Experiments with Undersmoothed IPW Estimators

Nima Hejazi and Mark van der Laan

Contents

Introduction

This document summarizes the results of experiments conducted in constructing an efficient IPW estimator, based on undersmoothing estimates of the propensity score g_n via the highly adaptive lasso. Throughout, we report the results of experiments that construct several variants of an IPW estimator:

1. Truncation: (1) *minimal*, fixed at $\delta = 0.01$; (2) *profile*, in which the optimal value of the truncation parameter δ is chosen after a choice of λ is determined by minimizing the selection criterion; or
   (3) *joint*, in which a combination of δ and λ is chosen so as to minimize the selection criterion.
2. Maximum L_1-norm multiplier: fixed at 25 when Q_0 is used and 125 when Q_n is estimated via HAL.

Data-generating mechanism #1a: no positivity issues

```r
# get truth and load results data
sim_truth <- get_truth(n_samp = 1e7, tsm_contrast = 1, dgp_type = "1a")
psi_true <- sim_truth$true_psi
eff_bound <- sim_truth$effic_bound
```

$$W_1 \sim U(0.2, 0.8); W_2 \sim Bern(0.6)$$

$$A \sim Bern\left(\frac{1}{1 + \exp(2 \cdot W_1 - W_2 - W_1 \cdot W_2)}\right)$$

$$Y = A \cdot (W_1 + W_2 + W_1 * W_2) + (1 - A) \cdot W_1 + \epsilon,$$
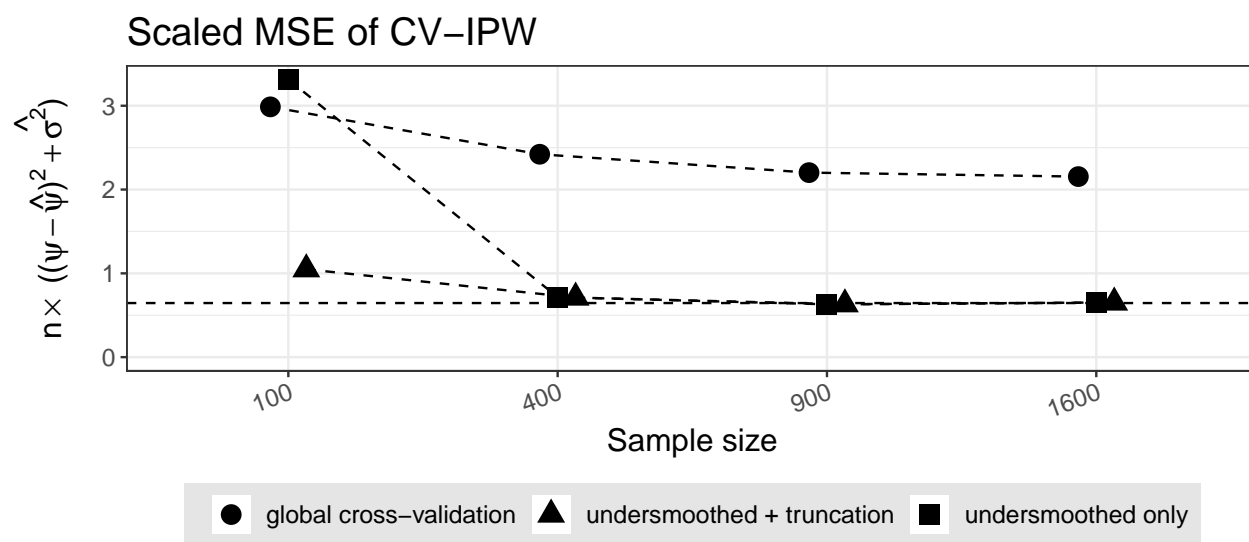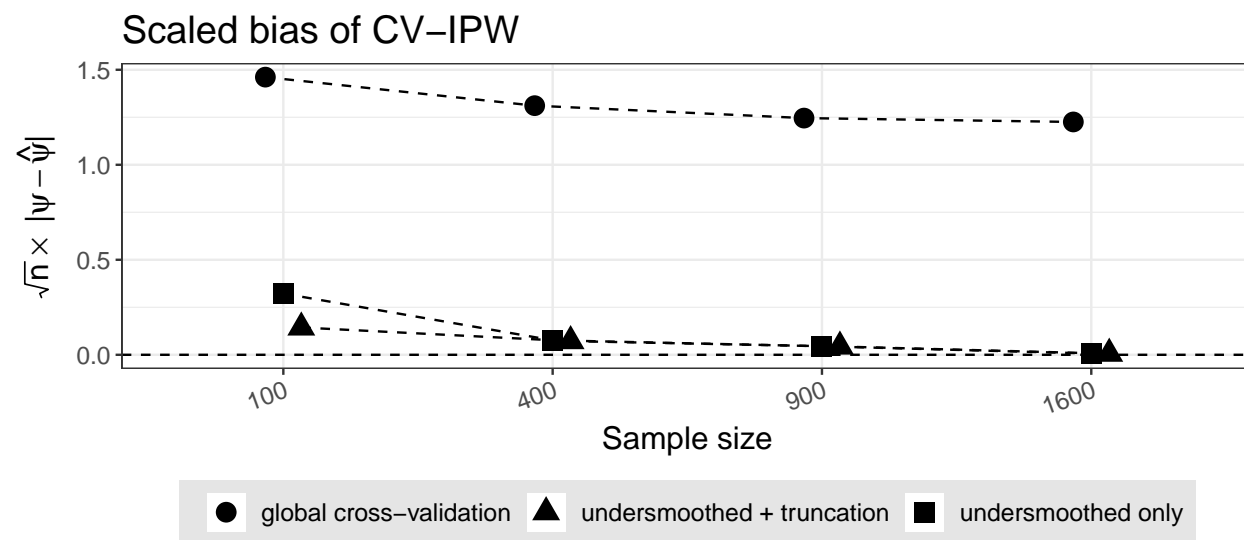
where ϵ $N(0, 0.1)$. This DGP does not admit issues with positivity, as the true propensity score g_0 has its minimum at 0.31.

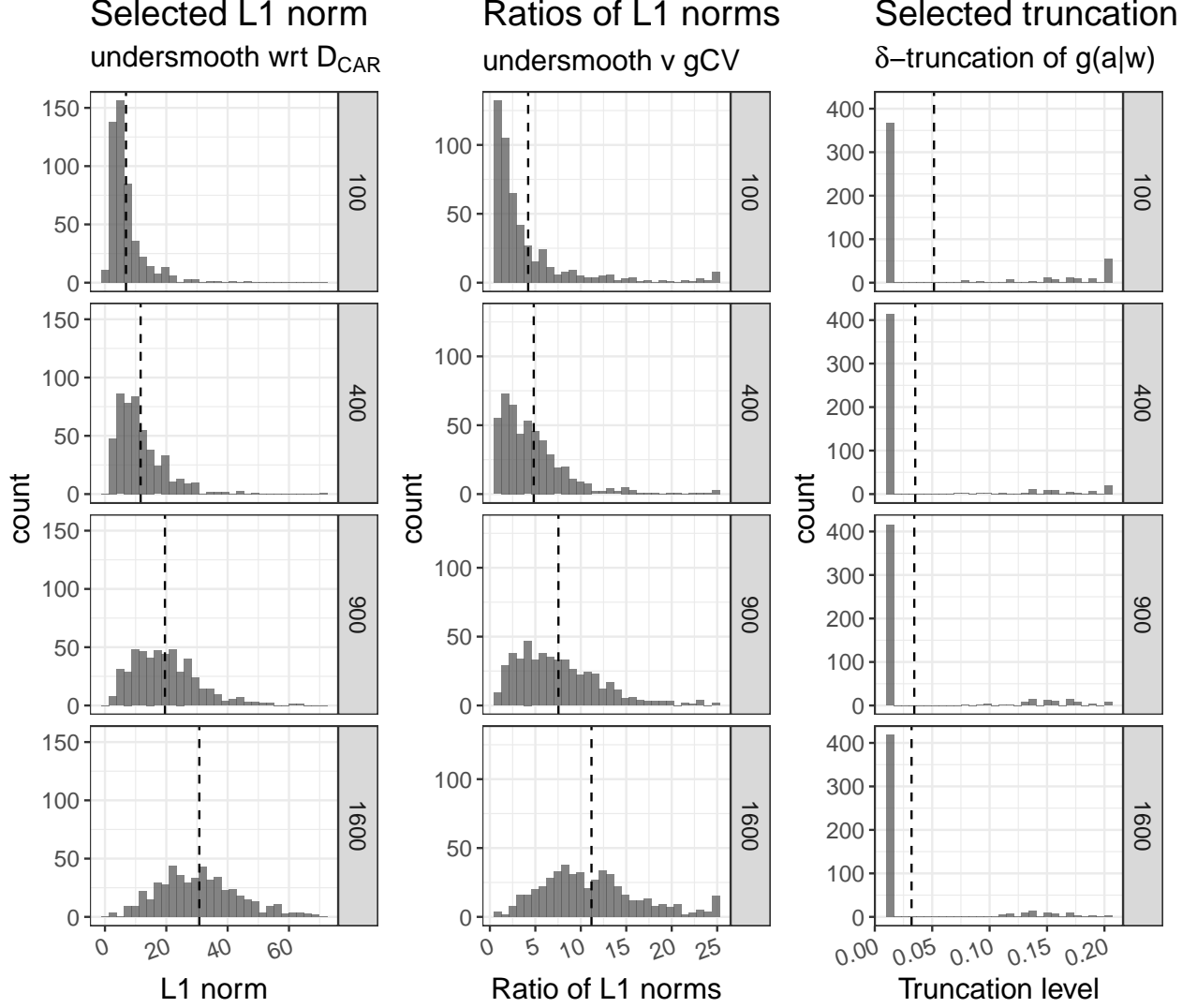## Experiment 1: Q_0 + profile minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Profile minimization over $\delta \in \{0.01, 0.02, \ldots, 0.2\}$
- Maximum L1-norm multiplier: 25

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0144	0.1437	0.0511	0.0105	1.0486	undersmoothed + truncation
400	0.0036	0.0720	0.0120	0.0018	0.7107	undersmoothed + truncation
900	0.0015	0.0435	0.0052	0.0007	0.6307	undersmoothed + truncation
1600	0.0001	0.0058	0.0029	0.0004	0.6515	undersmoothed + truncation
100	0.0322	0.3221	0.0641	0.0331	3.3133	undersmoothed only
400	0.0038	0.0764	0.0120	0.0018	0.7160	undersmoothed only
900	0.0015	0.0449	0.0052	0.0007	0.6305	undersmoothed only
1600	0.0002	0.0068	0.0029	0.0004	0.6516	undersmoothed only
100	0.1461	1.4613	0.0369	0.0299	2.9866	global cross-validation
400	0.0656	1.3111	0.0104	0.0061	2.4211	global cross-validation
900	0.0415	1.2457	0.0048	0.0024	2.2019	global cross-validation
1600	0.0306	1.2253	0.0027	0.0013	2.1547	global cross-validation

Scaled bias of CV−IPW

$\sqrt{n} \times |\psi - \hat{\psi}|$ versus Sample size

Legend: ● global cross−validation ▲ undersmoothed + truncation ■ undersmoothed only

Scaled MSE of CV−IPW

$n \times \left((\psi - \hat{\psi})^2 + \hat{\sigma^2}\right)$ versus Sample size

Legend: ● global cross−validation ▲ undersmoothed + truncation ■ undersmoothed only

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:
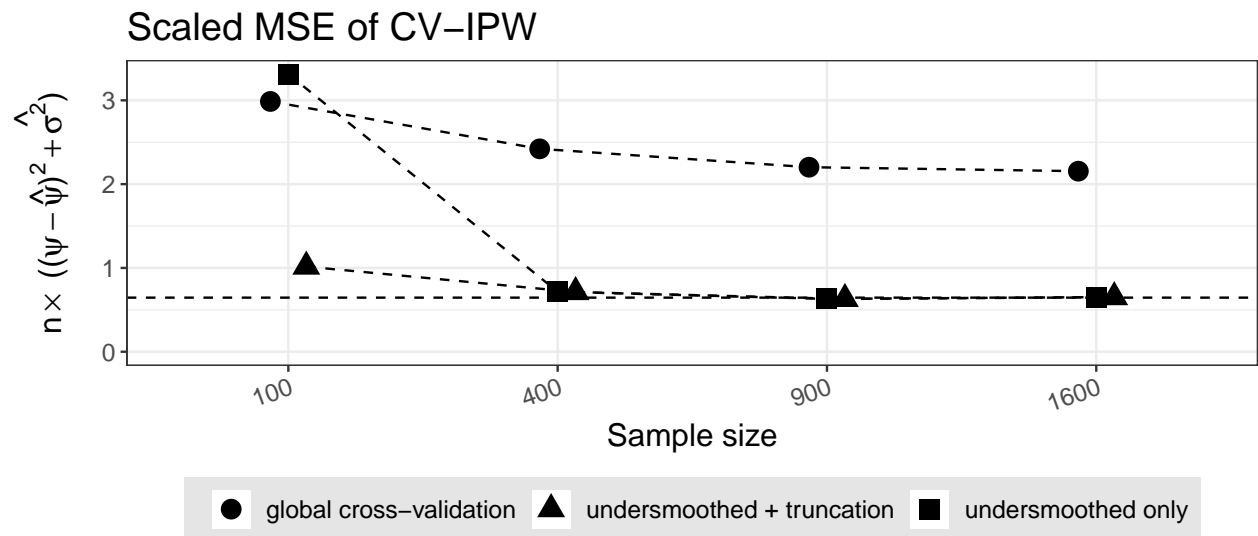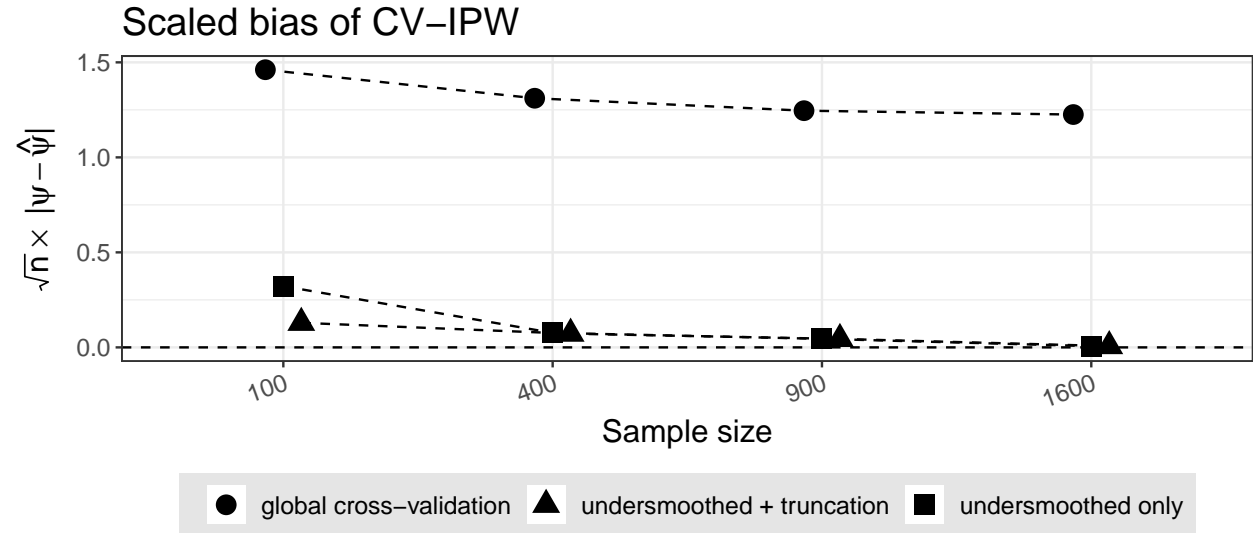
Experiment 2: $Q_0 +$ joint minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Joint minimization over $\delta \in \{0.01, 0.02, \ldots, 0.2\}$
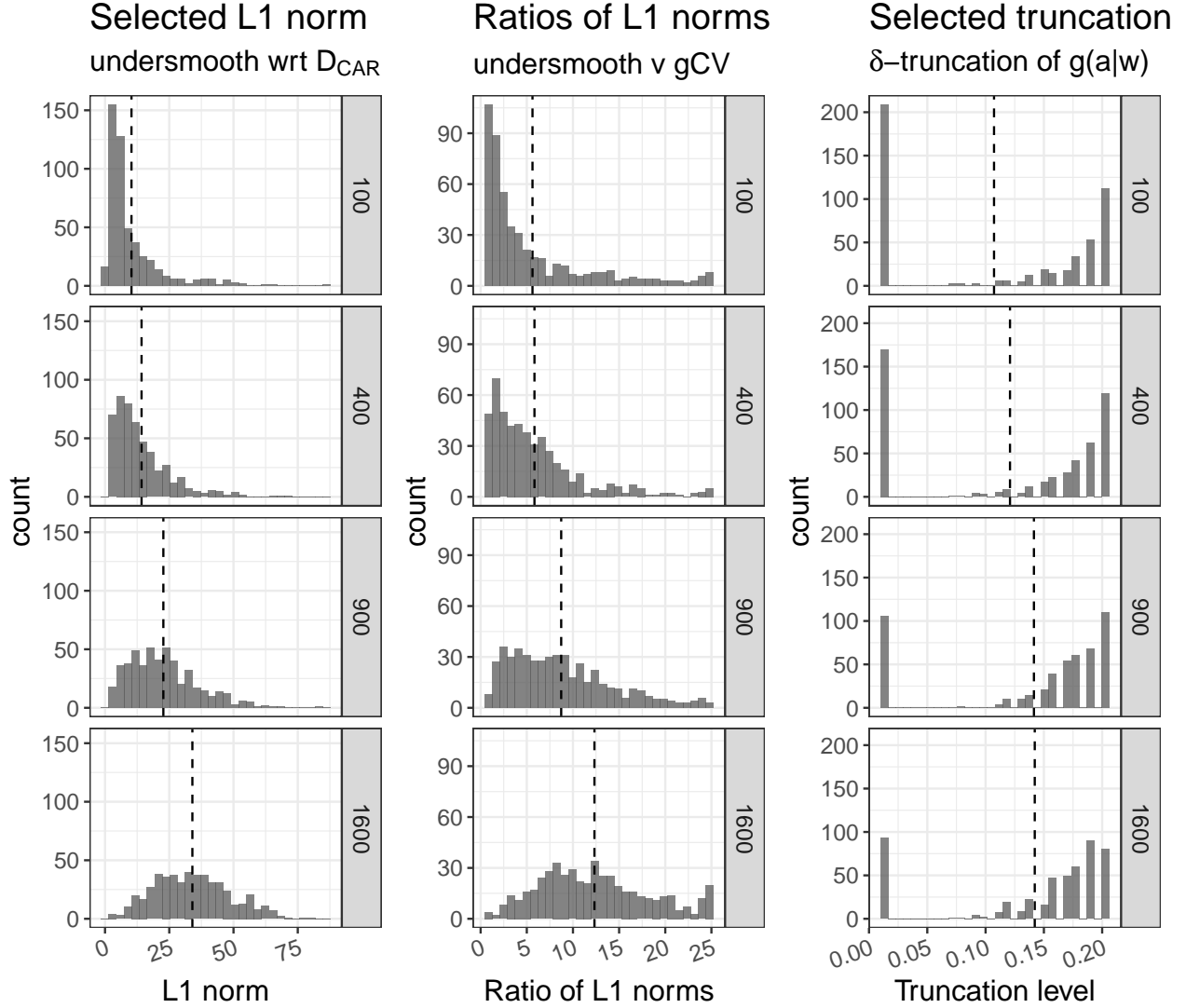- Maximum L1-norm multiplier: 25

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0129	0.1293	0.0508	0.0102	1.0161	undersmoothed + truncation
400	0.0036	0.0721	0.0119	0.0018	0.7121	undersmoothed + truncation
900	0.0015	0.0445	0.0052	0.0007	0.6298	undersmoothed + truncation
1600	0.0002	0.0068	0.0029	0.0004	0.6515	undersmoothed + truncation
100	0.0322	0.3221	0.0641	0.0331	3.3133	undersmoothed only
400	0.0038	0.0764	0.0120	0.0018	0.7160	undersmoothed only
900	0.0015	0.0449	0.0052	0.0007	0.6305	undersmoothed only
1600	0.0002	0.0068	0.0029	0.0004	0.6516	undersmoothed only
100	0.1461	1.4613	0.0369	0.0299	2.9866	global cross-validation

n_obs	bias	n_bias	avg_var	mse	nmse	label
400	0.0656	1.3111	0.0104	0.0061	2.4211	global cross-validation
900	0.0415	1.2457	0.0048	0.0024	2.2019	global cross-validation
1600	0.0306	1.2253	0.0027	0.0013	2.1547	global cross-validation

Scaled bias of CV−IPW

Scaled MSE of CV−IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:
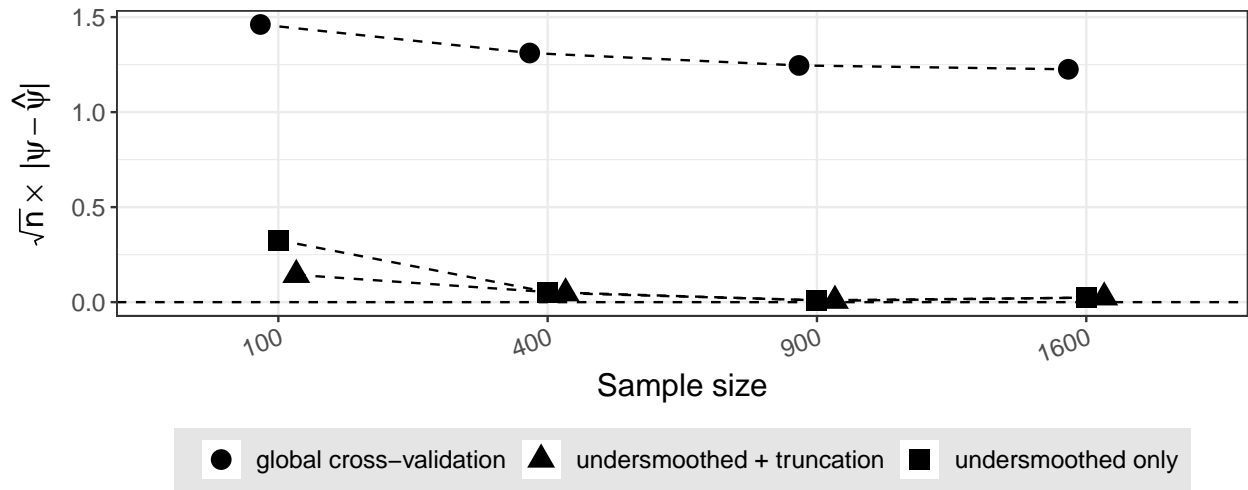
Experiment 3: Q_n + profile minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Profile minimization over $\delta \in \{0.01, 0.02, \ldots, 0.2\}$
- Maximum L1-norm multiplier: 125
- Q_n fit via HAL

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:
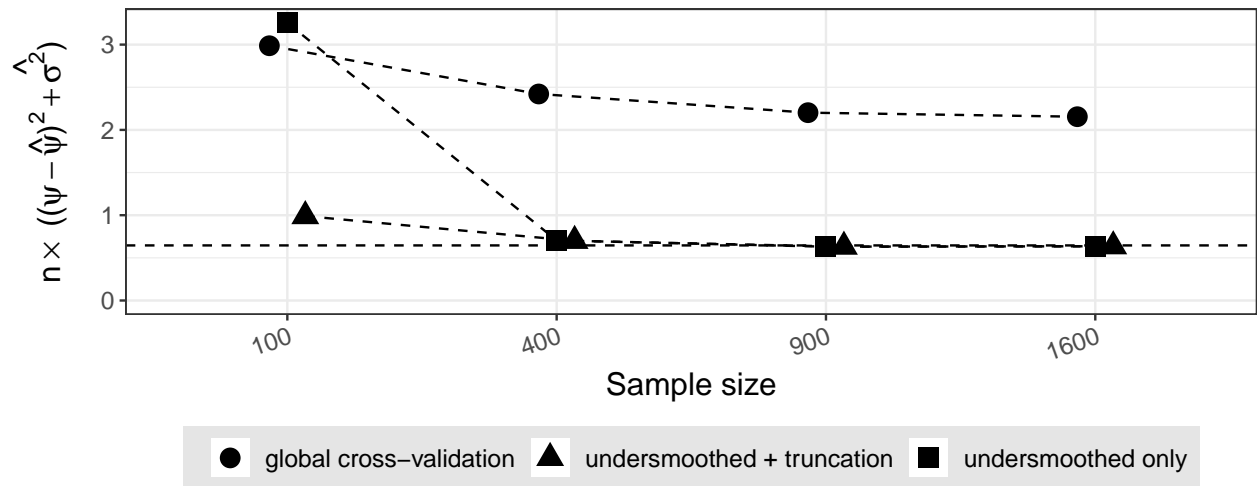
n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0143	0.1435	0.0511	0.0099	0.9894	undersmoothed + truncation
400	0.0023	0.0467	0.0119	0.0017	0.6992	undersmoothed + truncation
900	0.0002	0.0074	0.0052	0.0007	0.6311	undersmoothed + truncation
1600	0.0006	0.0252	0.0029	0.0004	0.6359	undersmoothed + truncation
100	0.0326	0.3263	0.0642	0.0326	3.2552	undersmoothed only
400	0.0026	0.0530	0.0120	0.0018	0.7046	undersmoothed only
900	0.0003	0.0097	0.0052	0.0007	0.6310	undersmoothed only

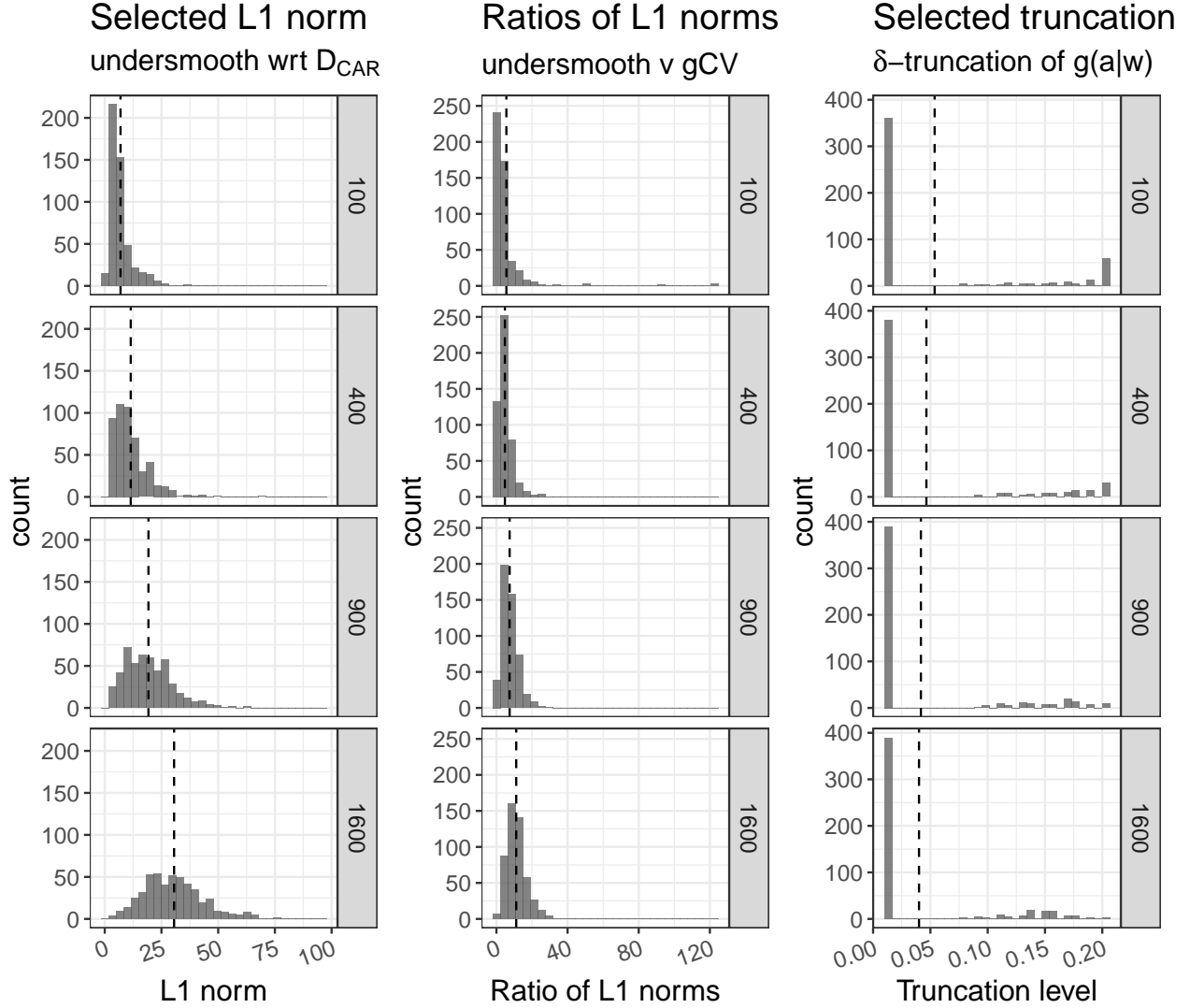| n_obs	bias	n_bias	avg_var	mse	nmse	label
1600	0.0006	0.0233	0.0029	0.0004	0.6358	undersmoothed only
100	0.1461	1.4613	0.0369	0.0299	2.9866	global cross-validation
400	0.0656	1.3111	0.0104	0.0061	2.4211	global cross-validation
900	0.0415	1.2457	0.0048	0.0024	2.2019	global cross-validation
1600	0.0306	1.2253	0.0027	0.0013	2.1547	global cross-validation

Scaled bias of CV–IPW

Scaled MSE of CV–IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:

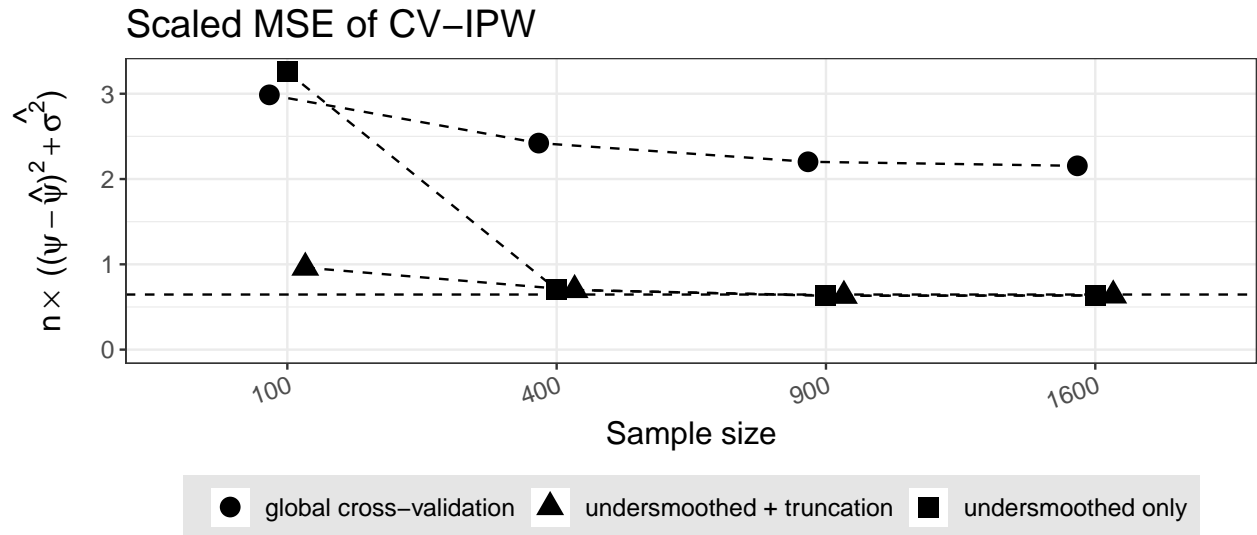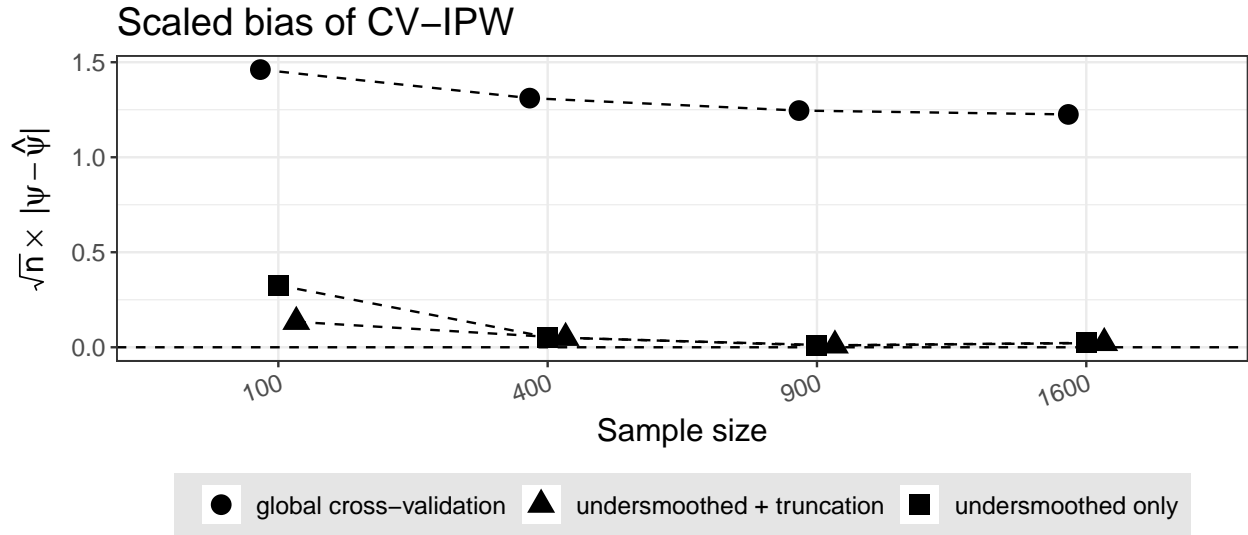| Selected L1 norm	Ratios of L1 norms	Selected truncation
undersmooth wrt D_CAR	undersmooth v gCV	δ–truncation of g(a|w)

Experiment 4: Q_n + joint minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Joint minimization over $\delta \in \{0.01, 0.02, \ldots, 0.2\}$
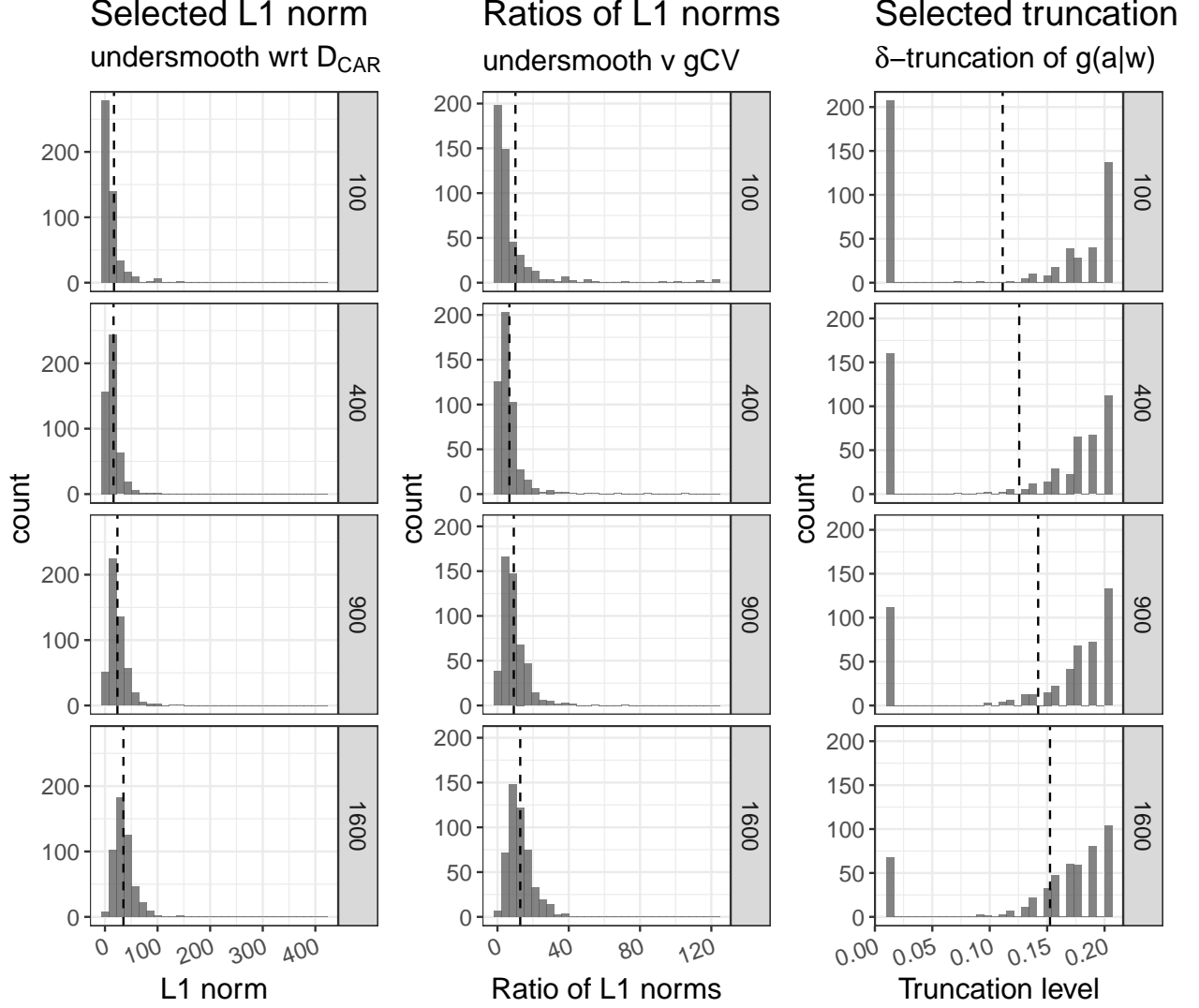- Maximum L1-norm multiplier: 125
- Q_n fit via HAL

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0134	0.1337	0.0511	0.0096	0.9612	undersmoothed + truncation
400	0.0025	0.0495	0.0118	0.0017	0.6989	undersmoothed + truncation
900	0.0003	0.0097	0.0052	0.0007	0.6314	undersmoothed + truncation
1600	0.0006	0.0225	0.0029	0.0004	0.6352	undersmoothed + truncation
100	0.0326	0.3263	0.0642	0.0326	3.2552	undersmoothed only
400	0.0026	0.0530	0.0120	0.0018	0.7046	undersmoothed only
900	0.0003	0.0097	0.0052	0.0007	0.6310	undersmoothed only
1600	0.0006	0.0233	0.0029	0.0004	0.6358	undersmoothed only

n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.1461	1.4613	0.0369	0.0299	2.9866	global cross-validation
400	0.0656	1.3111	0.0104	0.0061	2.4211	global cross-validation
900	0.0415	1.2457	0.0048	0.0024	2.2019	global cross-validation
1600	0.0306	1.2253	0.0027	0.0013	2.1547	global cross-validation

Scaled bias of CV−IPW

Scaled MSE of CV−IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:

Selected L1 norm	Ratios of L1 norms	Selected truncation
undersmooth wrt D$_{CAR}$	undersmooth v gCV	δ–truncation of g(a|w)

Data-generating mechanism #2a: severe positivity issues

$$W_1 \sim U(0, 0.6); W_2 \sim Bern(0.05)$$

$$A \sim Bern\left(\frac{1}{1 + \exp(2 \cdot W_1 - 5 \cdot W_2 - W_1 \cdot W_2)}\right)$$

$$Y = A \cdot (W_1 + W_2 + W_1 \cdot W_2) + (1 - A) \cdot W_1 + \epsilon,$$

where $\epsilon \sim N(0, 0.1)$. This DGP leads directly to issues with positivity, as the true propensity score g_0 has its minimum at 0.0067.
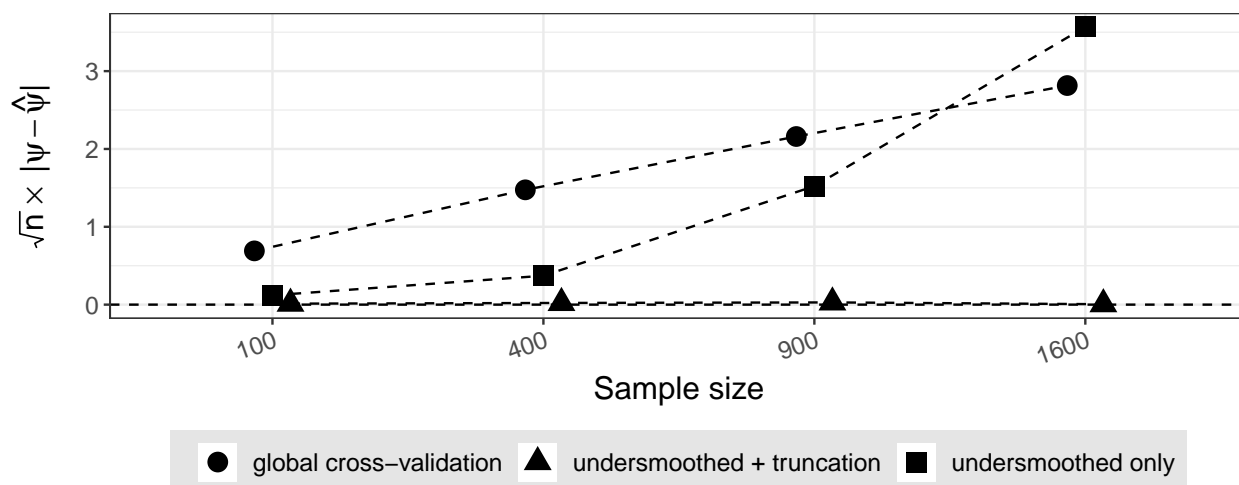
Experiment 1: Q_0 + profile minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Profile minimization over $\delta \in \{0.01, 0.02, \dots, 0.3\}$
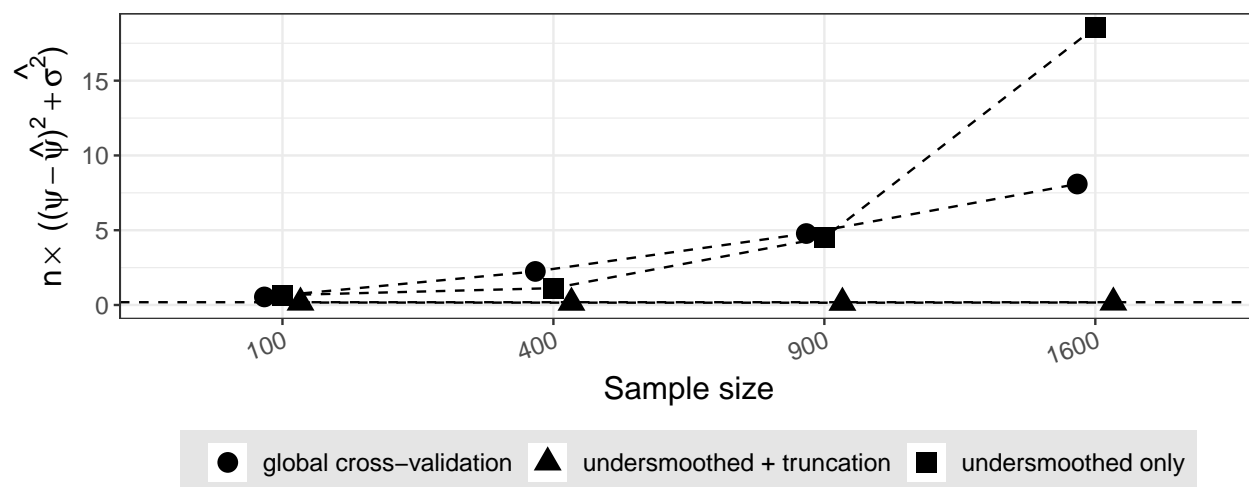- Maximum L1-norm multiplier: 40

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

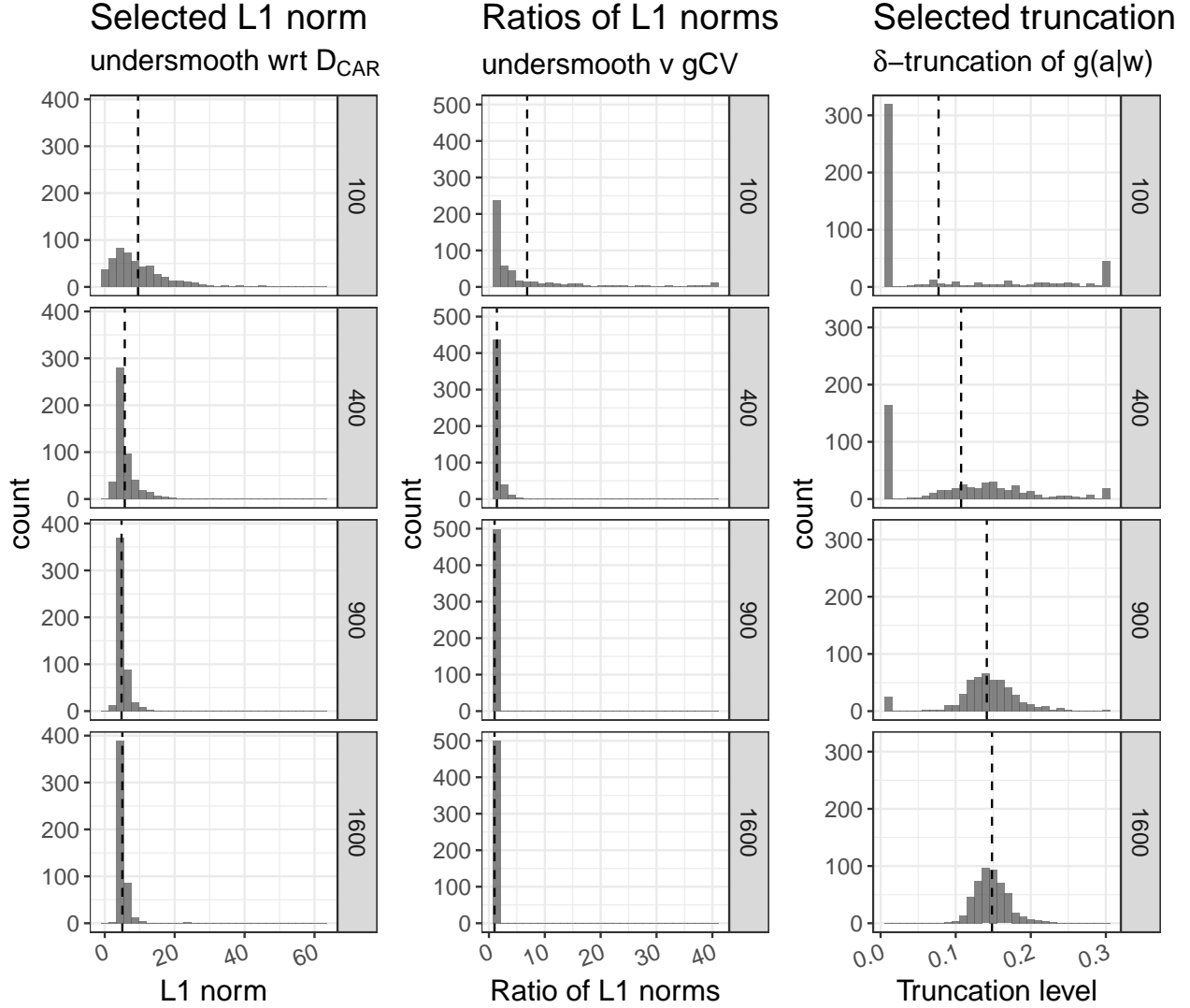| n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0013	0.0127	0.0029	0.0016	0.1634	undersmoothed + truncation
400	0.0011	0.0216	0.0007	0.0004	0.1561	undersmoothed + truncation
900	0.0010	0.0286	0.0003	0.0002	0.1515	undersmoothed + truncation
1600	0.0001	0.0055	0.0002	0.0001	0.1667	undersmoothed + truncation
100	0.0120	0.1200	0.0053	0.0066	0.6612	undersmoothed only
400	0.0186	0.3710	0.0013	0.0028	1.1261	undersmoothed only
900	0.0508	1.5227	0.0009	0.0050	4.5016	undersmoothed only
1600	0.0892	3.5672	0.0009	0.0116	18.5669	undersmoothed only
100	0.0691	0.6907	0.0011	0.0054	0.5392	global cross-validation
400	0.0737	1.4749	0.0003	0.0056	2.2445	global cross-validation
900	0.0720	2.1600	0.0002	0.0053	4.7776	global cross-validation
1600	0.0704	2.8147	0.0002	0.0051	8.0893	global cross-validation

Scaled bias of CV−IPW

Scaled MSE of CV−IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:

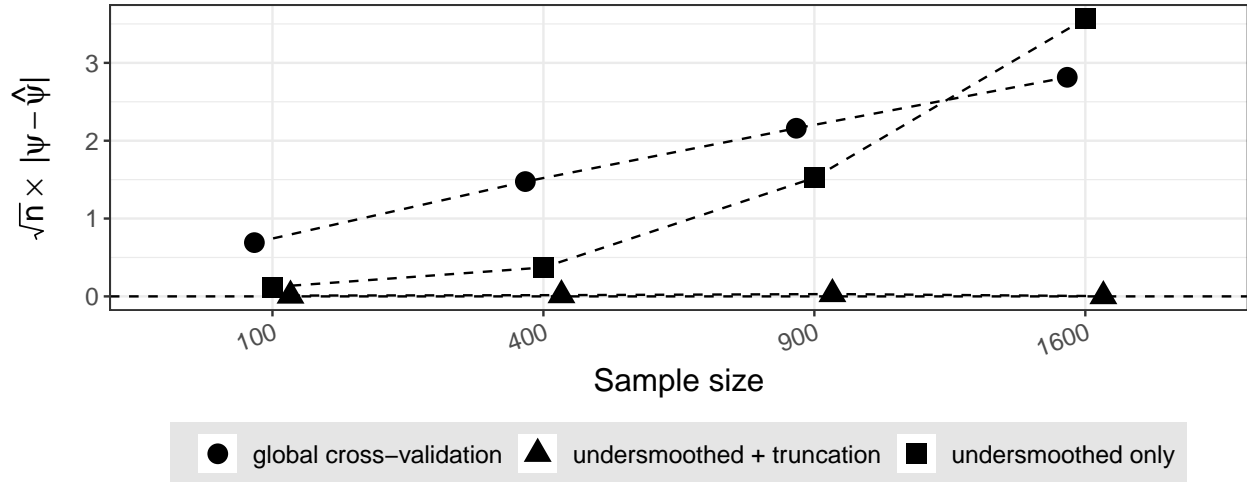| | Selected L1 norm | Ratios of L1 norms | Selected truncation |
| | undersmooth wrt D_{CAR} | undersmooth v gCV | δ–truncation of g(a|w) |

Experiment 2: Q_0 + joint minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Joint minimization over $\delta \in \{0.01, 0.02, \ldots, 0.3\}$
- Maximum L1-norm multiplier: 40

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:
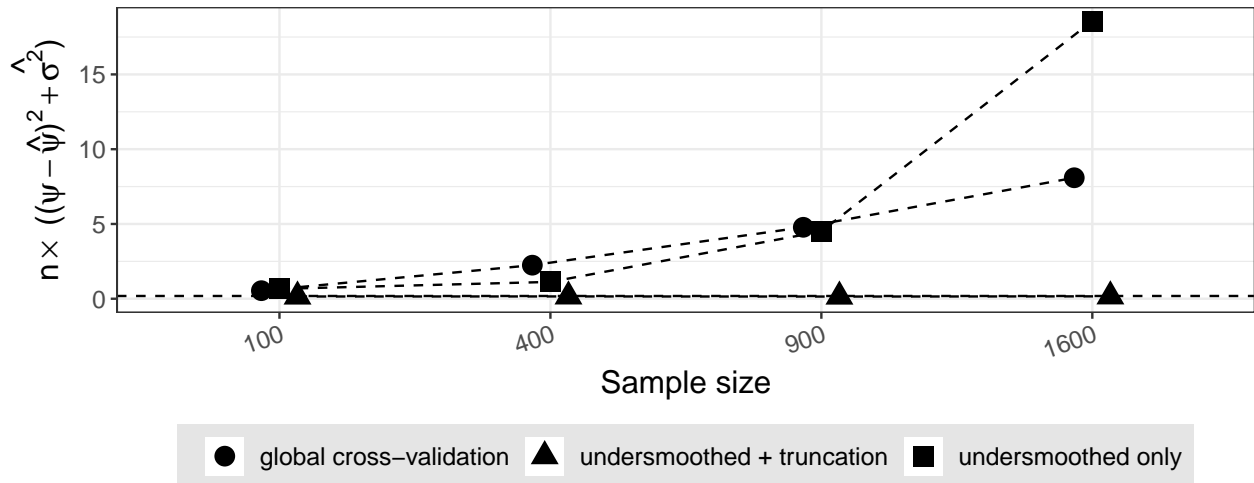
n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0010	0.0105	0.0025	0.0015	0.1498	undersmoothed + truncation
400	0.0008	0.0158	0.0006	0.0004	0.1550	undersmoothed + truncation
900	0.0010	0.0298	0.0003	0.0002	0.1465	undersmoothed + truncation
1600	0.0001	0.0034	0.0002	0.0001	0.1606	undersmoothed + truncation
100	0.0120	0.1200	0.0053	0.0066	0.6612	undersmoothed only
400	0.0186	0.3710	0.0013	0.0028	1.1261	undersmoothed only
900	0.0508	1.5227	0.0009	0.0050	4.5016	undersmoothed only
1600	0.0892	3.5672	0.0009	0.0116	18.5669	undersmoothed only
100	0.0691	0.6907	0.0011	0.0054	0.5392	global cross-validation

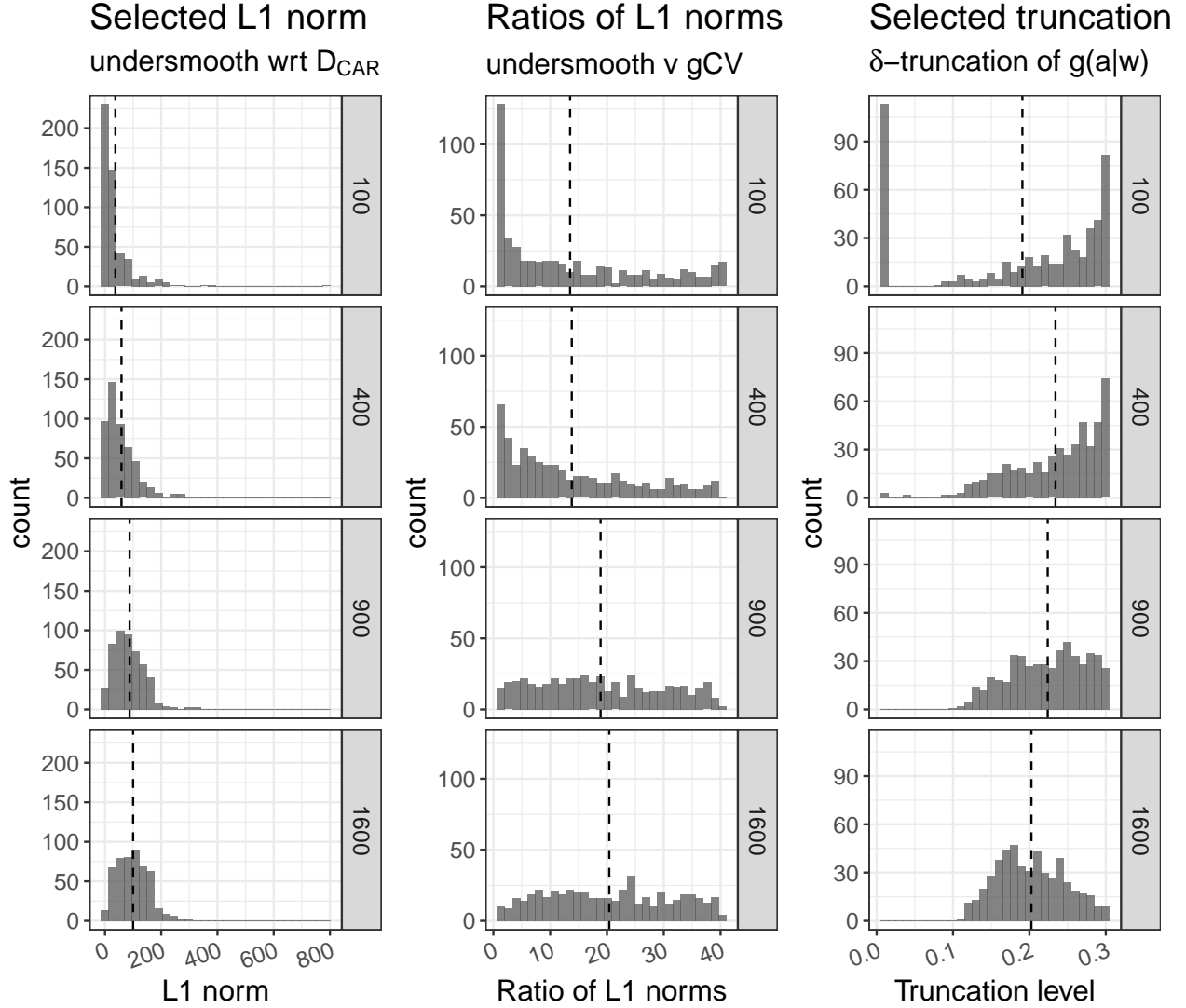| n_obs	bias	n_bias	avg_var	mse	nmse	label
400	0.0737	1.4749	0.0003	0.0056	2.2445	global cross-validation
900	0.0720	2.1600	0.0002	0.0053	4.7776	global cross-validation
1600	0.0704	2.8147	0.0002	0.0051	8.0893	global cross-validation

Scaled bias of CV−IPW

Scaled MSE of CV−IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:

Selected L1 norm — undersmooth wrt D_{CAR} | **Ratios of L1 norms** — undersmooth v gCV | **Selected truncation** — δ–truncation of $g(a|w)$

(histograms for sample sizes 100, 400, 900, 1600)

x-axes: L1 norm | Ratio of L1 norms | Truncation level

Data-generating mechanism #2b: moderate positivity issues

$$W_1 \sim U(0, 0.6); W_2 \sim Bern(0.05)$$

$$A \sim Bern\left(\frac{1}{1 + \exp(2 \cdot W_1 - 5 \cdot W_2 + 3 \cdot W_1 \cdot W_2 - W_1^4)}\right)$$

$$Y = A \cdot (W_1 + W_1 \cdot W_2) + (1 - A) \cdot W_2 + \epsilon,$$

where $\epsilon \; N(0, 0.1)$. This DGP leads directly to issues with positivity, as the true propensity score g_0 has its minimum at 0.0474.
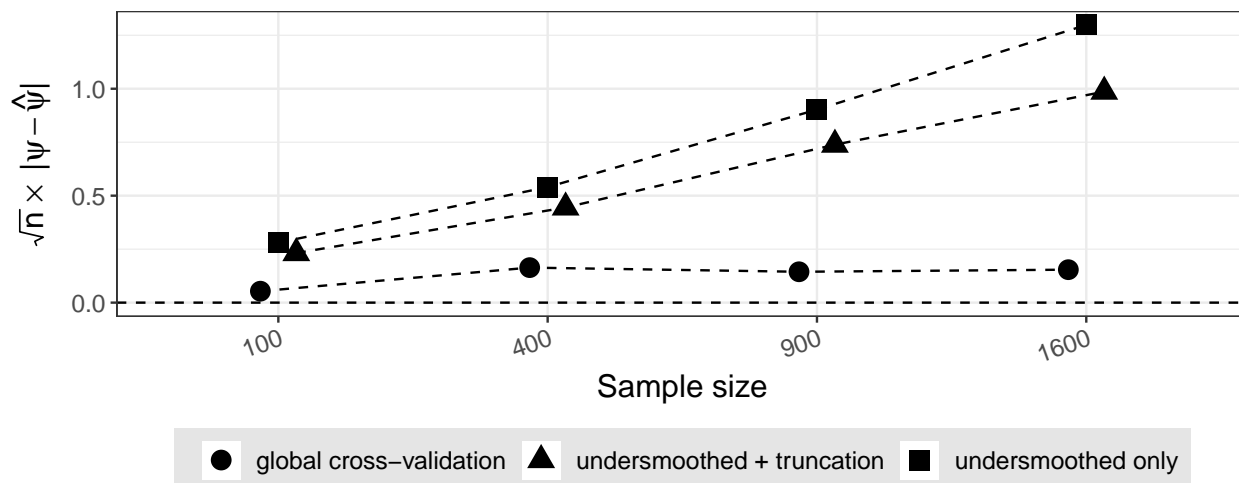
Experiment 1: Q_0 + profile minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Profile minimization over $\delta \in \{0.01, 0.02, \dots, 0.2\}$
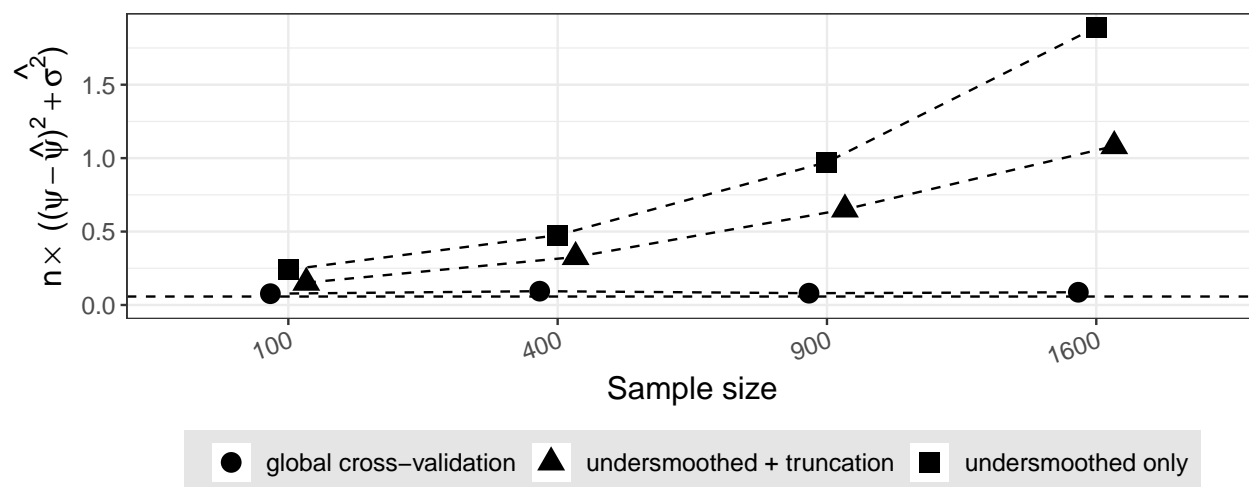- Maximum L1-norm multiplier: 30

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

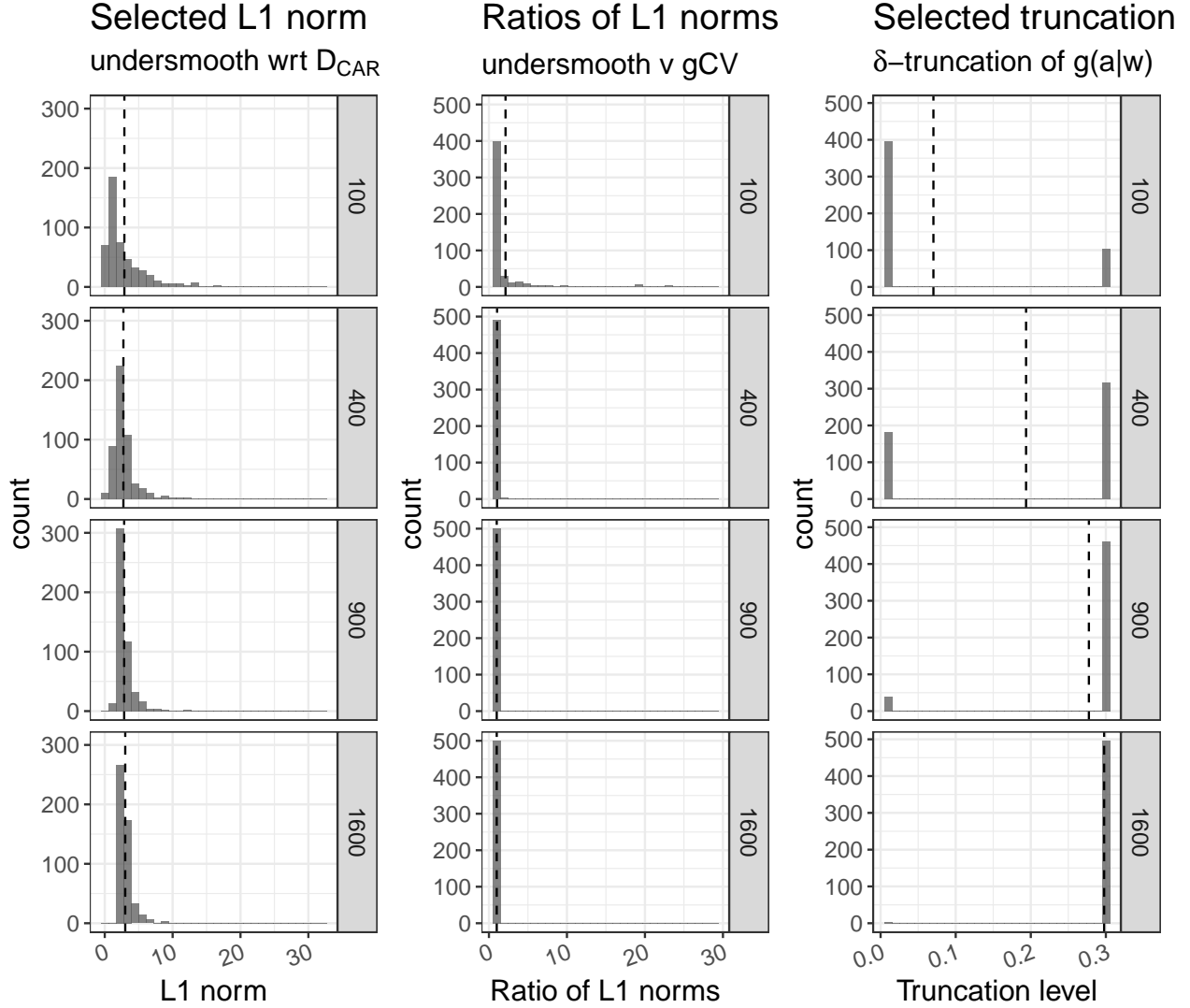| n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0232	0.2317	0.0016	0.0015	0.1520	undersmoothed + truncation
400	0.0222	0.4449	0.0004	0.0008	0.3269	undersmoothed + truncation
900	0.0246	0.7385	0.0002	0.0007	0.6507	undersmoothed + truncation
1600	0.0247	0.9864	0.0001	0.0007	1.0821	undersmoothed + truncation
100	0.0281	0.2805	0.0019	0.0024	0.2402	undersmoothed only
400	0.0270	0.5402	0.0004	0.0012	0.4752	undersmoothed only
900	0.0301	0.9023	0.0002	0.0011	0.9727	undersmoothed only
1600	0.0325	1.2981	0.0001	0.0012	1.8907	undersmoothed only
100	0.0054	0.0536	0.0012	0.0008	0.0770	global cross-validation
400	0.0082	0.1637	0.0003	0.0002	0.0940	global cross-validation
900	0.0048	0.1443	0.0002	0.0001	0.0804	global cross-validation
1600	0.0039	0.1541	0.0001	0.0001	0.0868	global cross-validation

Scaled bias of CV–IPW

Scaled MSE of CV–IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:

Selected L1 norm	Ratios of L1 norms	Selected truncation
undersmooth wrt D_{CAR}	undersmooth v gCV	δ–truncation of g(a|w)

## Experiment 2: Q_0 + joint minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Joint minimization over $\delta \in \{0.01, 0.02, \ldots, 0.2\}$
- Maximum L1-norm multiplier: 30

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.0225	0.2254	0.0016	0.0014	0.1445	undersmoothed + truncation
400	0.0222	0.4442	0.0004	0.0008	0.3266	undersmoothed + truncation
900	0.0246	0.7379	0.0002	0.0007	0.6500	undersmoothed + truncation
1600	0.0246	0.9849	0.0001	0.0007	1.0791	undersmoothed + truncation
100	0.0281	0.2805	0.0019	0.0024	0.2402	undersmoothed only
400	0.0270	0.5402	0.0004	0.0012	0.4752	undersmoothed only
900	0.0301	0.9023	0.0002	0.0011	0.9727	undersmoothed only
1600	0.0325	1.2981	0.0001	0.0012	1.8907	undersmoothed only
100	0.0054	0.0536	0.0012	0.0008	0.0770	global cross-validation

| n_obs	bias	n_bias	avg_var	mse	nmse	label
400	0.0082	0.1637	0.0003	0.0002	0.0940	global cross-validation
900	0.0048	0.1443	0.0002	0.0001	0.0804	global cross-validation
1600	0.0039	0.1541	0.0001	0.0001	0.0868	global cross-validation

Scaled bias of CV−IPW
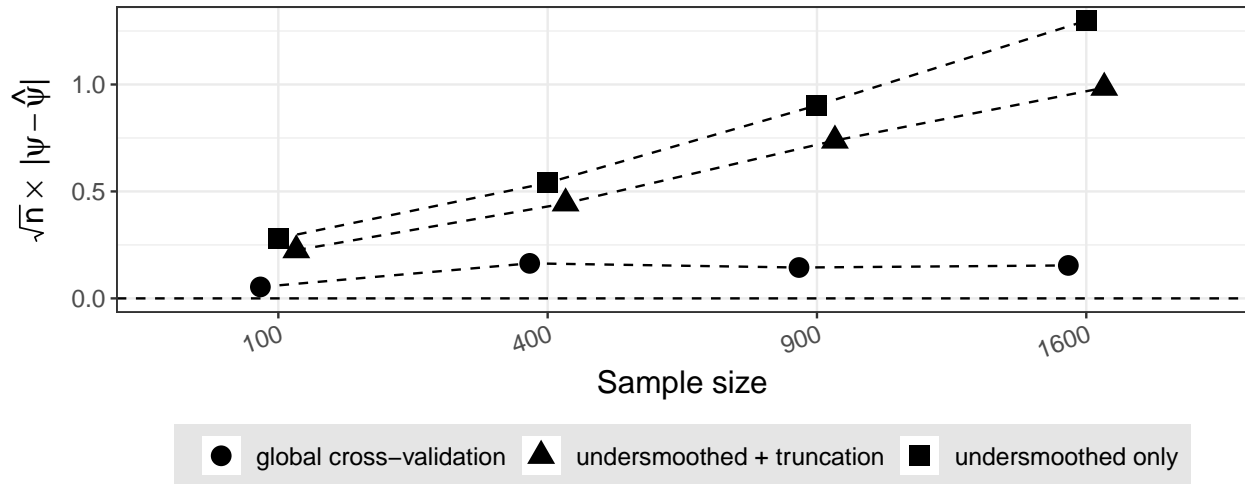
Scaled MSE of CV−IPW
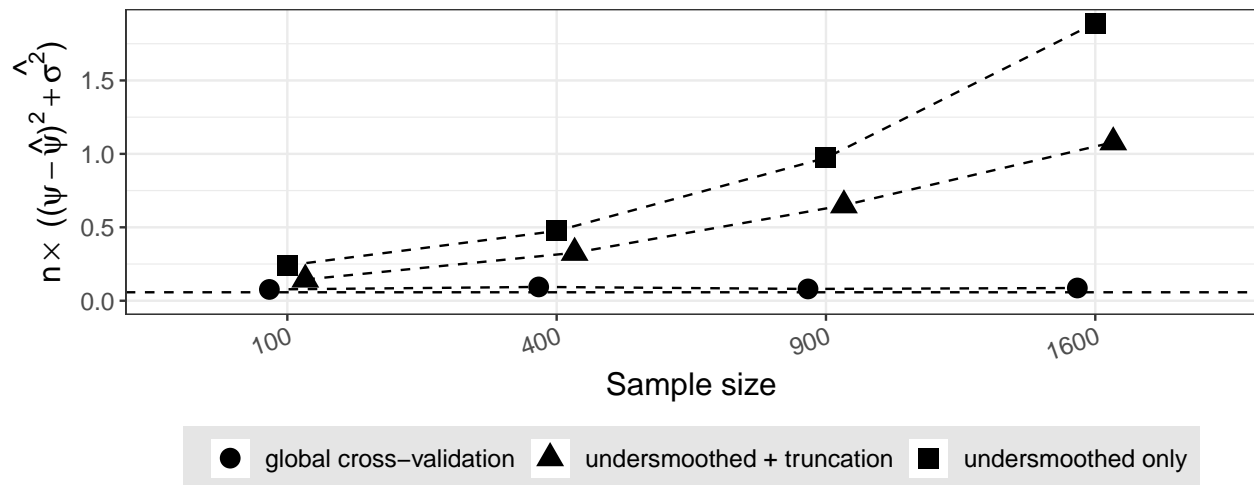
We now visualize estimator performance using custom metrics for the estimator that does allow truncation:
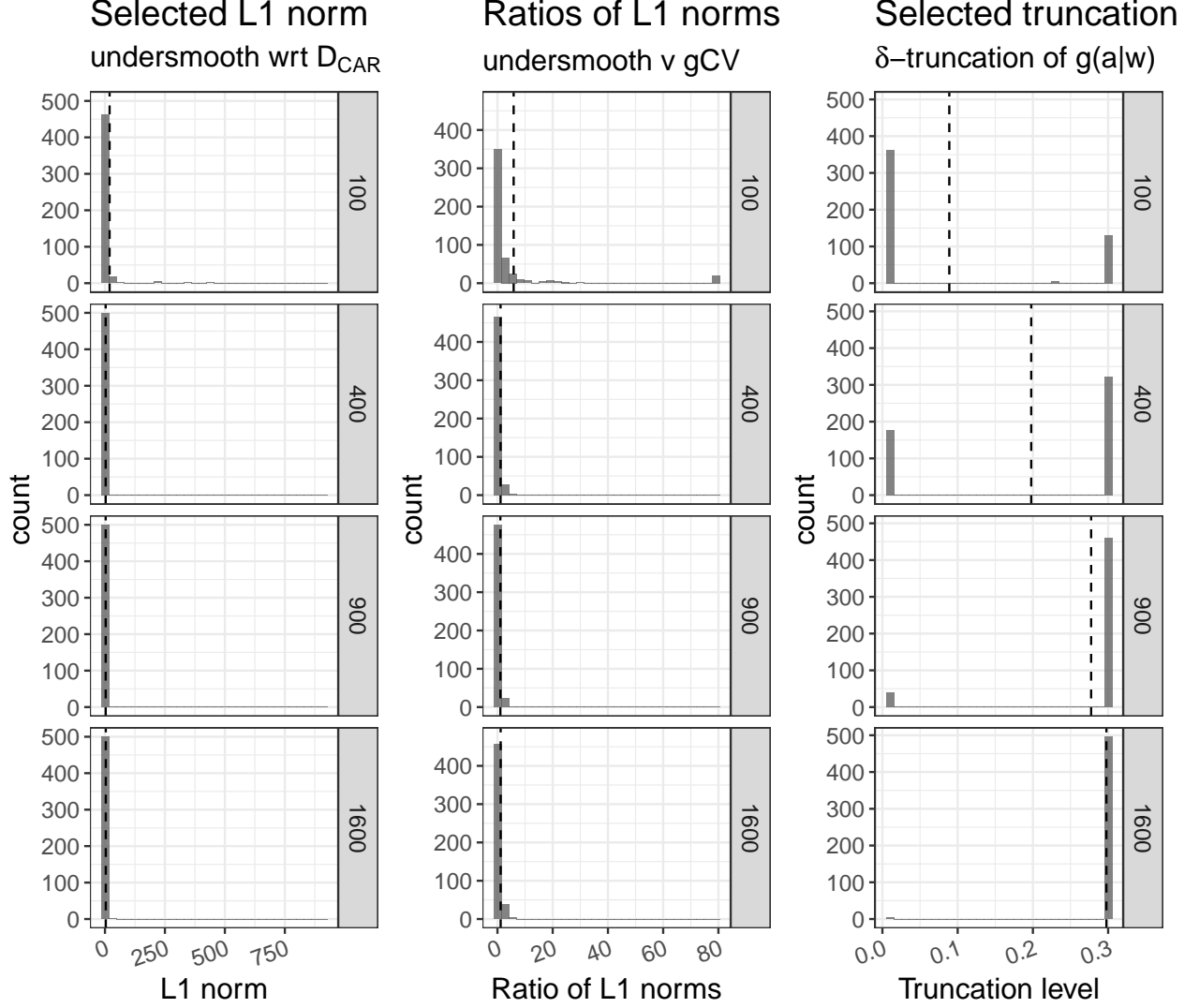
Data-generating mechanism #3: Ashkan's proposal

$$W_1 \sim U(-2, 2); W_2 \sim N(0, 0.5); W_3 \sim Bern(0.5)$$
$$W_4 \sim Bern(0.5); W_5 \sim U(-1, 1); W_6 \sim Bern(0.5)$$
$$W_7 \sim Bern(0.5); W_8 \sim Bern(0.5); W_9 \sim U(-1, 1)$$
$$A \sim Bern\left(\frac{1}{1 + \exp(W_2^2 - \exp(W_1/2) - W_1 \cdot W_2)}\right)$$
$$Y = A - 2 \cdot W_2^2 + 2 \cdot W_1 + 2 \cdot \frac{1}{n}\sum_{i=1}^{n} W_2^2 + W_2 + W_1 \cdot W_2 + \epsilon,$$
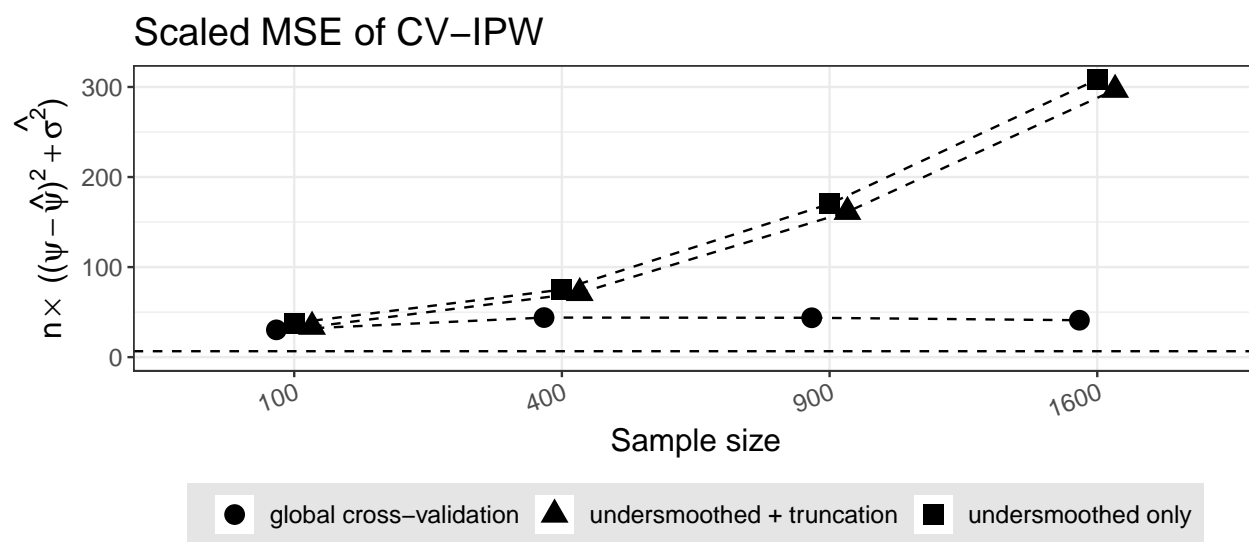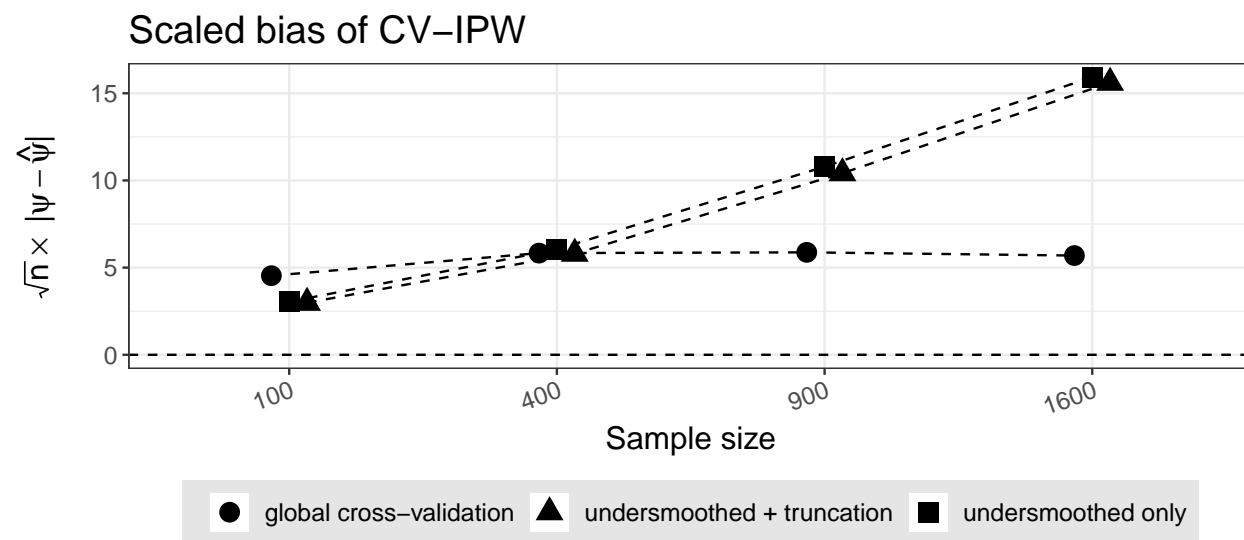
where ϵ $N(0, 0.25)$. This DGP does not admit issues with positivity, as the true propensity score g_0 has its minimum at 0.1348.
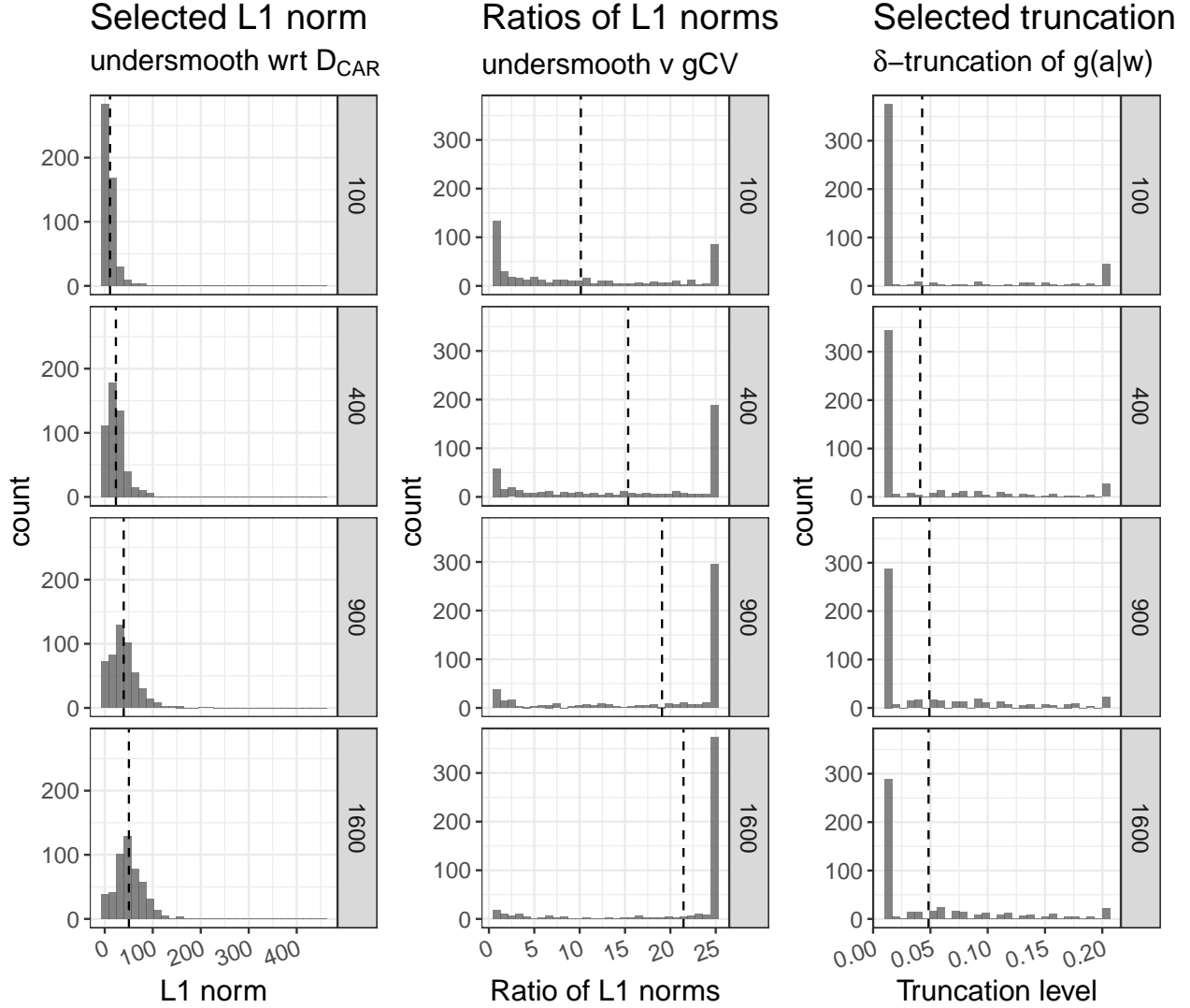
Experiment 1: Q_0 + profile minimization

- Horvitz-Thompson (nonstabilized) IPW estimator
- Profile minimization over $\delta \in \{0.01, 0.02, \dots, 0.2\}$
- Maximum L1-norm multiplier: 25

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

n_obs	bias	n_bias	avg_var	mse	nmse	label
100	0.2984	2.9839	1.8495	0.3378	33.7786	undersmoothed + truncation
400	0.2900	5.7991	0.2326	0.1781	71.2388	undersmoothed + truncation
900	0.3472	10.4160	0.0869	0.1793	161.3393	undersmoothed + truncation
1600	0.3901	15.6040	0.0341	0.1856	296.9281	undersmoothed + truncation
100	0.3042	3.0419	2.0595	0.3791	37.9087	undersmoothed only
400	0.3023	6.0455	0.2484	0.1875	75.0193	undersmoothed only
900	0.3598	10.7930	0.0939	0.1891	170.1565	undersmoothed only
1600	0.3981	15.9255	0.0360	0.1927	308.3211	undersmoothed only
100	0.4537	4.5375	0.1591	0.3032	30.3182	global cross-validation
400	0.2916	5.8317	0.0471	0.1099	43.9404	global cross-validation
900	0.1959	5.8758	0.0228	0.0486	43.7265	global cross-validation
1600	0.1423	5.6916	0.0136	0.0256	41.0054	global cross-validation

Scaled bias of CV–IPW

Scaled MSE of CV–IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:

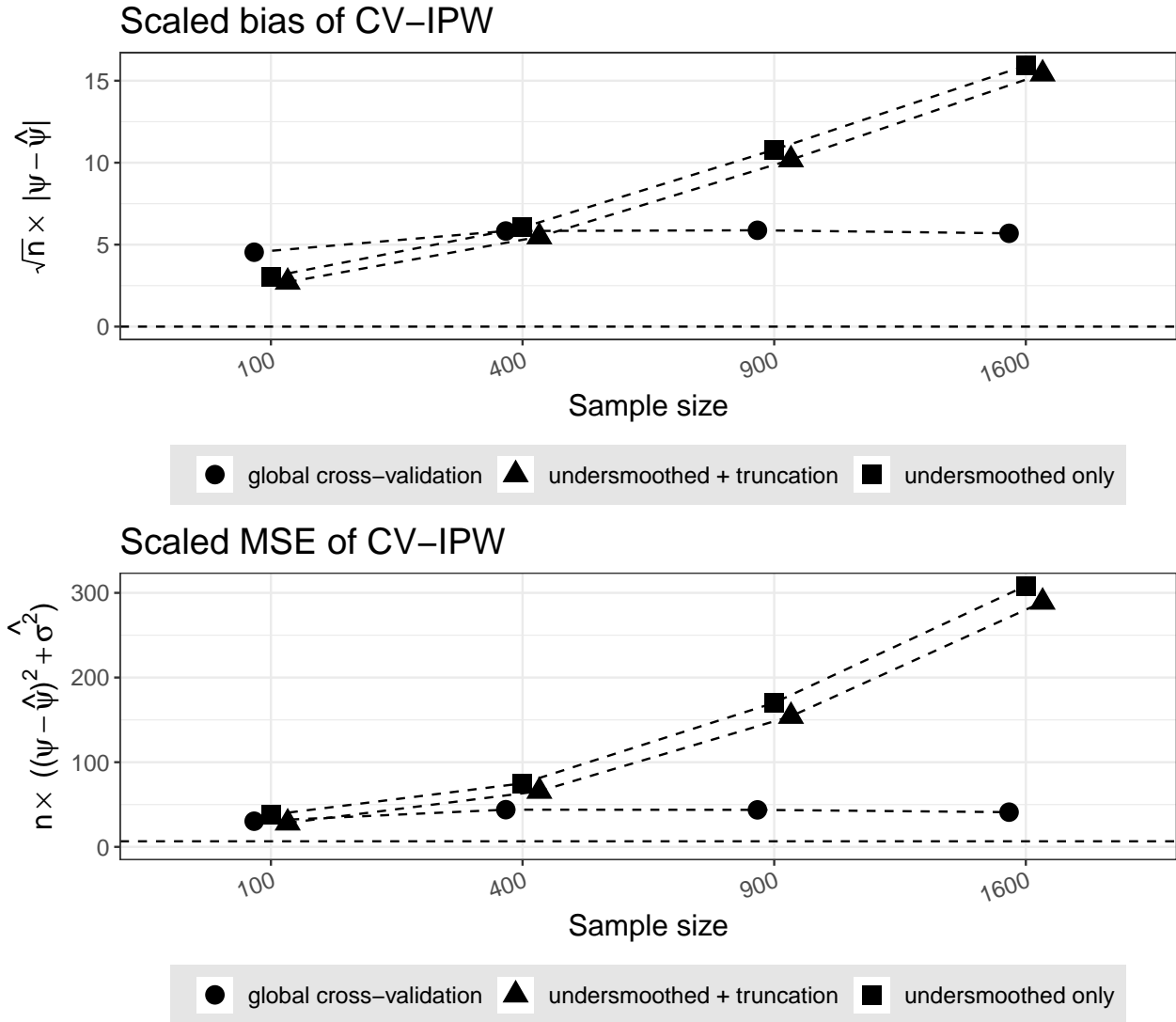| | Selected L1 norm — undersmooth wrt D_{CAR} | Ratios of L1 norms — undersmooth v gCV | Selected truncation — δ-truncation of g(a\|w) |

Experiment 2: Q_0 + **joint minimization**

- Horvitz-Thompson (nonstabilized) IPW estimator
- Joint minimization over $\delta \in \{0.01, 0.02, \ldots, 0.2\}$
- Maximum L1-norm multiplier: 25

The following table and plots summarize and visualize the performance of the estimator variants in terms of standard metrics:

n__obs	bias	n__bias	avg_var	mse	nmse	label
100	0.2715	2.7147	1.2195	0.2830	28.2986	undersmoothed + truncation
400	0.2741	5.4813	0.2340	0.1644	65.7535	undersmoothed + truncation
900	0.3394	10.1811	0.0853	0.1714	154.2363	undersmoothed + truncation
1600	0.3854	15.4162	0.0351	0.1807	289.1227	undersmoothed + truncation
100	0.3042	3.0419	2.0595	0.3791	37.9087	undersmoothed only
400	0.3023	6.0455	0.2484	0.1875	75.0193	undersmoothed only
900	0.3598	10.7930	0.0939	0.1891	170.1565	undersmoothed only
1600	0.3981	15.9255	0.0360	0.1927	308.3211	undersmoothed only
100	0.4537	4.5375	0.1591	0.3032	30.3182	global cross-validation

n_obs	bias	n_bias	avg_var	mse	nmse	label
400	0.2916	5.8317	0.0471	0.1099	43.9404	global cross-validation
900	0.1959	5.8758	0.0228	0.0486	43.7265	global cross-validation
1600	0.1423	5.6916	0.0136	0.0256	41.0054	global cross-validation

Scaled bias of CV–IPW

Scaled MSE of CV–IPW

We now visualize estimator performance using custom metrics for the estimator that does allow truncation:

Selected L1 norm
undersmooth wrt D_{CAR}

Ratios of L1 norms
undersmooth v gCV

Selected truncation
δ−truncation of g(a|w)