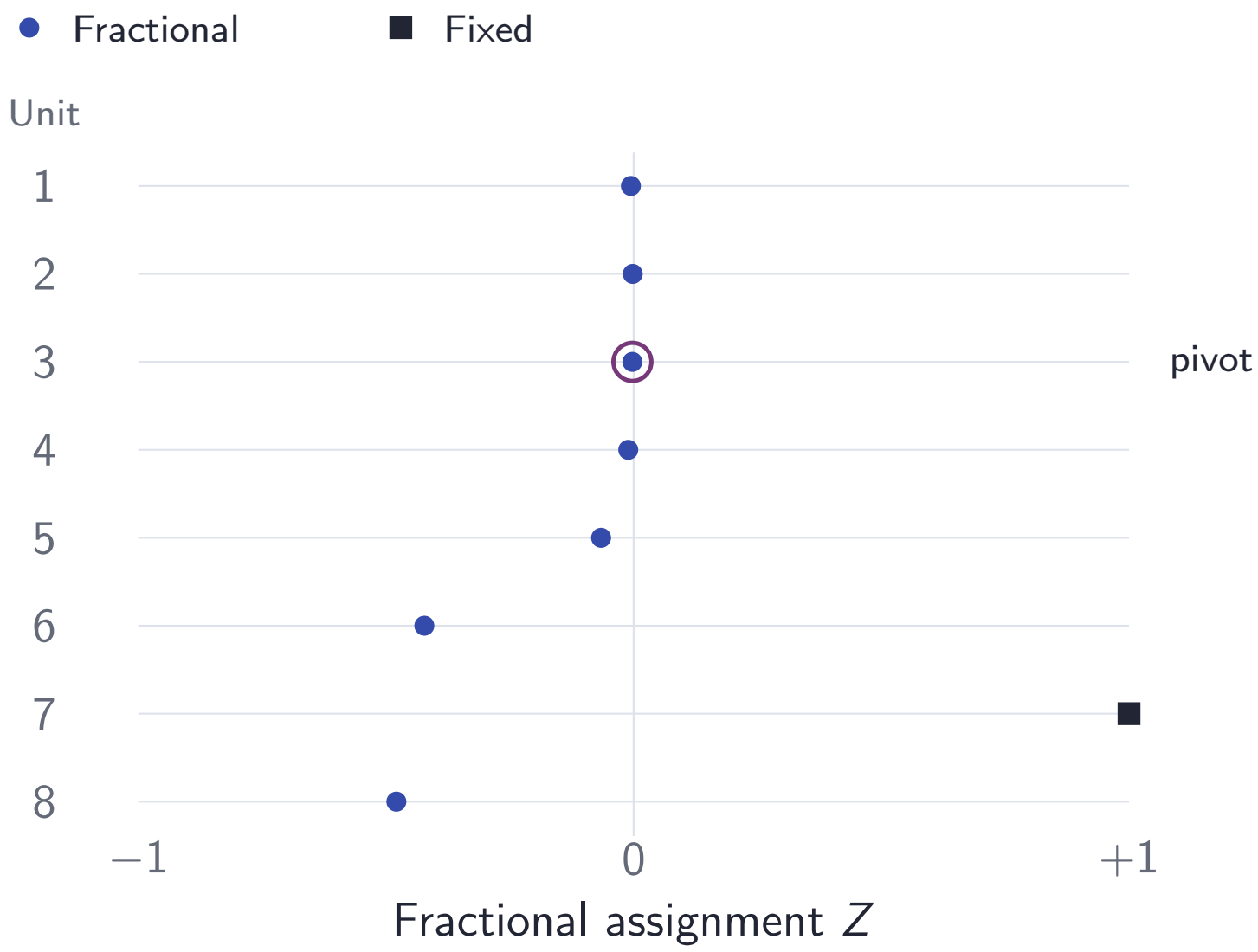
# Two steps of the Gram-Schmidt walk



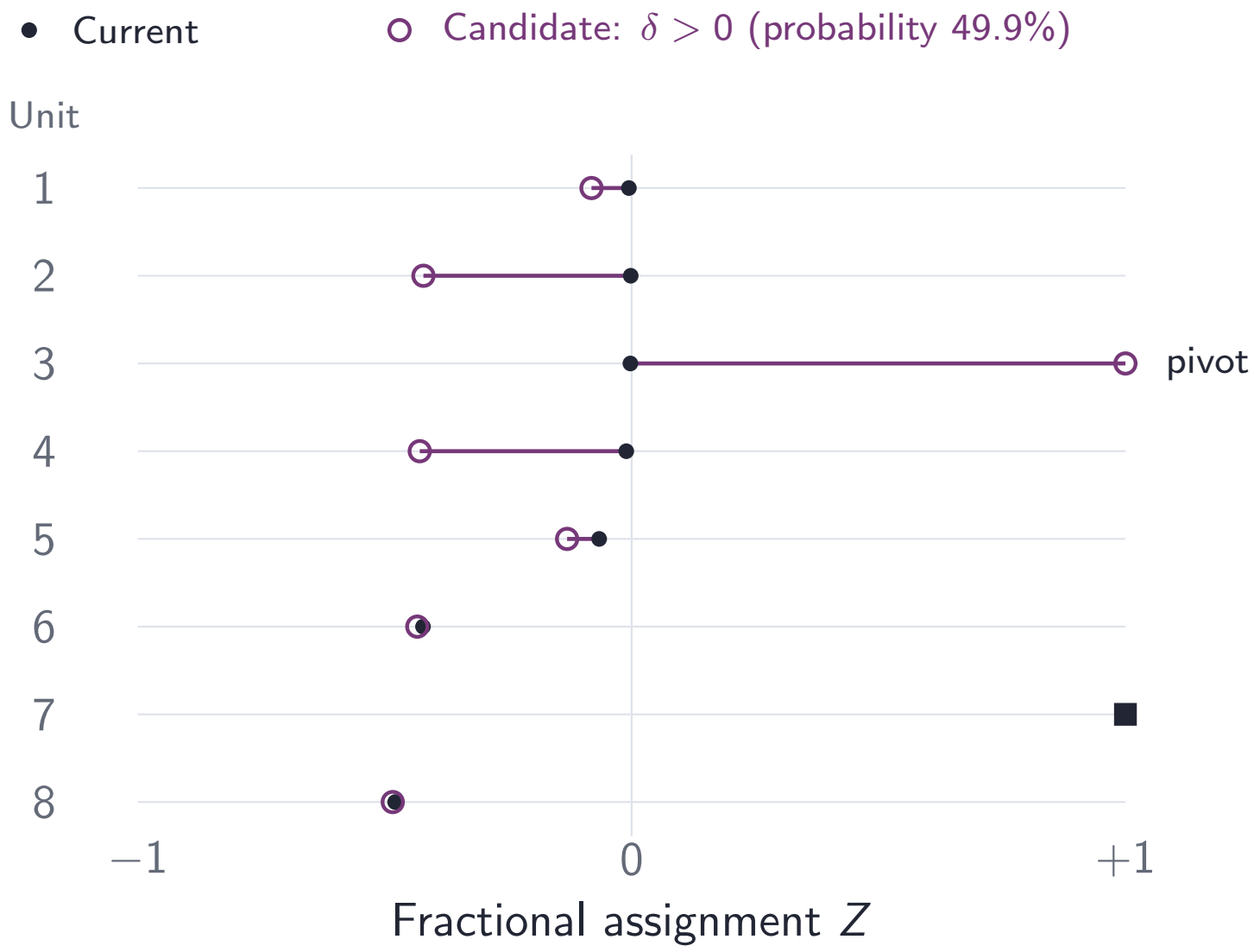
# Two steps of the Gram-Schmidt walk



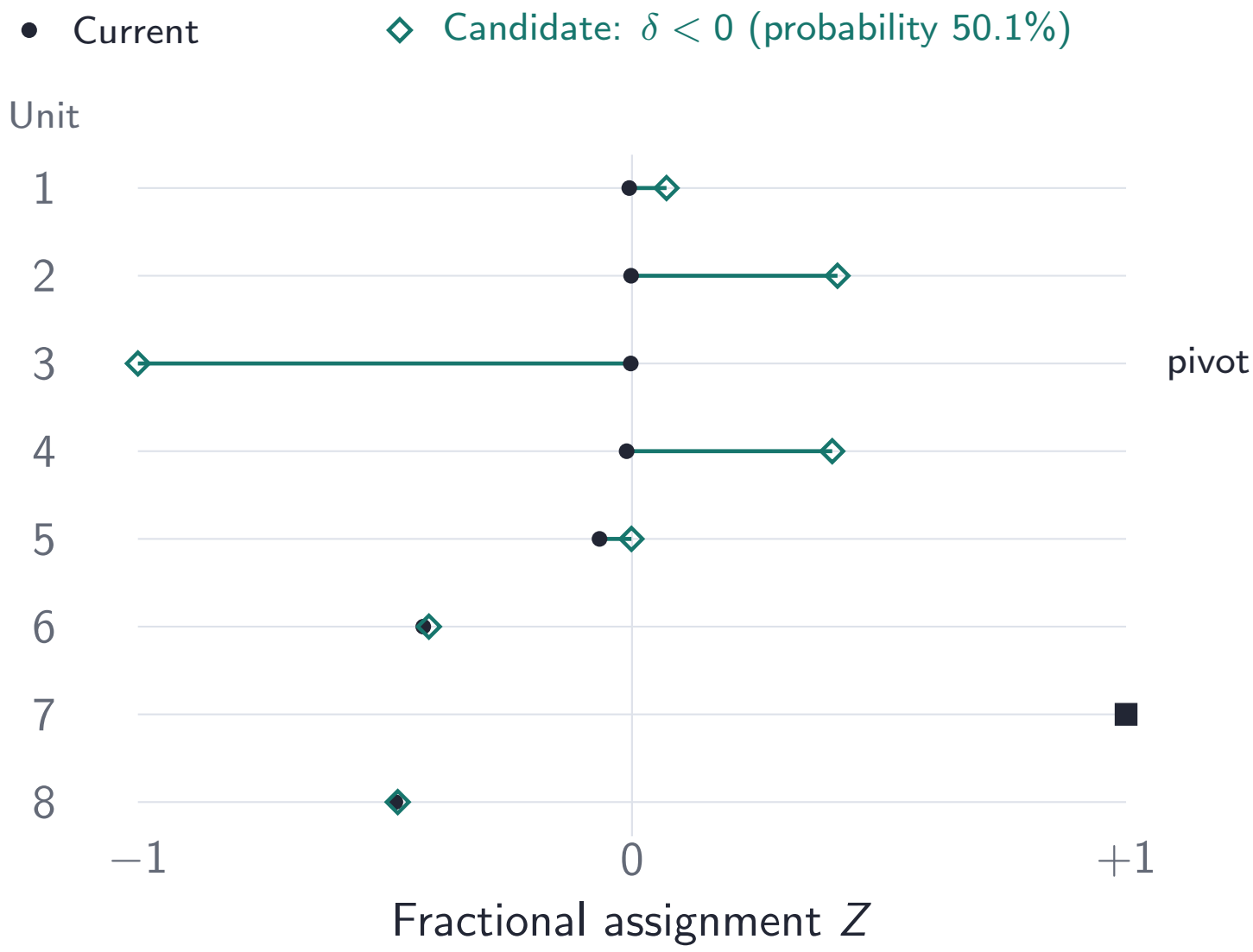
# Two steps of the Gram-Schmidt walk



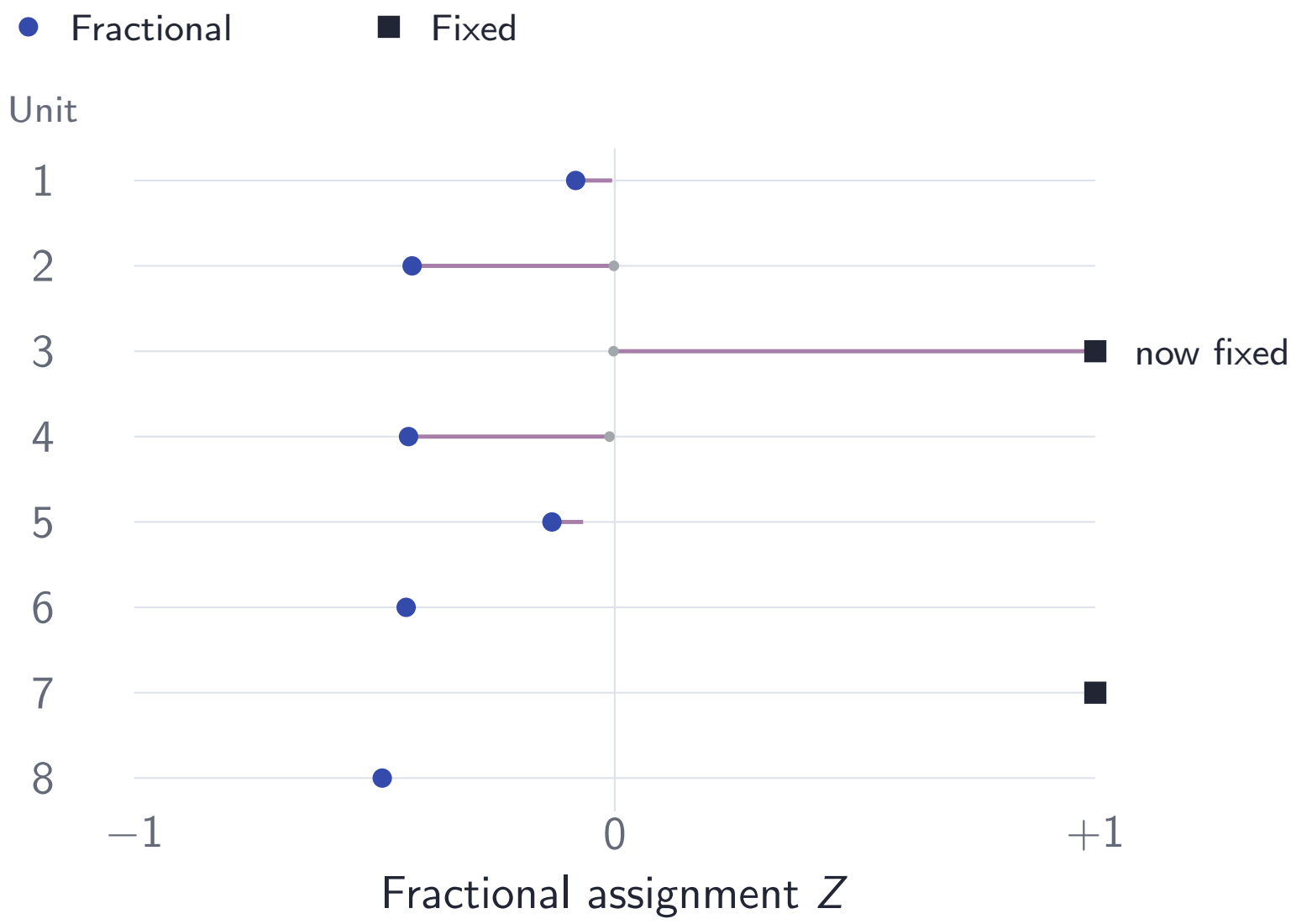
# Two steps of the Gram-Schmidt walk



# Two steps of the Gram-Schmidt walk



# Two steps of the Gram-Schmidt walk



# Exploiting Smoothness with GSW

# Smooth GSW

Want to use kernelized GSW to balance functions in $H^s(\mathbb{R}^d)$.

$$\mathcal{P}_d^s = \{P : \|m_P\|_{H^s(\mathbb{R}^d)} \leq 1\}.$$

Kernelized GSW for the Matérn RKHS with parameter $\nu$ controls functions in Sobolev space $H^{d/2+\nu}$. Set $\nu_n = 1/\log n$.

Run GSW on matrix $\varphi I_n + (1-\varphi)(W_n + 11')/2$ for $\varphi \in (0,1)$.

The identity term adds robustness, encouraging additional randomization to protect against unmodeled imbalances.

# Exploiting Smoothness

**Theorem.** (Adaptivity) Let $\sigma_W$ be as above. For $n \geq 2$ and $d \geq 1$:

- *Rough regime.* If $0 < 2s \leq d$, then

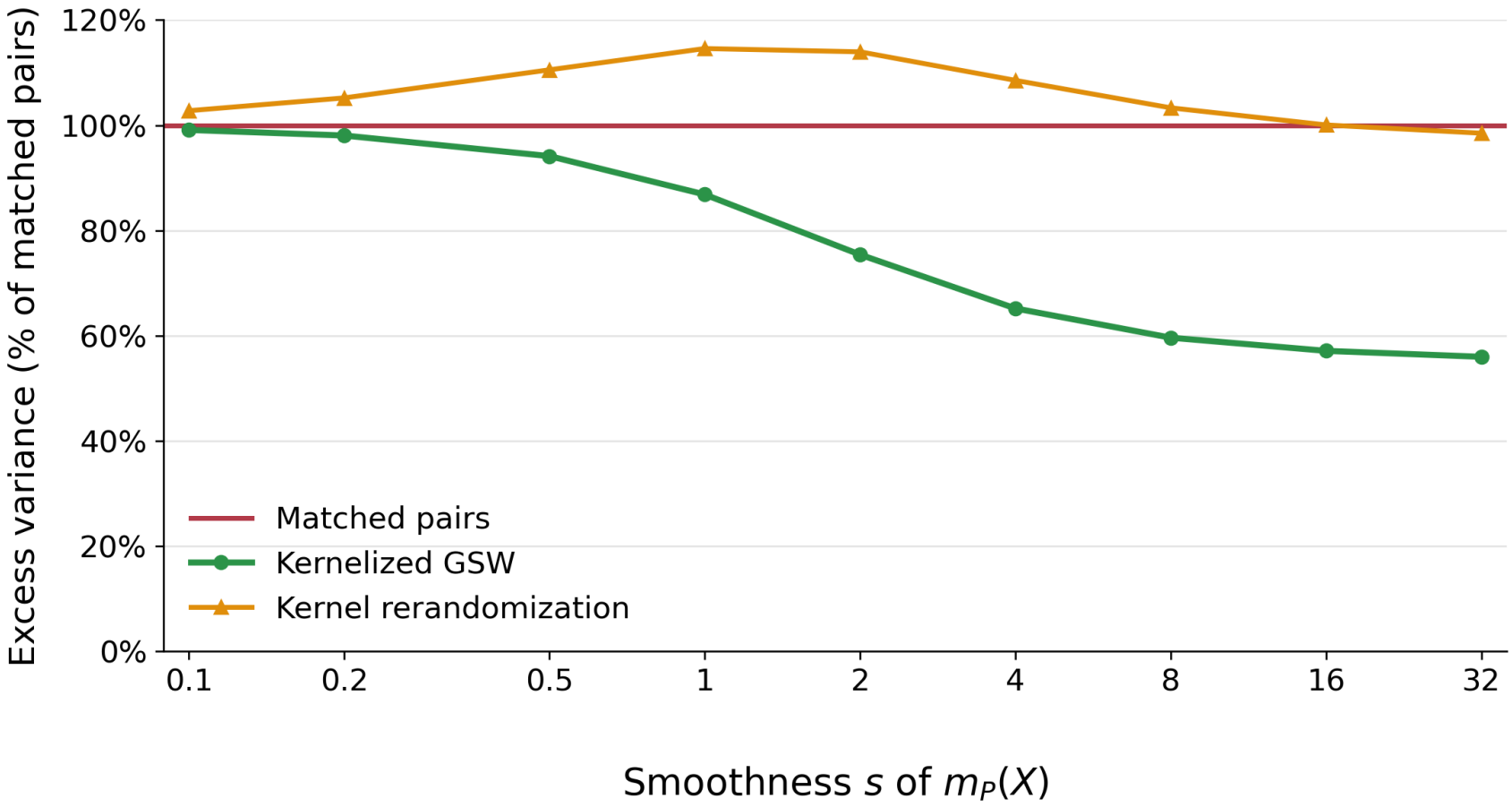
$$\sup_{P \in \mathcal{P}_d^s} \mathcal{V}_n(\sigma_W, P) \lesssim n^{-2s/d}.$$

- *Smooth regime.* If $2s > d$, then

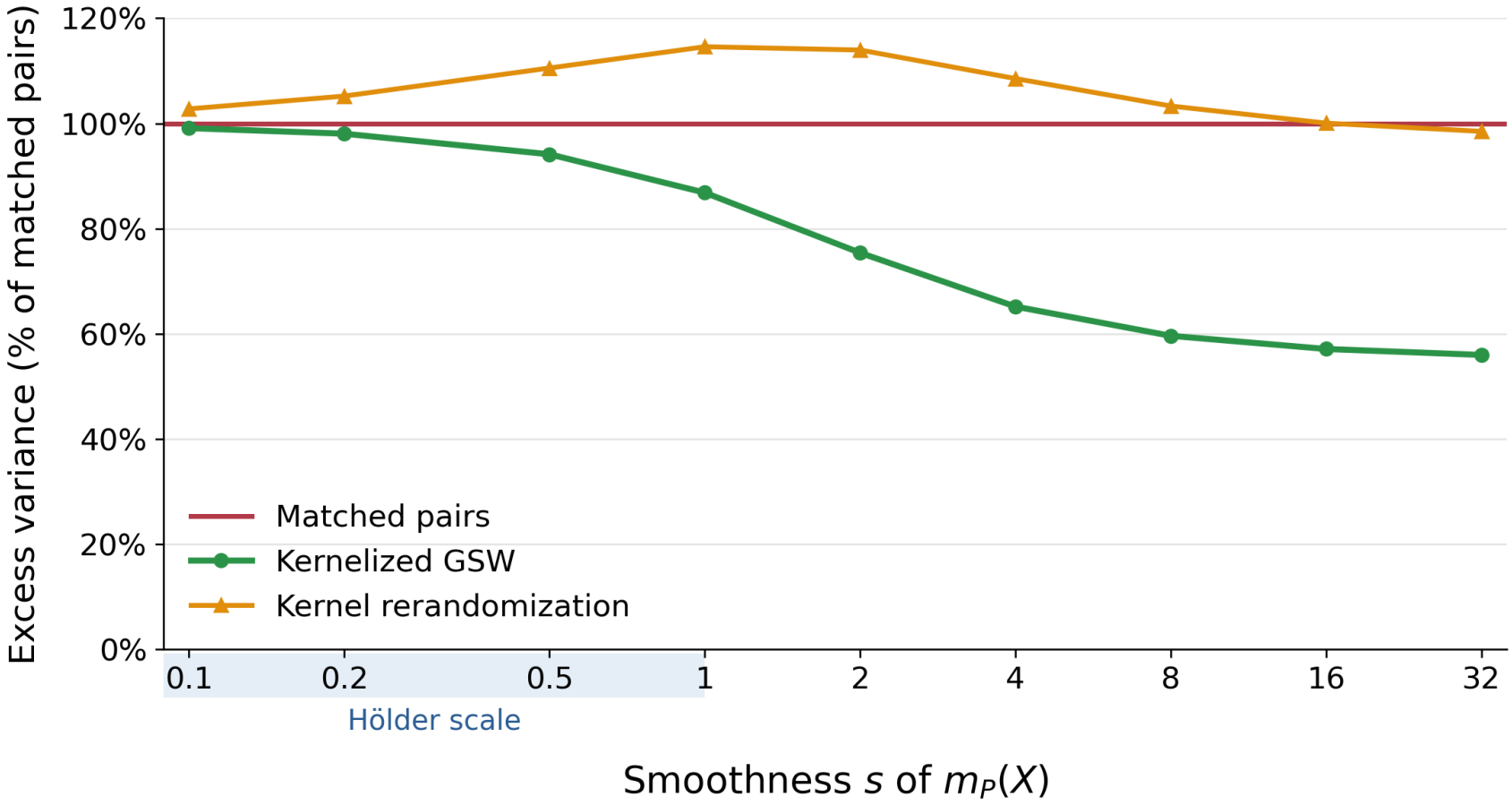
$$\sup_{P \in \mathcal{P}_d^s} \mathcal{V}_n(\sigma_W, P) \lesssim n^{-1}.$$

The kernelized design **adaptively** exploits smoothness without prior knowledge of $s > 0$, up to the parametric rate $n^{-1}$.

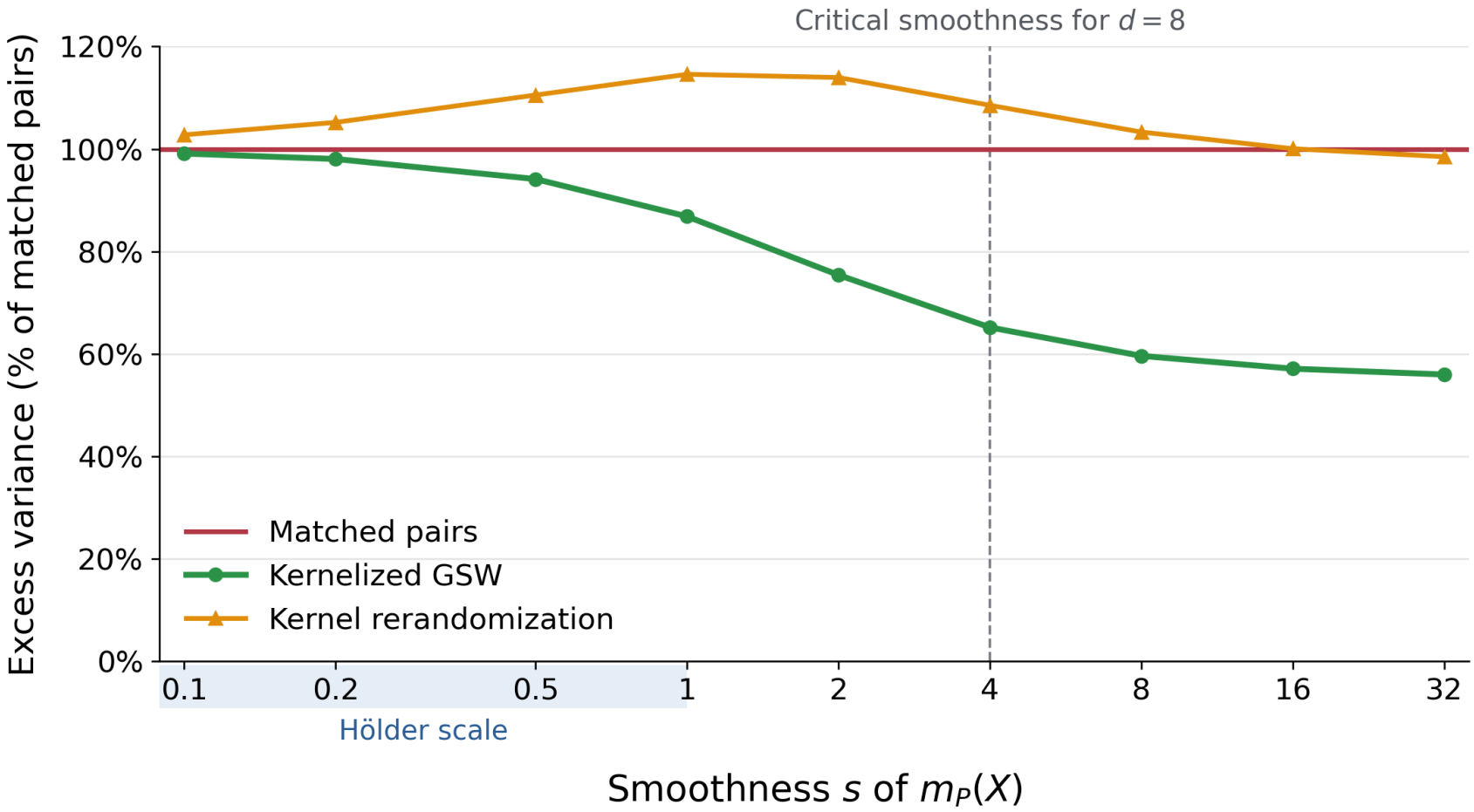
# Simulation: Exploiting Smoothness



# Simulation: Exploiting Smoothness



# Simulation: Exploiting Smoothness



# Curse of Dimensionality

Is kernelized GSW efficient beyond $d = o(\log n)$? No.

**Theorem.** (Impossibility) Fix $s > 0$. For every $n \geq 2$ and $d \geq 1$,

$$\inf_{\sigma} \sup_{P \in \mathcal{P}_d^s} \mathcal{V}_n(\sigma, P) \gtrsim n^{-2s/d}.$$

As before, for fixed $s > 0$, $n^{-2s/d} \to 0$ only if $d = o(\log n)$.

**Asymptotic Inefficiency.** If $d \geq c \log n$ for $c > 0$, then

$$\liminf_{n \to \infty} \inf_{\sigma} \sup_{P \in \mathcal{P}_d^s} \mathcal{V}_n(\sigma, P) > 0.$$

No design can guarantee asymptotic efficiency, even over smooth outcome models, beyond very low dimension.

# Efficient Design Beyond Low Dimension

# Beyond Semiparametric Efficiency

To guarantee asymptotic efficiency, we need to uniformly balance full-dimensional smooth functions $g \in H^s(\mathbb{R}^d)$.

For fixed $s > 0$, this is **impossible** unless $d = o(\log n)$.

Suggests we should modify our efficiency target. For example, balance functions in a restricted-complexity class $\mathcal{G}$.

**Example.** (Main Effects) Let $\mathcal{G}$ contain $g(X) = \sum_{j=1}^{d} g_j(X_j)$.

Let $m_{\mathcal{G}}$ be the $L^2(P_X)$ projection of $m_P$ onto $\mathcal{G}$. The **restricted efficiency target** is $V^*(P) + \Delta_{\mathcal{G}}(P)$, where

$$\Delta_{\mathcal{G}}(P) = 4E_P[(m_P(X) - m_{\mathcal{G}}(X))^2].$$

# Non-Parametric Additive GSW

Let $W$ be a univariate Matérn kernel with smoothness parameter $\nu_n = 1/\log n$. Run GSW with $K(x, x') = 1 + \frac{1}{d}\sum_{j=1}^{d} W(x_j, x_j')$.

Write $m_{\mathcal{G}}(X) = a + \sum_{j=1}^{d} m_j(X_j)$. Impose $\sum_{j=1}^{d} \|m_j\|^2_{H^s(\mathbb{R})} \leq 1$.

**Theorem.** (Additive GSW) For suitable $\varphi_n$, design $\sigma_K$ has

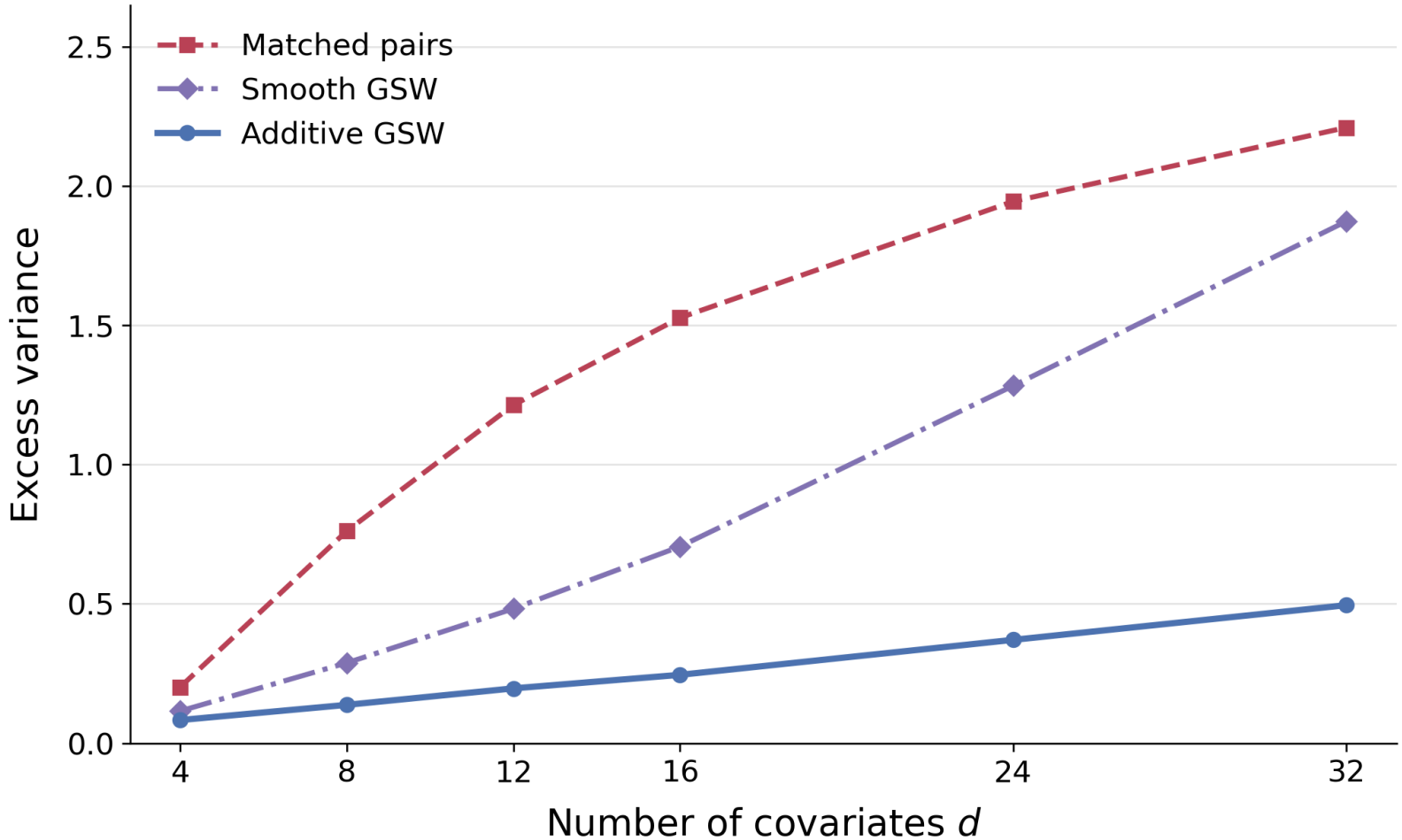
$$\mathcal{V}_n(\sigma_K, P) \leq \Delta_{\mathcal{G}}(P) + C\left(\frac{d\log n}{n}\right)^{\min(s,1/2)}.$$

Restricted efficiency is possible for $d = o(n)$ up to logarithms.

**Well-spec.** Let $\mathcal{P}^s_{\text{add}}$ have $\Delta_{\mathcal{G}}(P) = 0$ and $\sum_{j=1}^{d} \|m_j\|^2_{H^s(\mathbb{R})} \leq 1$.

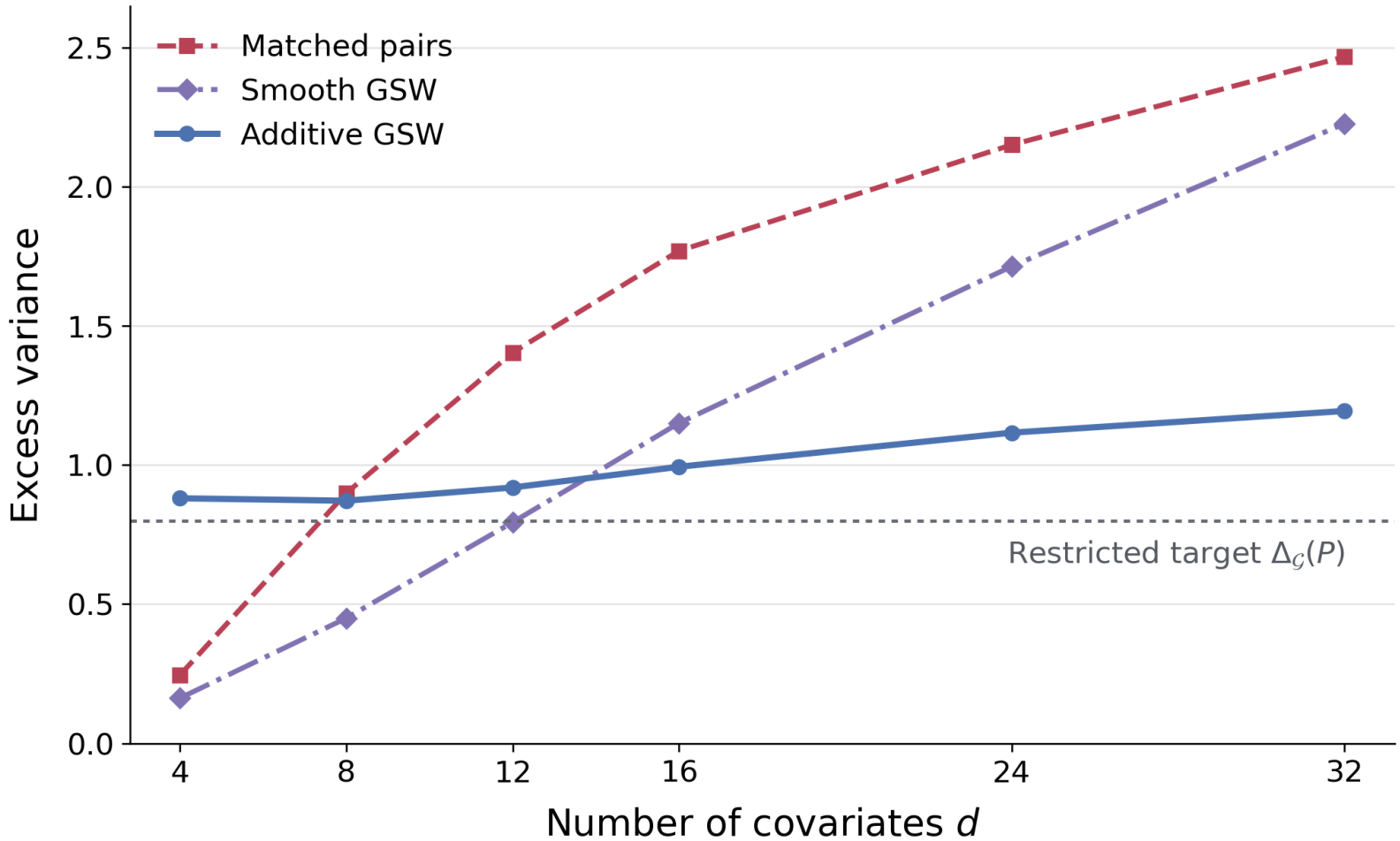
$$\inf_{\sigma} \sup_{P \in \mathcal{P}^s_{\text{add}}} \mathcal{V}_n(\sigma, P) \to 0 \quad \text{if} \quad d = o(n/\log n).$$

# Simulation: Well Specification




$n = 240$. Well Specification.

# Simulation: Omitted Interaction




$n = 240$. Omitted Interaction.

# Robustifying Additive GSW

When $m_P(X)$ has substantial covariate interactions, additive GSW can do worse than matched pairs in low dimensions.

Can combine GSW with matched pairs to robustify the design.

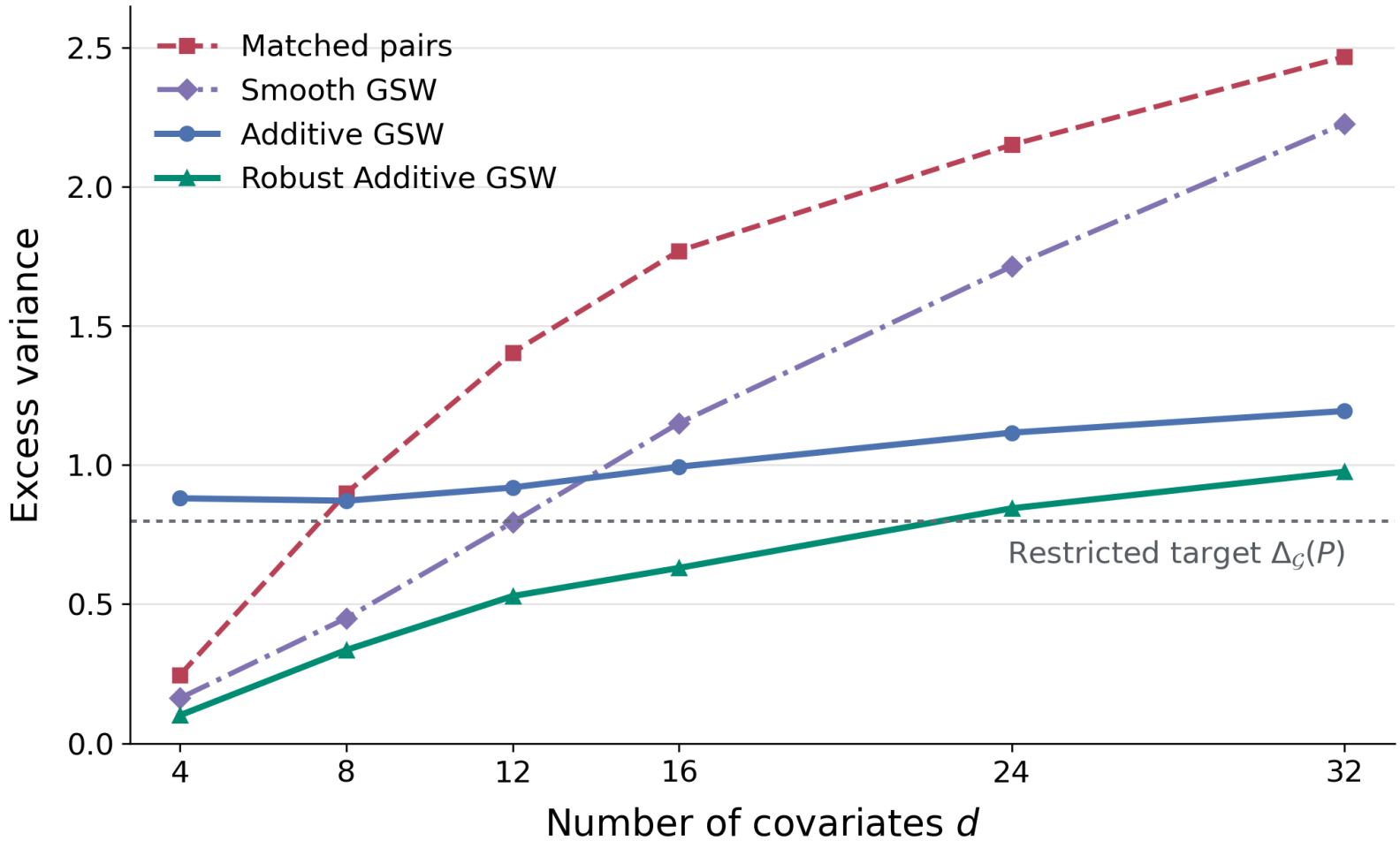
Matched pairs flips iid coins $T_p \in \{\pm 1\}$ for each pair $(i_p, j_p)$.

Treat $i_p$ if $T_p = 1$ and $j_p$ otherwise. Then imbalance in $g$ is

$$\frac{1}{n}\sum_{i=1}^{n} Z_i g(X_i) = \frac{1}{n}\sum_{p=1}^{n/2} T_p\big(g(X_{i_p}) - g(X_{j_p})\big).$$

**Matched GSW.** Draw $(T_p)_{p=1}^{n/2} \in \{\pm 1\}^{n/2}$ using kernelized GSW.

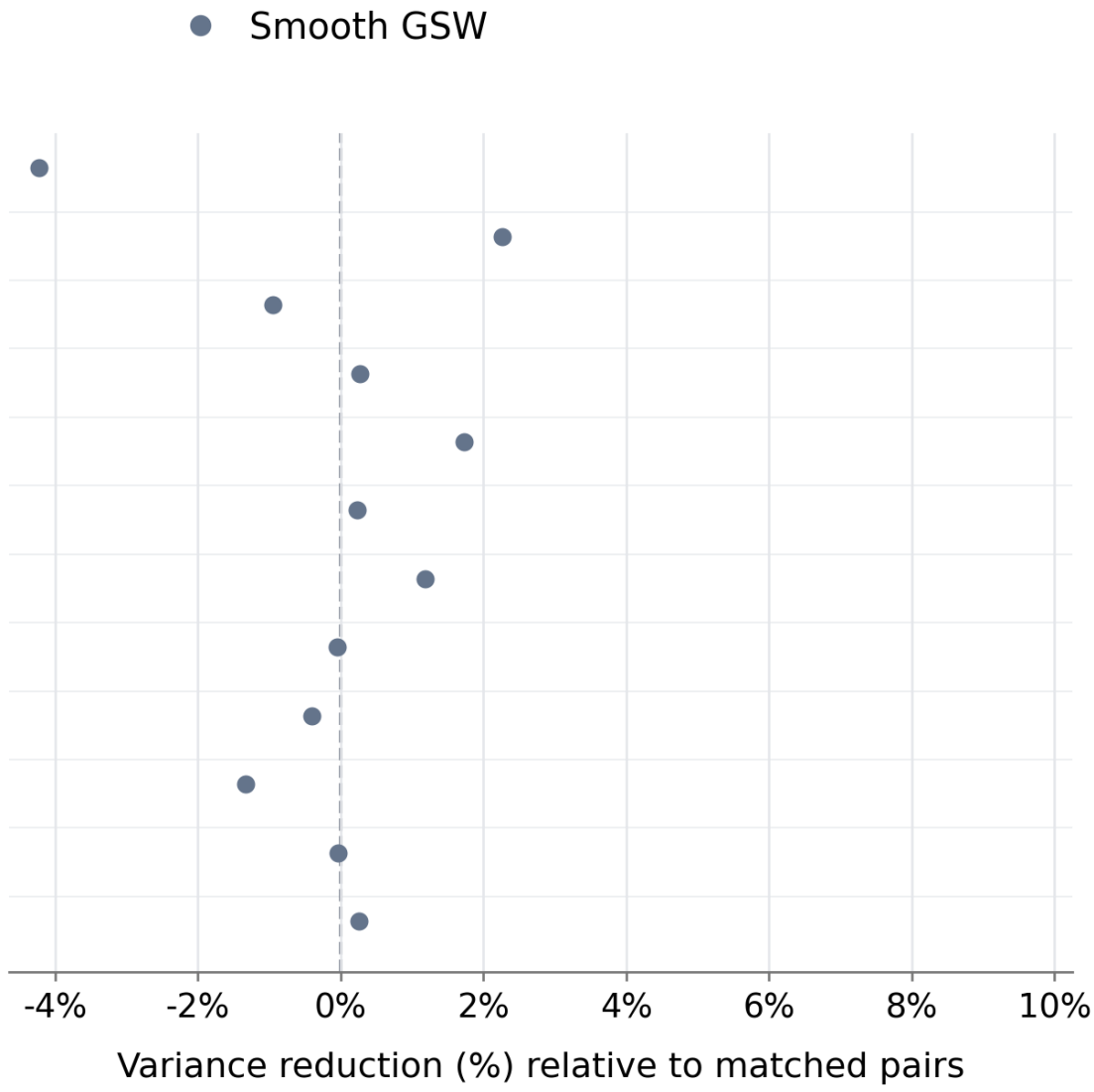
# Simulation: Omitted Interaction




$n = 240$. Omitted Interaction.

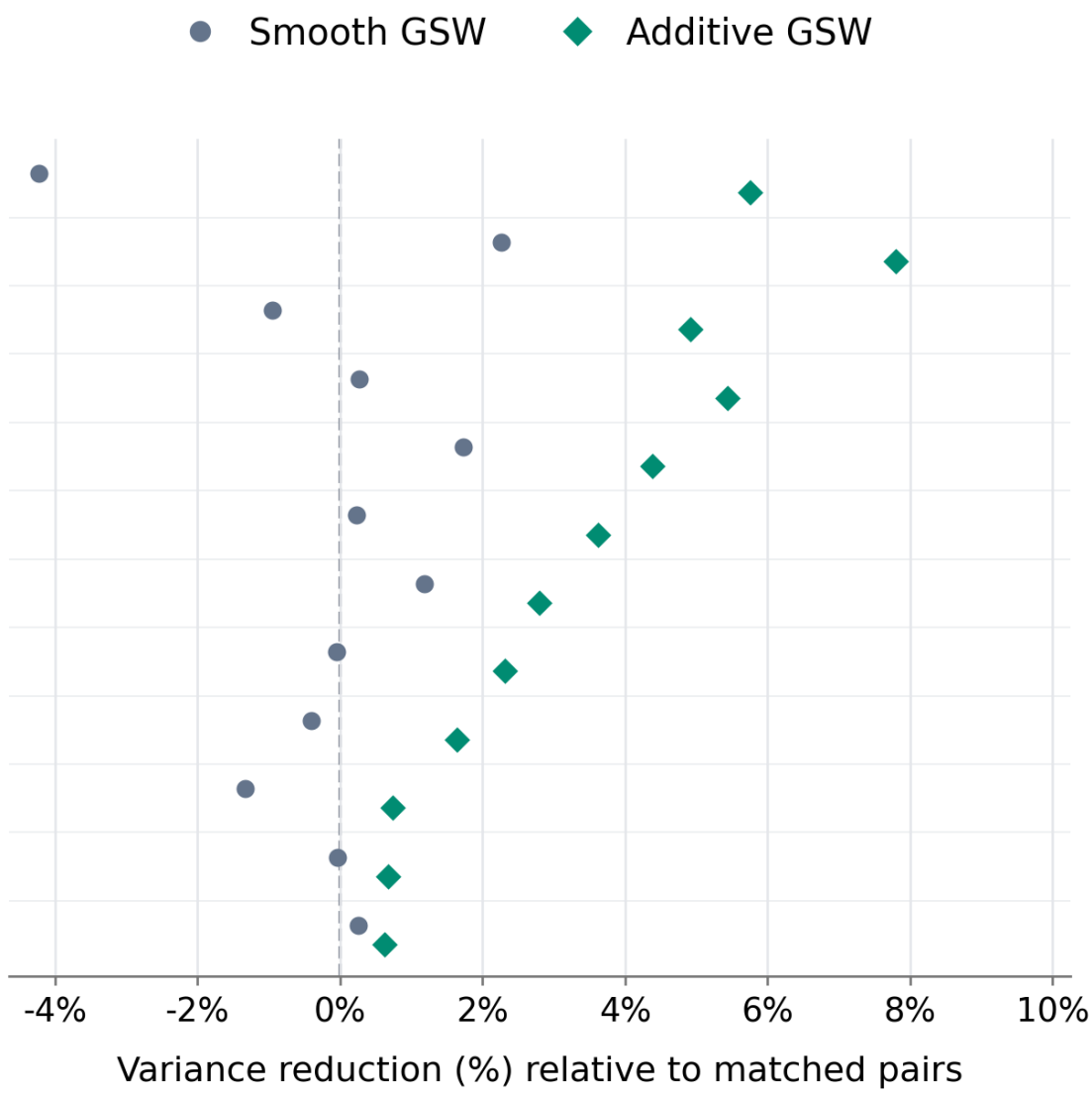
# Empirical Application: Unit-Level GSW

| Experiment | $n$ | $d$ |
|---|---|---|
| Augsburg ’22 | 124 | 21 |
| BenYishay ’20 | 96 | 22 |
| Augenblick ’26 | 828 | 11 |
| Bhandari ’23 | 444 | 31 |
| Blattman ’23 | 832 | 51 |
| Evsyukova ’25 | 364 | 12 |
| Haushofer ’16 | 940 | 14 |
| Greenstone ’25 | 292 | 15 |
| Beam ’16 | 96 | 16 |
| Gautam ’25 | 108 | 15 |
| Bandiera ’21 | 548 | 7 |
| Gupta ’26 | 1,200 | 14 |



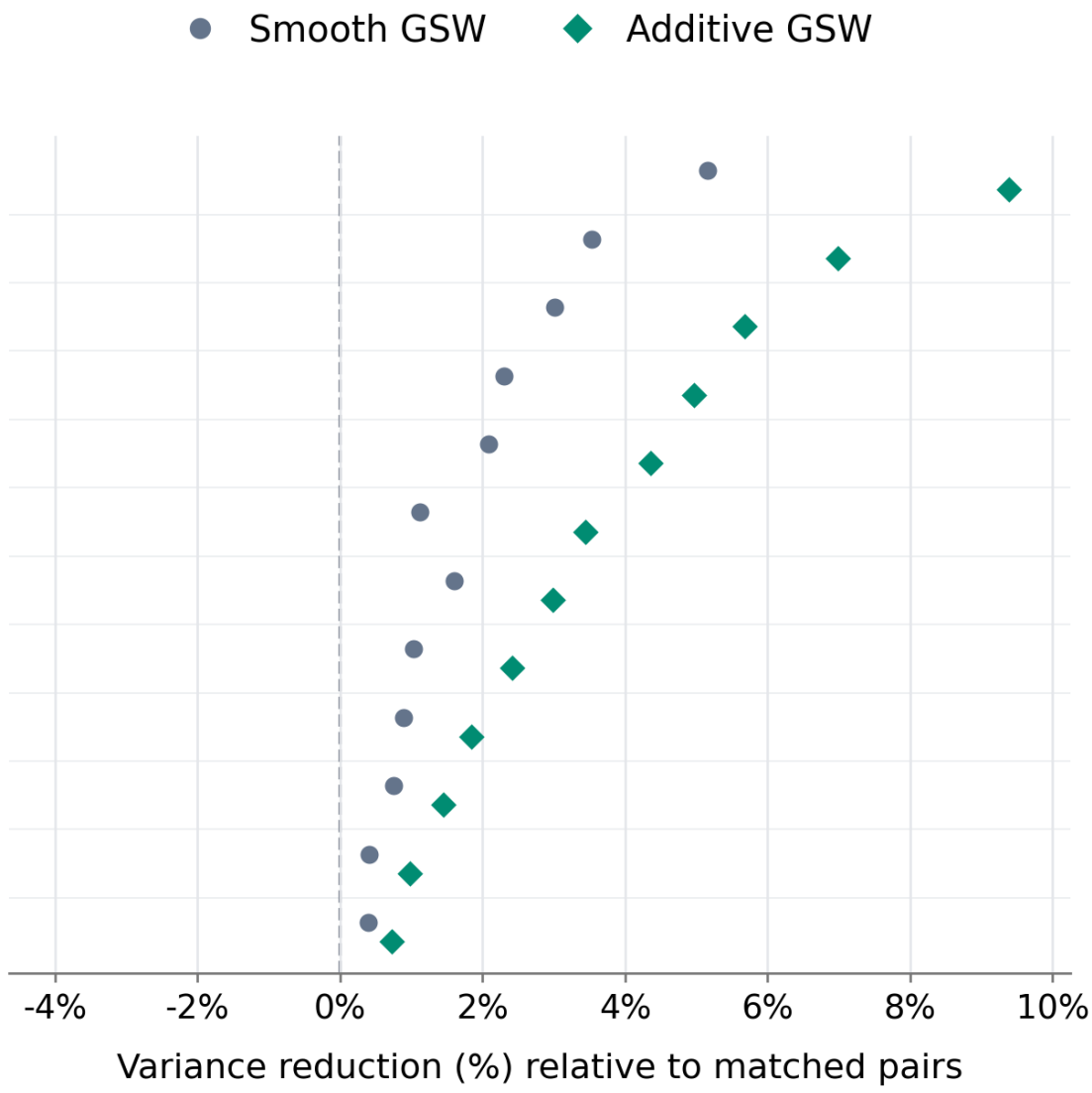
# Empirical Application: Unit-Level GSW

| Experiment | $n$ | $d$ |
|---|---|---|
| Augsburg ’22 | 124 | 21 |
| BenYishay ’20 | 96 | 22 |
| Augenblick ’26 | 828 | 11 |
| Bhandari ’23 | 444 | 31 |
| Blattman ’23 | 832 | 51 |
| Evsyukova ’25 | 364 | 12 |
| Haushofer ’16 | 940 | 14 |
| Greenstone ’25 | 292 | 15 |
| Beam ’16 | 96 | 16 |
| Gautam ’25 | 108 | 15 |
| Bandiera ’21 | 548 | 7 |
| Gupta ’26 | 1,200 | 14 |

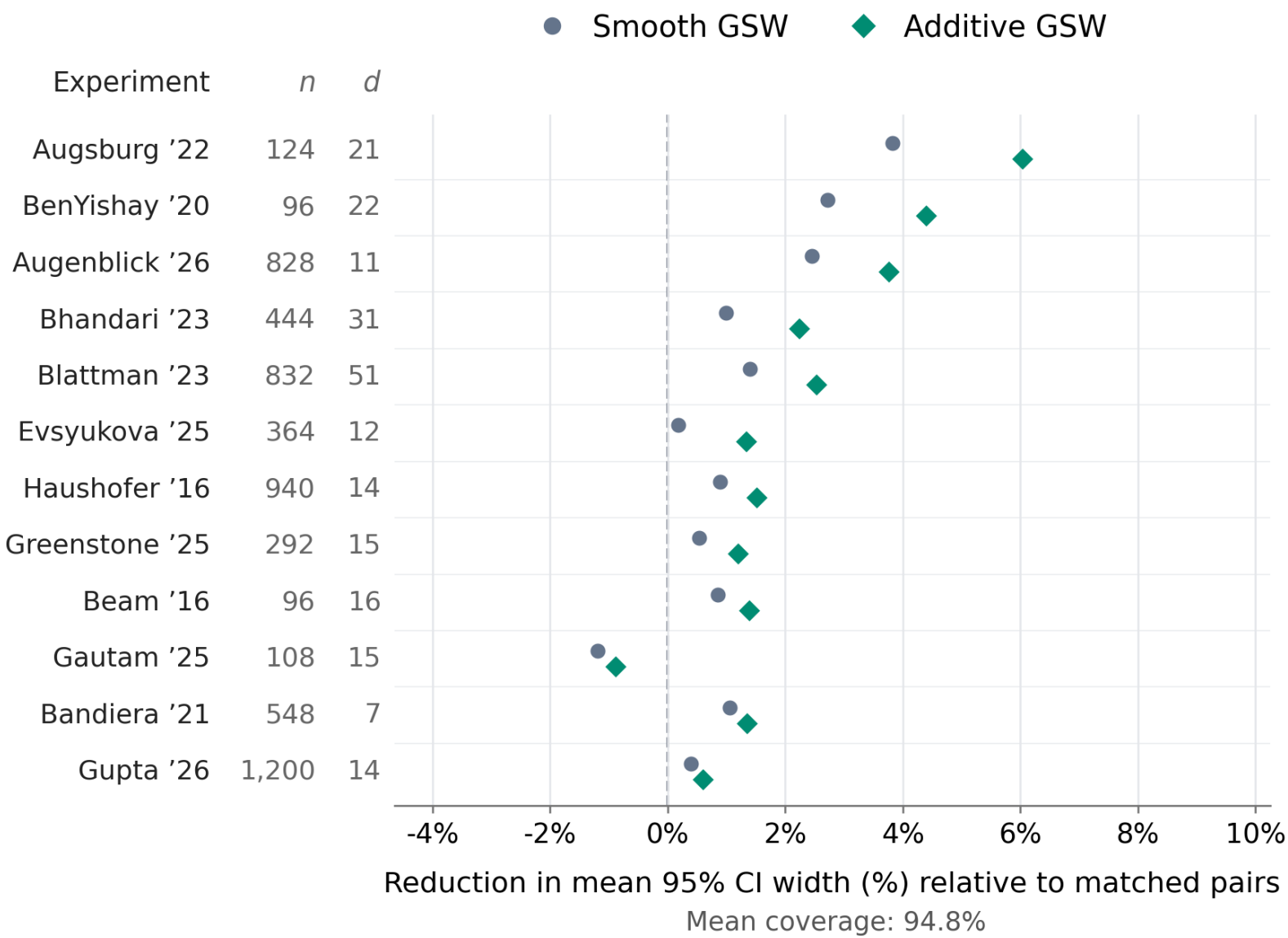


# Empirical Application: Matched GSW

| Experiment | $n$ | $d$ |
|---|---|---|
| Augsburg ’22 | 124 | 21 |
| BenYishay ’20 | 96 | 22 |
| Augenblick ’26 | 828 | 11 |
| Bhandari ’23 | 444 | 31 |
| Blattman ’23 | 832 | 51 |
| Evsyukova ’25 | 364 | 12 |
| Haushofer ’16 | 940 | 14 |
| Greenstone ’25 | 292 | 15 |
| Beam ’16 | 96 | 16 |
| Gautam ’25 | 108 | 15 |
| Bandiera ’21 | 548 | 7 |
| Gupta ’26 | 1,200 | 14 |



# Empirical Application: Inference



# Conclusion

Matched-pairs designs (Gosset, 1931) are asymptotically inefficient, even over linear models, unless $d = o(\log n)$.

**Impossibility Theorem.** The same $d = o(\log n)$ barrier applies to every design when we only impose smoothness.

Additive GSW attains restricted efficiency with $d = o(n)$, up to logs. Balances nonparametric main effects $g(x) = \sum_{j=1}^{d} g_j(x_j)$.

In empirical application, matched additive GSW uniformly attained the lowest variance, inference worked well.

**Recommendation.** Use matched additive GSW. Or other designs that target lower-complexity (e.g. bivariate) function spaces.