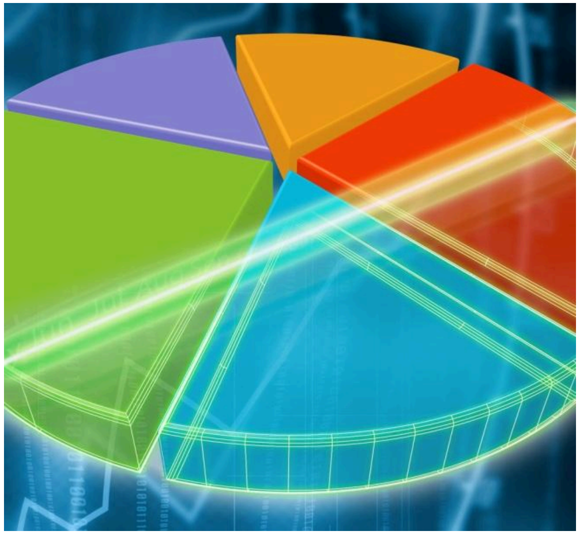
European Commission



Hiking Mount Toblerone: Advanced Methods for Random Balance Design

Elmar Plischke, Stefano Tarantola, Thierry Mara

Outline

- Random Balance Design (RBD)
- De-biasing
- Random and quasi-random permutations (QRP)
- RBD with QRP
- Case study

Background

Let us consider the input/output relation:

$$Y = f(X_1, X_2, \ldots, X_k)$$

where:

X_i are random variables

f is a numerical simulation model

Y is the model output

Basic set-up of RBD

1. Create a uniform sample $u \in [0,1]$ sampling from a periodic curve

$$u = \pi^{-1} \cos^{-1}\left(-\cos\left(\pi\omega(2s-1)\right)\right)$$

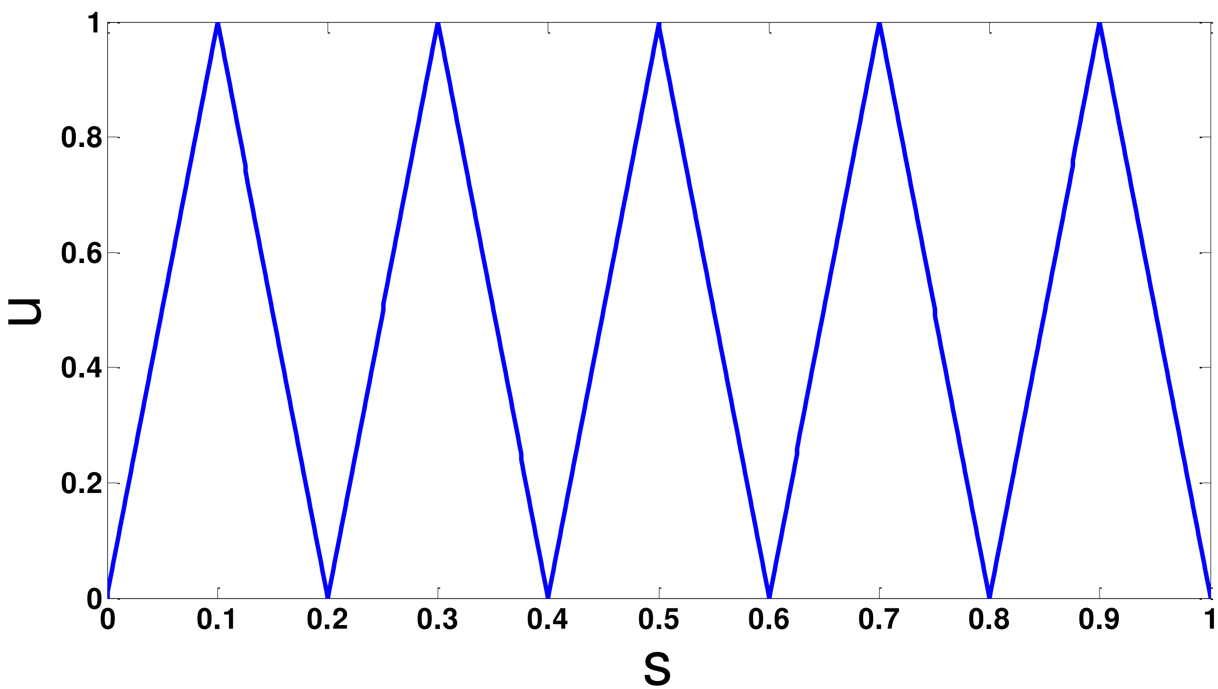
Where: $s \in [0,1]$

ω is an integer called basic frequency

Basic set-up of RBD

$$u = \pi^{-1} \cos^{-1}\left(-\cos\left(\pi\omega(2s-1)\right)\right)$$

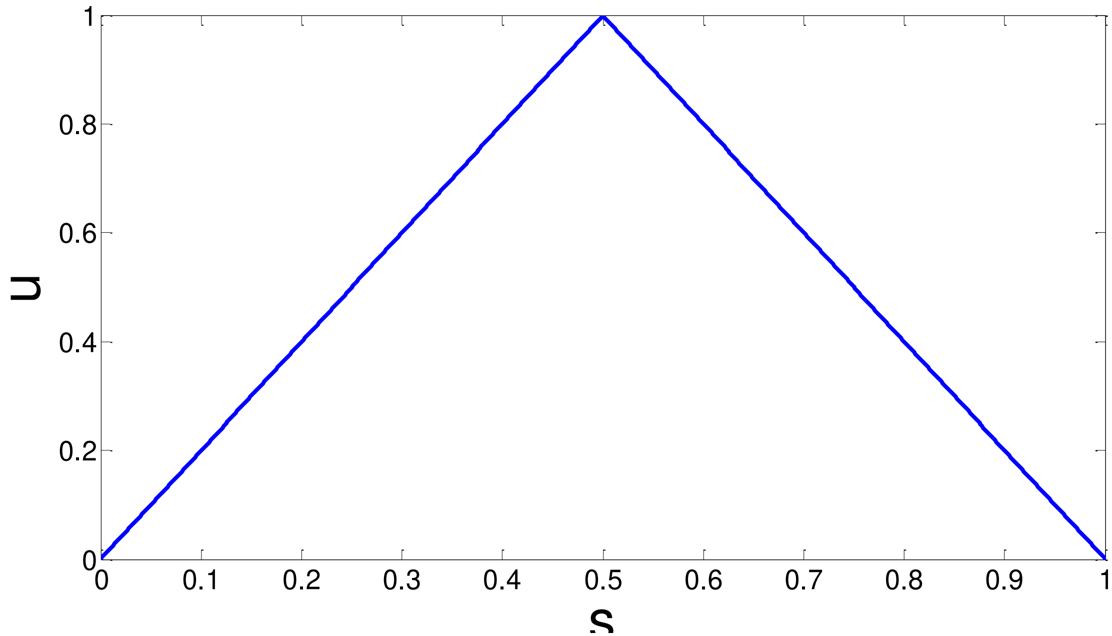
Example: ω = 5;



Basic set-up of RBD

Most common case ω = 1:

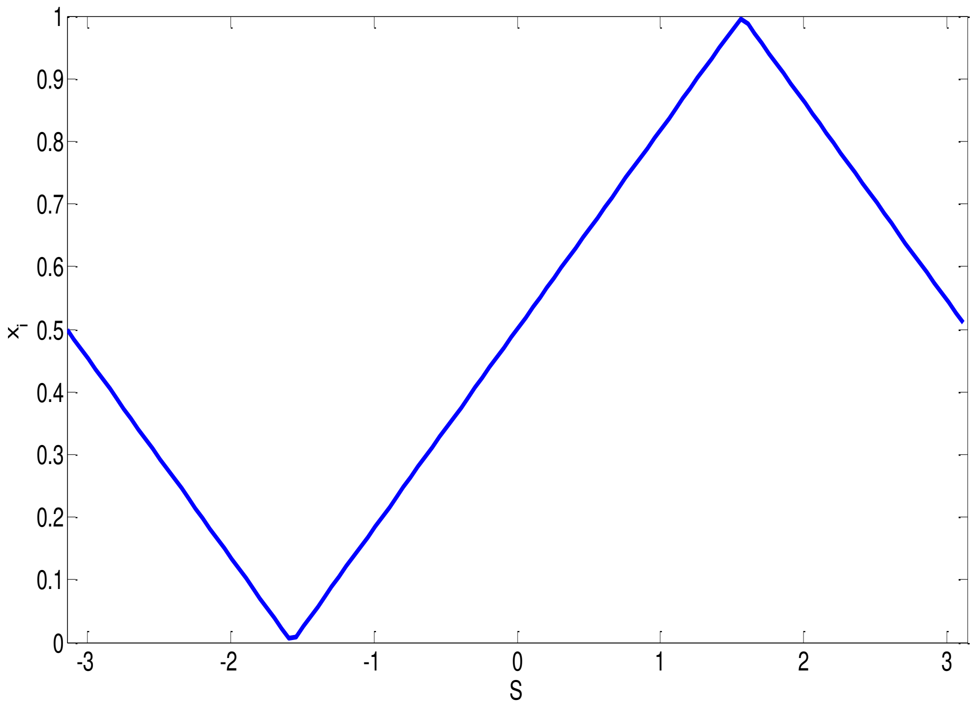
$$u = 1 - |2s - 1| \qquad s \in [0,1]$$



Basic set-up of RBD

In the original formulation:

$$u_i(s) = \frac{1}{2} + \frac{1}{\pi}\arcsin\left(\sin \omega_i s\right) \qquad s \in (-\pi; \pi) \;\; \omega_i = 1 \;\; i = 1,2,\ldots,k$$



Basic set-up of RBD

$$Y = f(X_1, X_2) = 3X_1 + 5X_2 \qquad (X_1, X_2) \in [0;1]^2 \qquad s \in (-\pi; \pi)$$

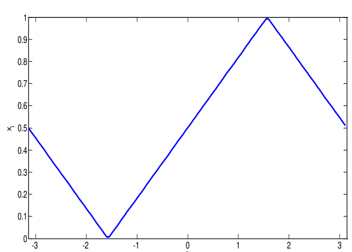
1) Sample N times from –π to π equidistantly (vector s) (ex N=8)

2) Feed *s* through the periodic curve: $u(s) = \frac{1}{2} + \frac{1}{\pi}\arcsin(\sin \omega s)$

3) Generate k indep. random permutations of u: $p_i(u) = u_i$

4) Transform u_i into x_i (not needed if iid U[0,1])

5) Evaluate model output y $y = f(x_1, x_2)$



s	u	x1 u1	x2 u2	y
-2.7489	0.3750	0.625	0.875	6.25
-1.9635	0.1250	0.125	0.625	3.50
-1.1781	0.1250	0.875	0.125	3.25
-0.3927	0.3750	0.375	0.125	1.75
0.3927	0.6250	0.125	0.875	4.75
1.1781	0.8750	0.875	0.375	4.50
1.9635	0.8750	0.625	0.375	3.75
2.7489	0.6250	0.375	0.625	4.25

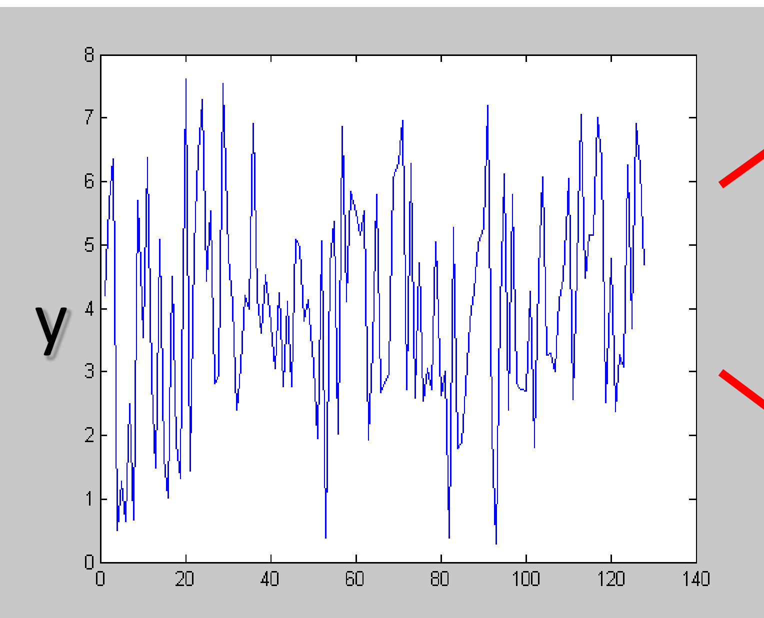
f

Basic set-up of RBD

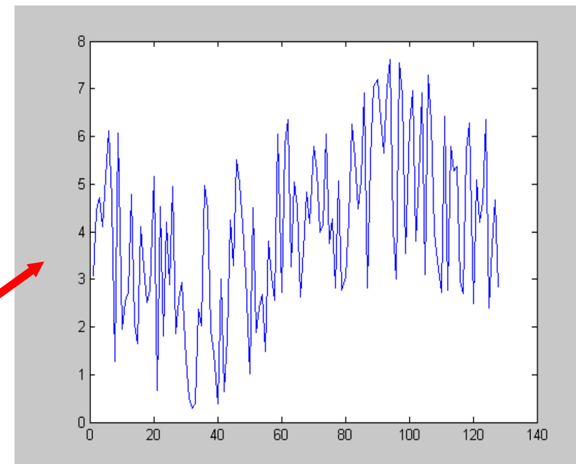
- Re-order y by applying inverse permutations p_1^{-1} and p_2^{-1}

Example with N=128

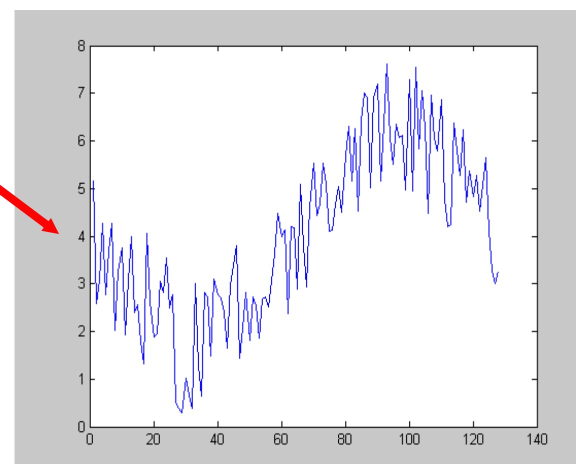
y



N pts



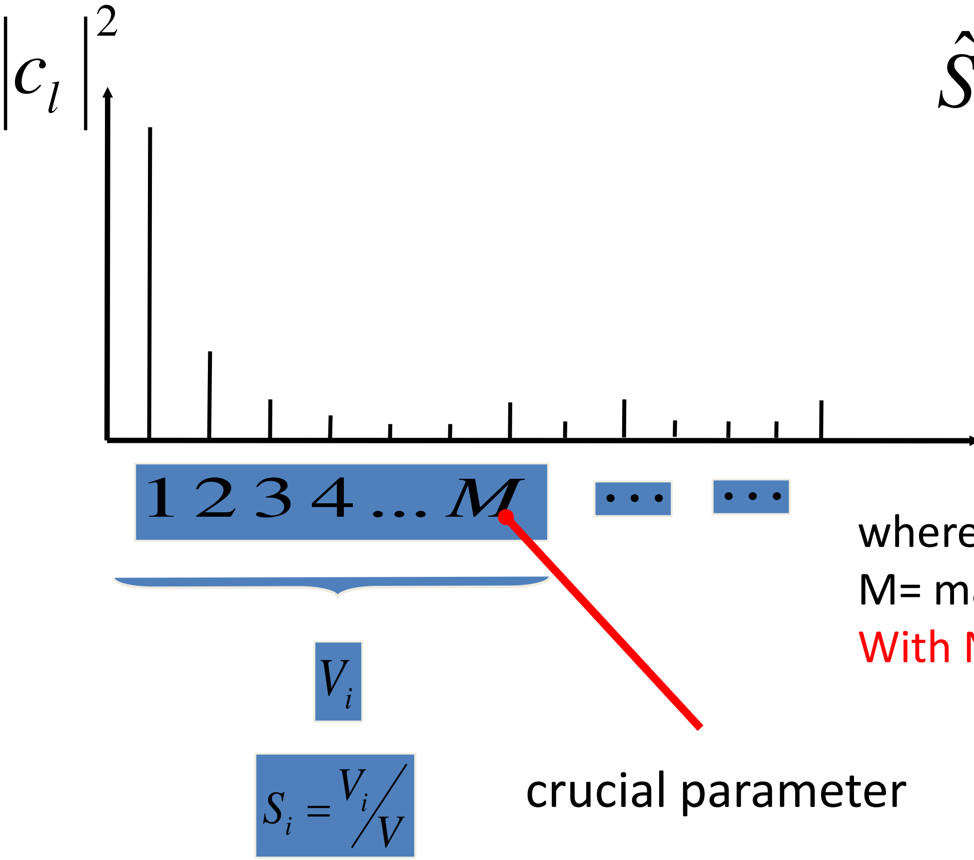
y after p_1^{-1}



y after p_2^{-1}

Basic set-up of RBD

- Looking for resonances of re-ordered y at low frequencies



$$\hat{S}_i = \frac{\sum_{l=1}^{M} |c_l|^2}{\sum_{l=1}^{N} |c_l|^2}, \qquad i = 1,2,\ldots,k$$

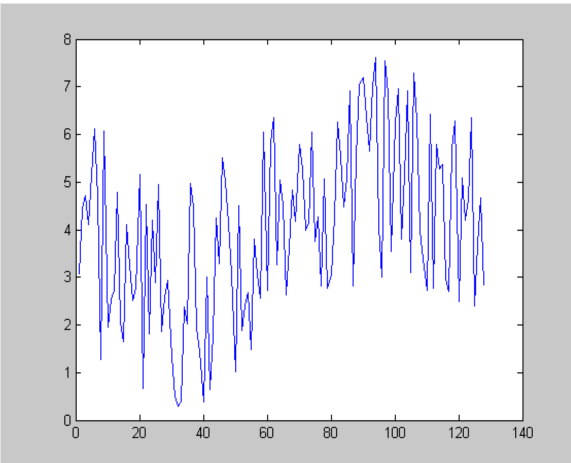
where c_l are the Fourier coefficients of reordered y

M= max n. of harmonics (usually 4 or 6)

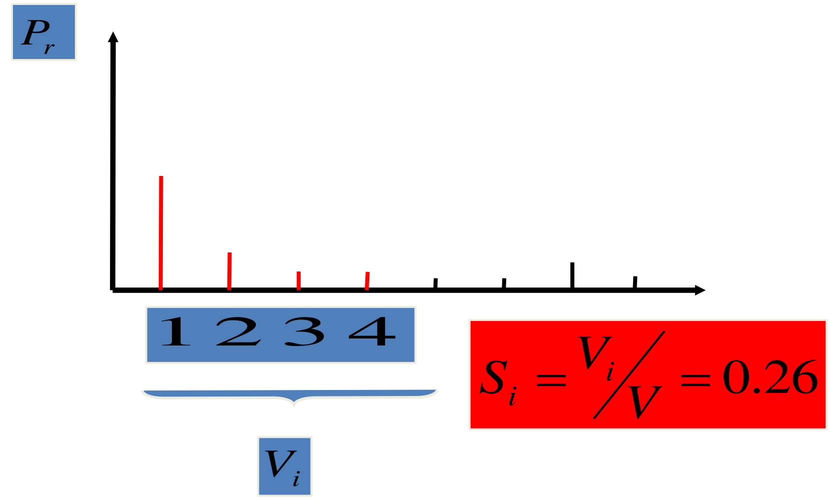
With N points all $\hat{S}_i$ can be computed

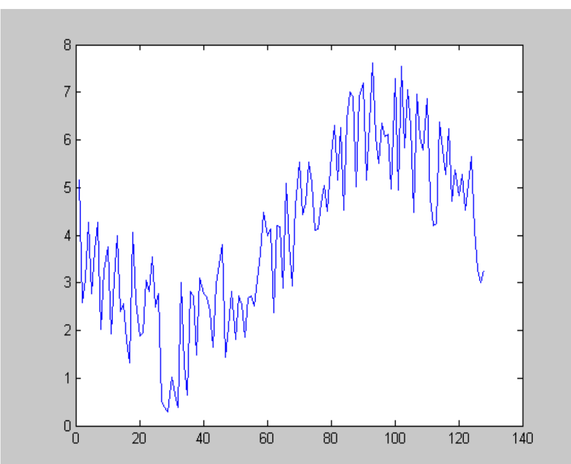
Basic set-up of RBD

$$y = f(x_1, x_2) = 3x_1 + 5x_2$$

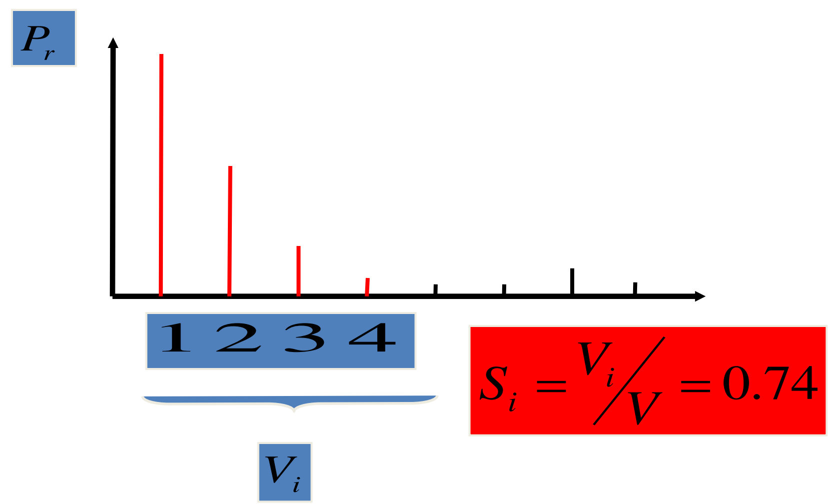


y re-ordered
wrt s_1





y re-ordered
wrt s_2



Limitations of RBD

1. Sensitivity indices (especially small indices) are biased with respect to analytical values

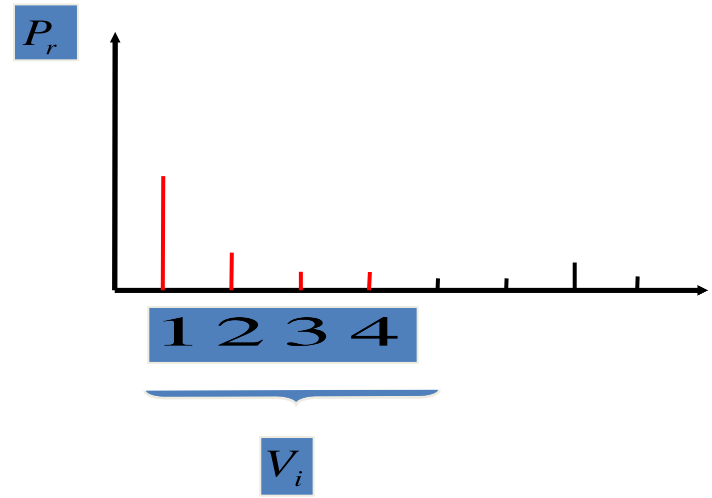
2. Sample design in use does not necessarily cover the sample space uniformly. Estimates are affected by large random error

Improvements of RBD can be achieved by controlling these two drawbacks.

Bias in RBD

The factors X_{-i} are randomly sampled.

The remaining part of variance V_{-i} appears at all frequencies as random noise. A fraction of this noise overlaps to the signal at the lower harmonics.



Tissot and Prieur (2012) propose a bias correction formula based on the assumption that the unexplained variance V_{-i} has a white noise

$$S_i^{DB} = \frac{N\hat{S}_i - 2M}{N - 2M}$$

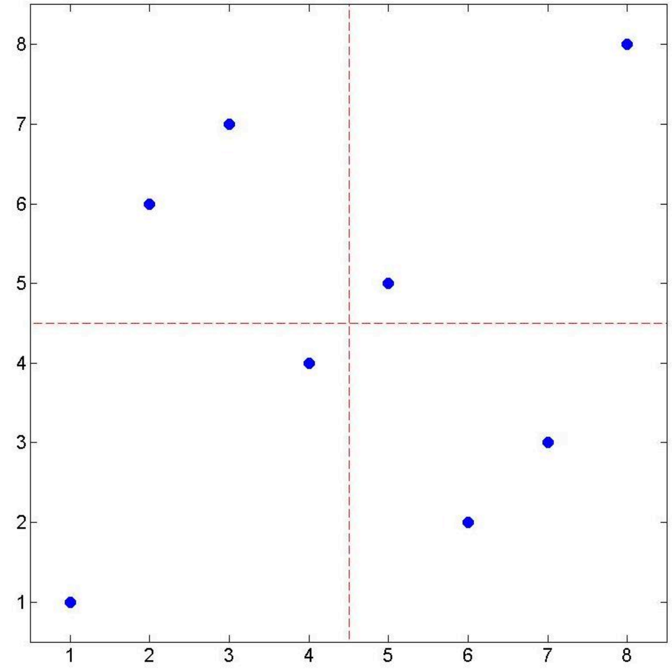
We will see a case study later

Large random error

2. Sample design in use does not necessarily cover the sample space with good uniform properties

Instead of random permutations, use permutations obtained from low-discrepancy sequences

s	p1	p2
-2.7489	1	1
-1.9635	5	5
-1.1781	3	7
-0.3927	7	3
0.3927	2	6
1.1781	6	2
1.9635	4	4
2.7489	8	8



u1	u2
-2.7489	-2.7489
0.3927	0.3927
-1.1781	1.9635
1.9635	-1.1781
-1.9635	1.1781
1.1781	-1.9635
-0.3927	-0.3927
2.7489	2.7489

quasi-random balance design

Quasi-Random Balance Design

$$y = \sum_{i=1}^{6} X_i \quad X_i \sim U[0,1]$$

We tested available Sobol' sequence generators:

GNU scientific Library
Numerical Recipes
MatLab
Joe and Kuo
Broda Ltd.

$$N = 512$$

Quasi-Random Source	S_1	S_2	S_3	S_4	S_5	S_6
Analytical Values	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667
GSL	0.1621	0.1632	0.1623	0.1627	0.1624	0.1628
NR	0.1627	0.1664	0.1614	0.1618	0.1867	0.1765
MatLab	0.1689	0.1683	0.1679	0.1677	0.1663	0.1657
Joe&Kuo	0.1630	0.1622	0.1624	0.1623	0.1624	0.1632
Broda	0.1640	0.1626	0.1621	0.1644	0.1623	0.1622
Simple Random	0.1760	0.2137	0.2622	0.1842	0.1860	0.1557
Simple Random	0.1832	0.2022	0.1595	0.1873	0.2164	0.1846

Test Case

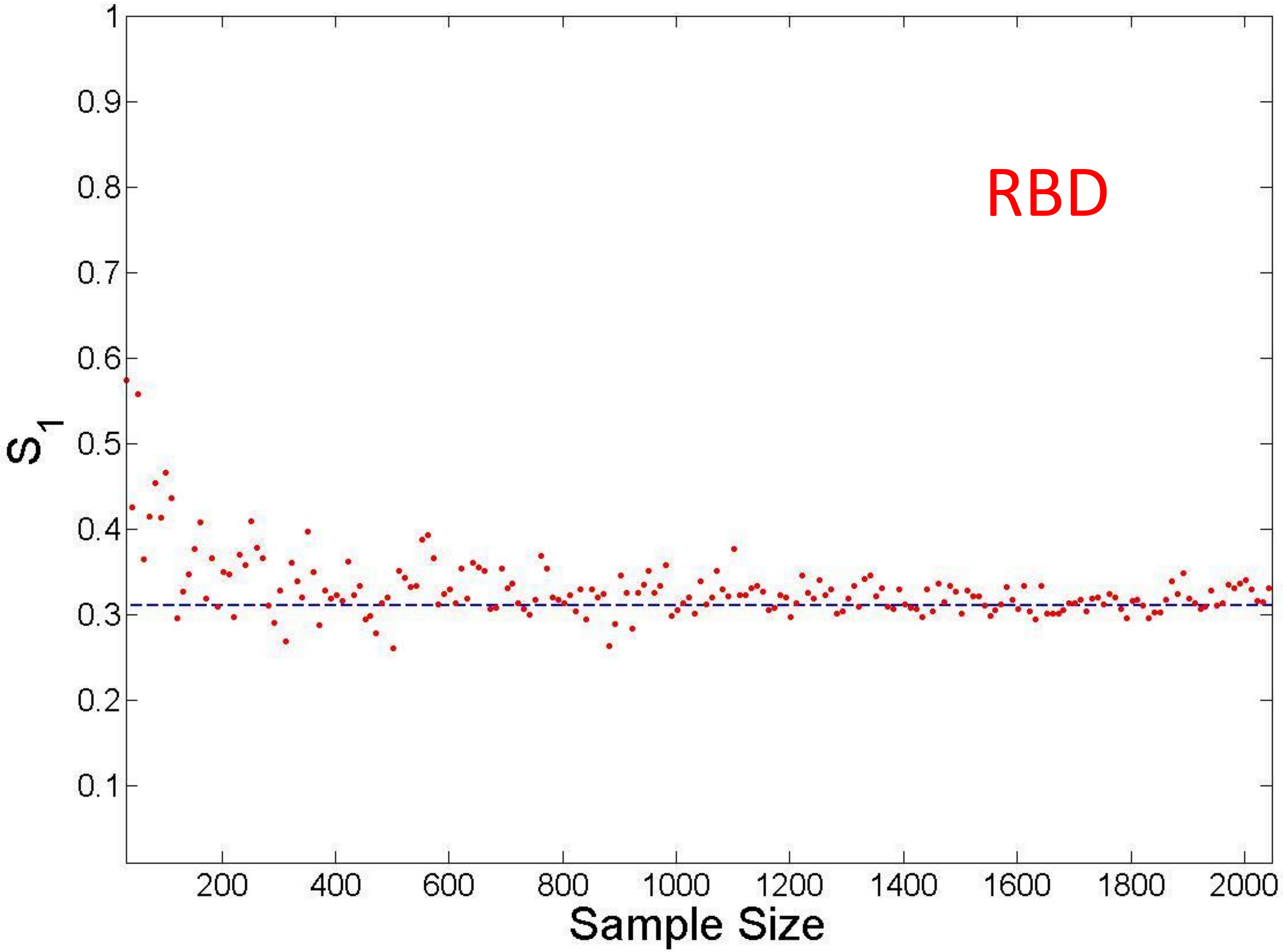
Ishigami test function

$$y = \sin X_1 + 7 \sin^2 X_2 + 0.1 X_3^4 \sin X_1 \qquad X_i \sim U[-\pi, \pi]$$

Analytic main effects:

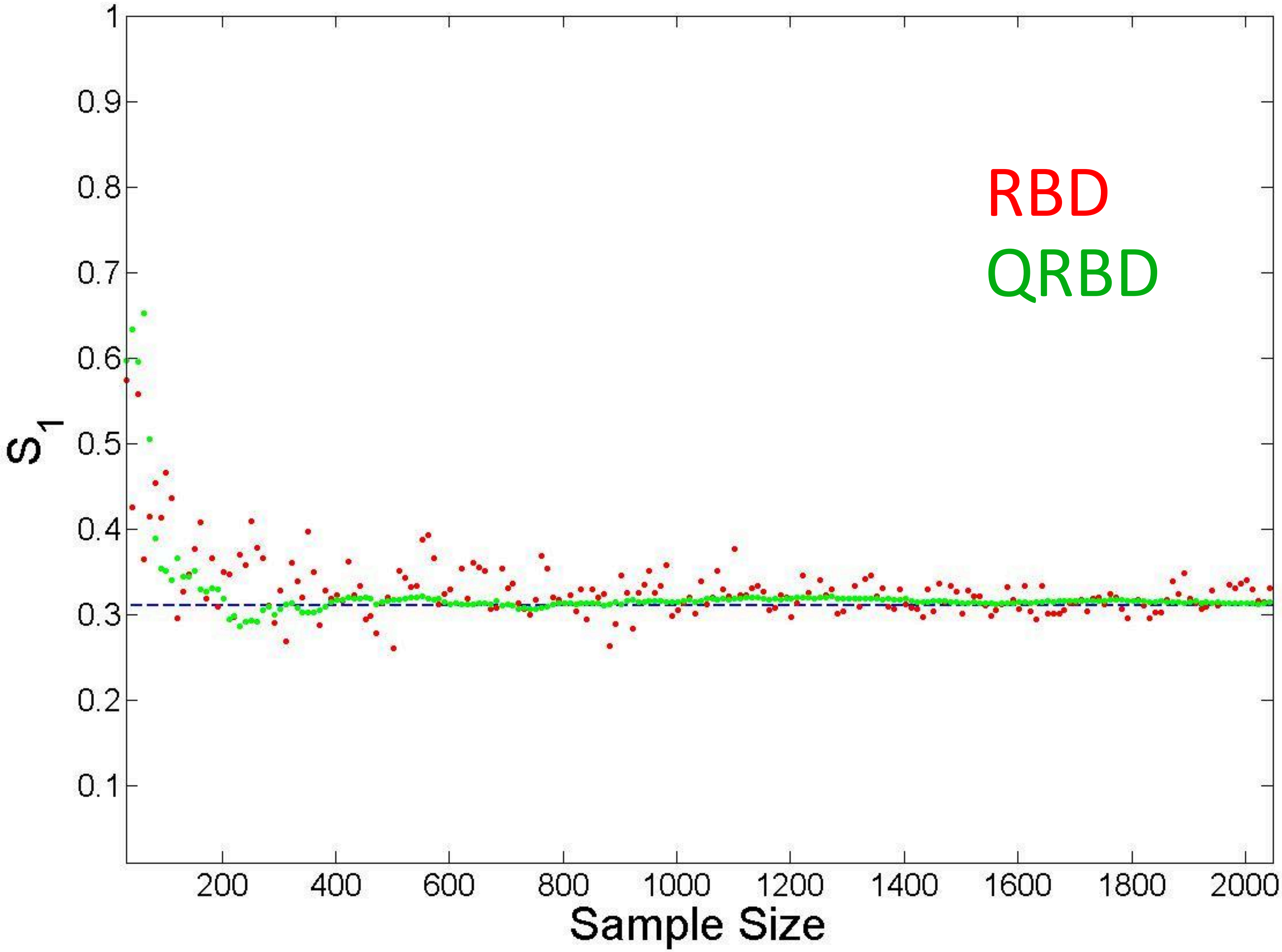
$$S_1 = 0.3138$$
$$S_2 = 0.4424$$
$$S_3 = 0$$

X_1



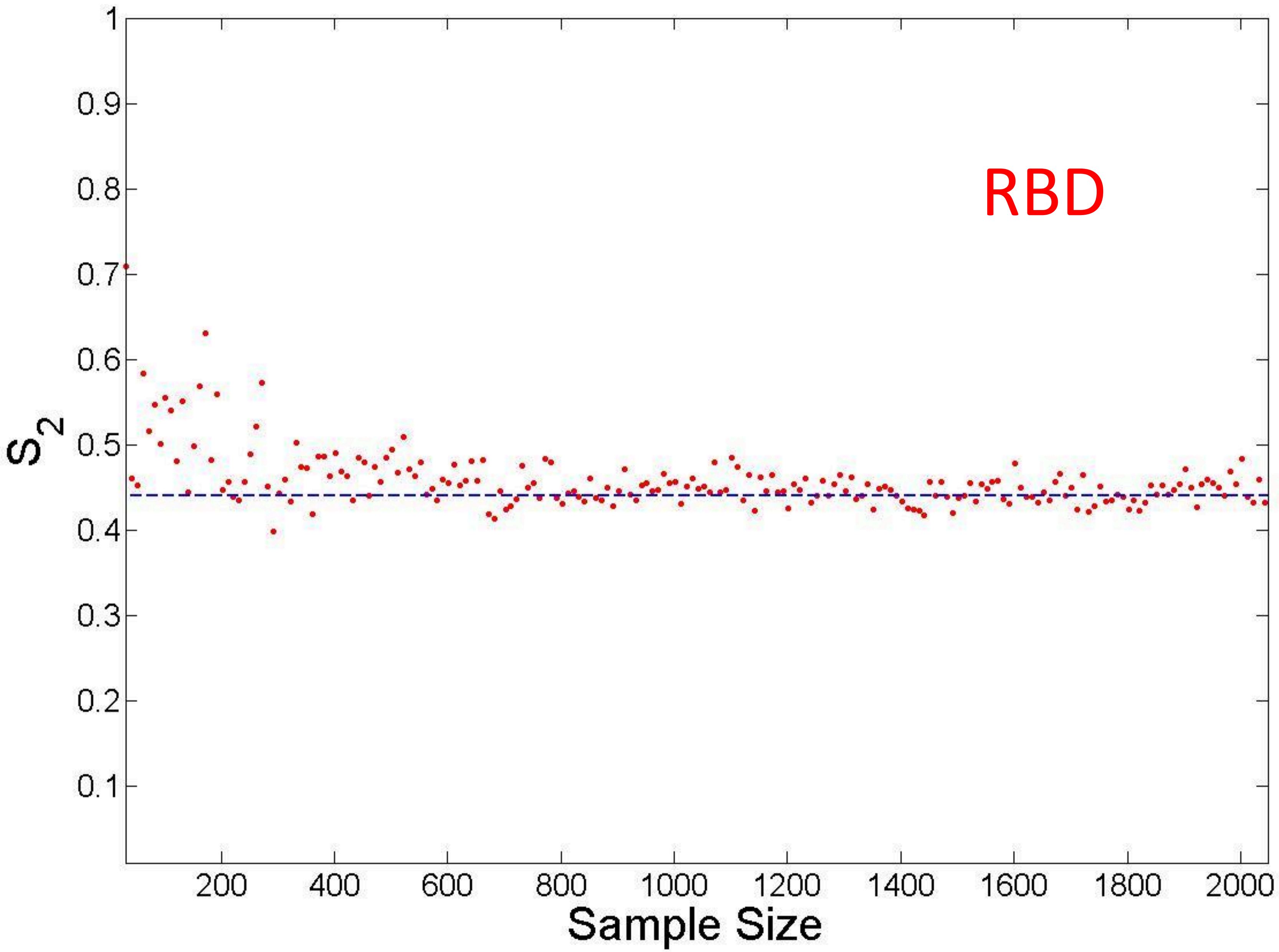
RBD
S1
Sample Size
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
200
400
600
800
1000
1200
1400
1600
1800
2000

X_1



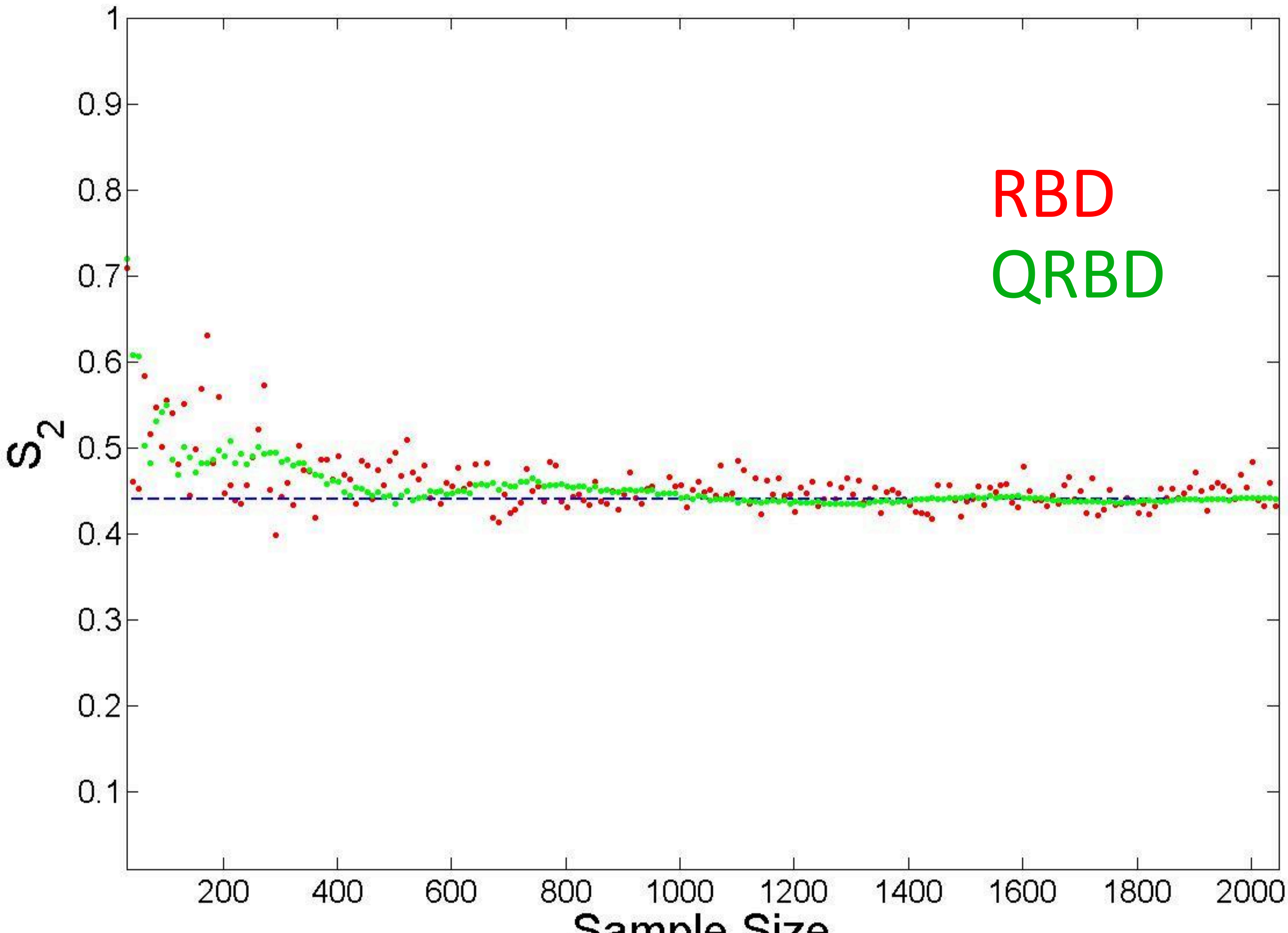
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
S_1
200
400
600
800
1000
1200
1400
1600
1800
2000
Sample Size
RBD
QRBD

X_2



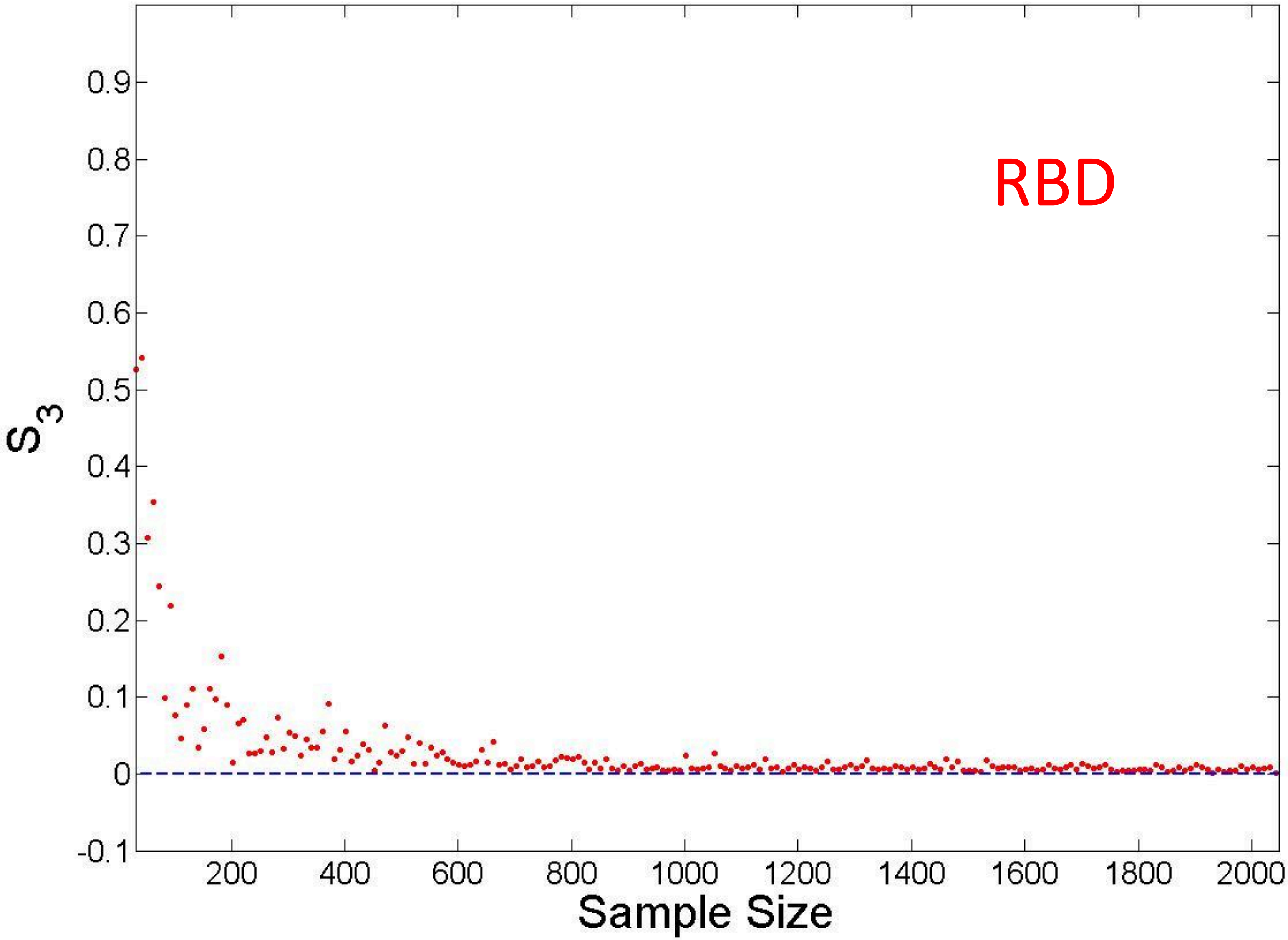
RBD
S2
Sample Size
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
200
400
600
800
1000
1200
1400
1600
1800
2000

X_2

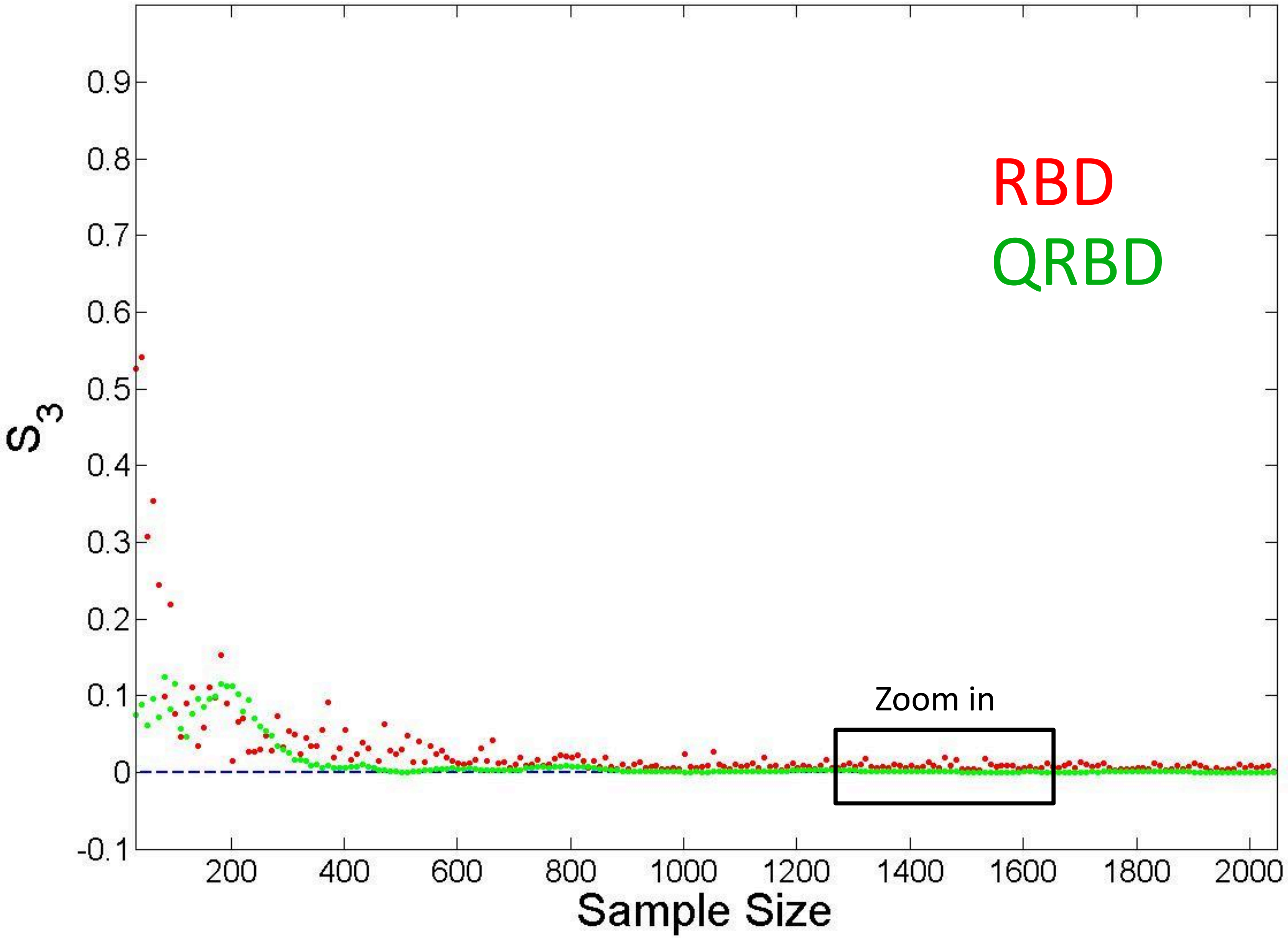


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
S2
200
400
600
800
1000
1200
1400
1600
1800
2000
Sample Size
RBD
QRBD

X_3

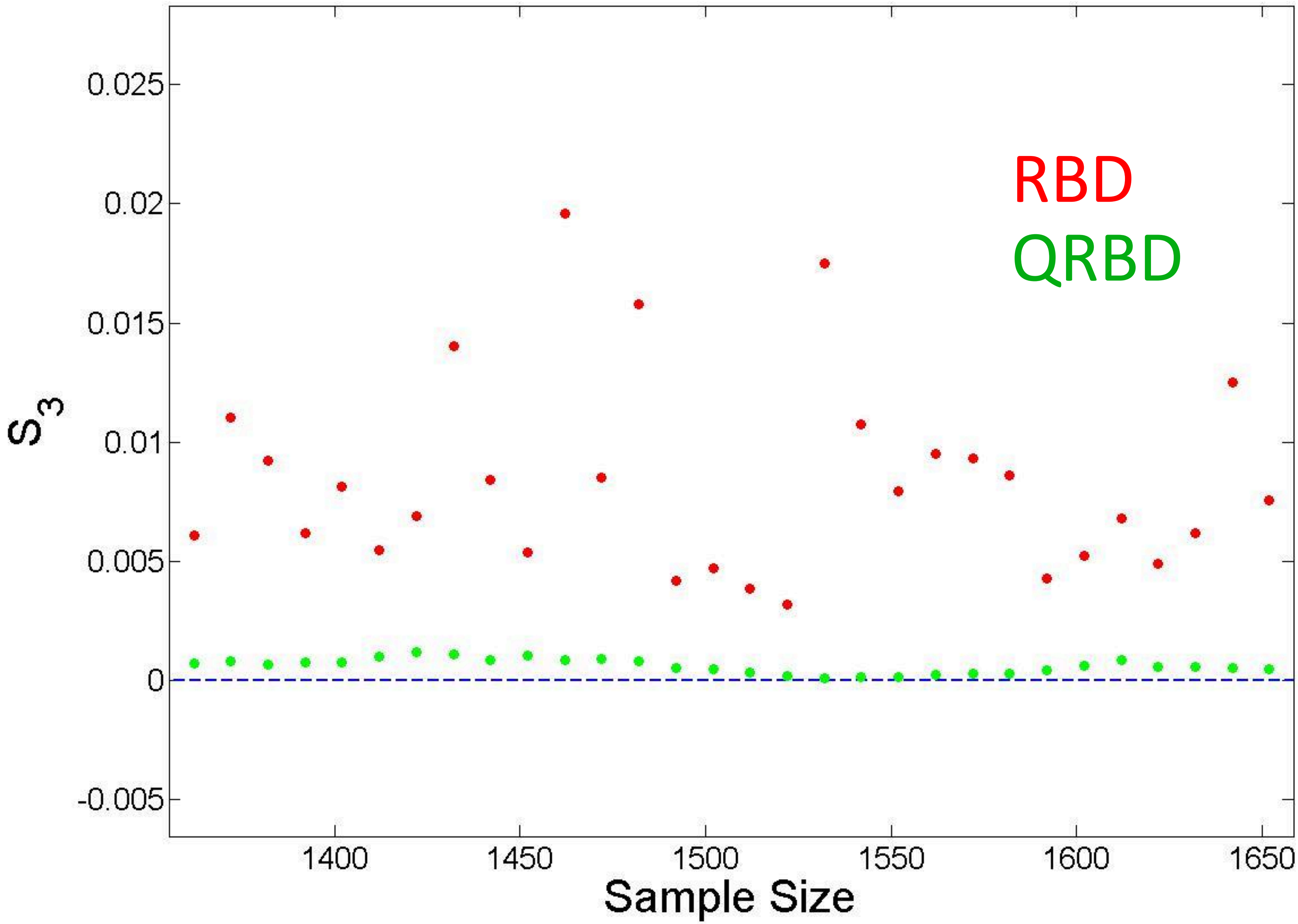

RBD
S_3
Sample Size
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
200
400
600
800
1000
1200
1400
1600
1800
2000

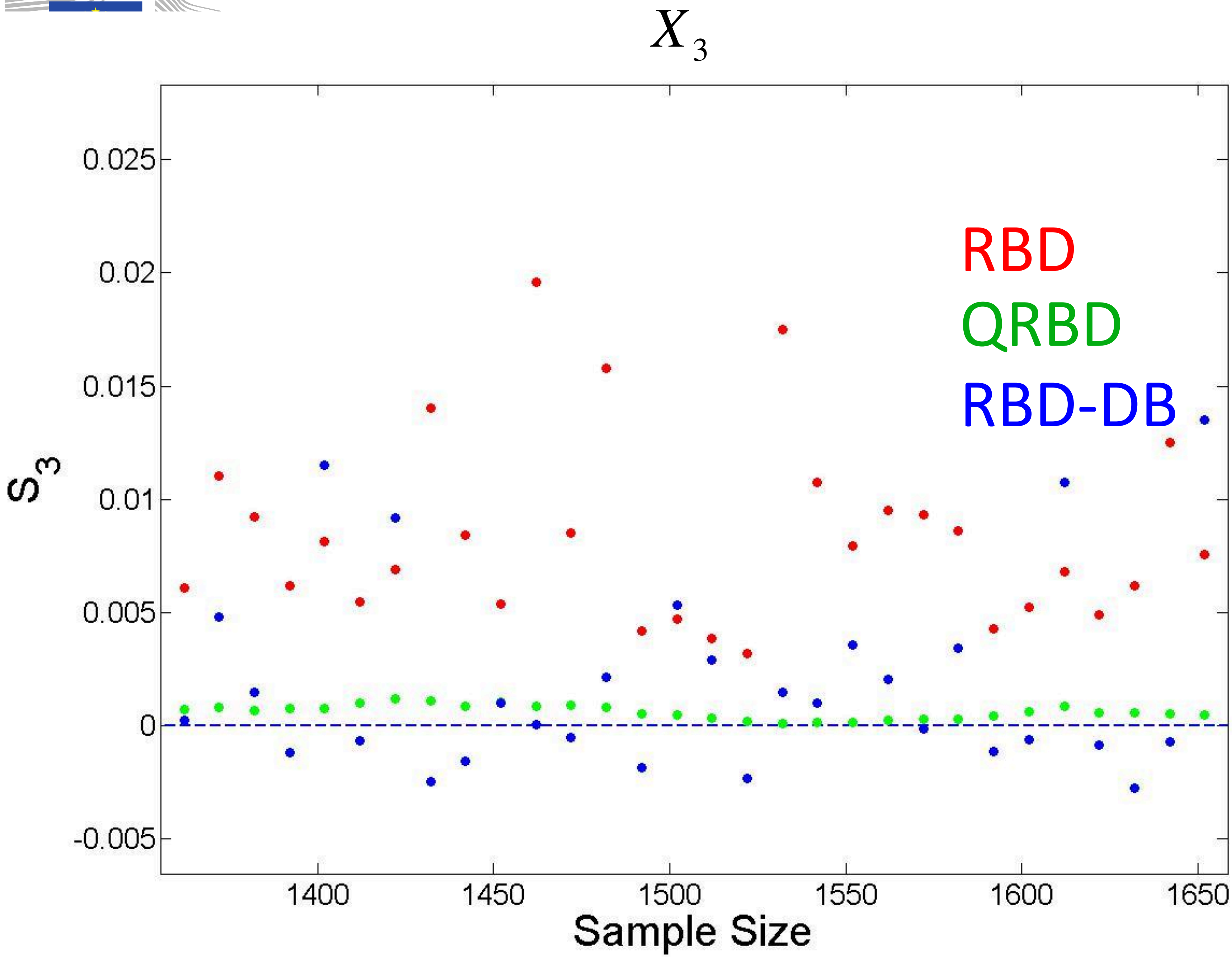
X_3



RBD
QRBD
Zoom in
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
S_3
200
400
600
800
1000
1200
1400
1600
1800
2000
Sample Size

X_3


0.025
0.02
0.015
0.01
0.005
0
-0.005
S_3
1400
1450
1500
1550
1600
1650
Sample Size
RBD
QRBD



X_3
RBD
QRBD
RBD-DB
S_3
0.025
0.02
0.015
0.01
0.005
0
-0.005
1400
1450
1500
1550
1600
1650
Sample Size

Second Test Case: a discontinuous function

$$y = \sum_i X_i - \frac{\gamma}{2},$$

$$\gamma = 1 \quad if \quad \exists x_i : 0 < x_i < 1/2$$

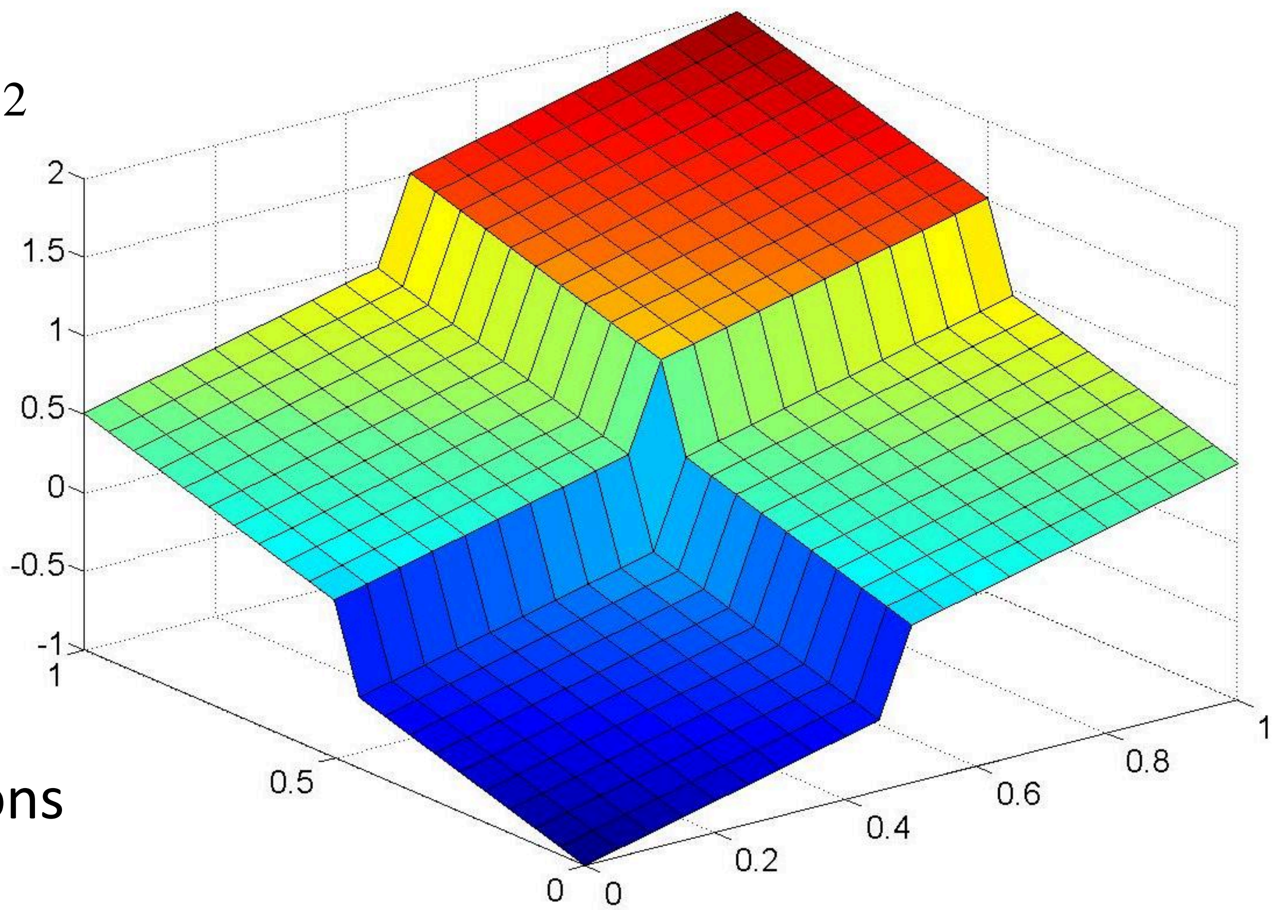
$$\gamma = 2 \quad if \quad \exists x_{i,j} : 0 < x_{i,j} < 1/2$$

$$\gamma = 3 \quad if \quad \exists x_{i,j,l} : 0 < x_{i,j,l} < 1/2$$

...

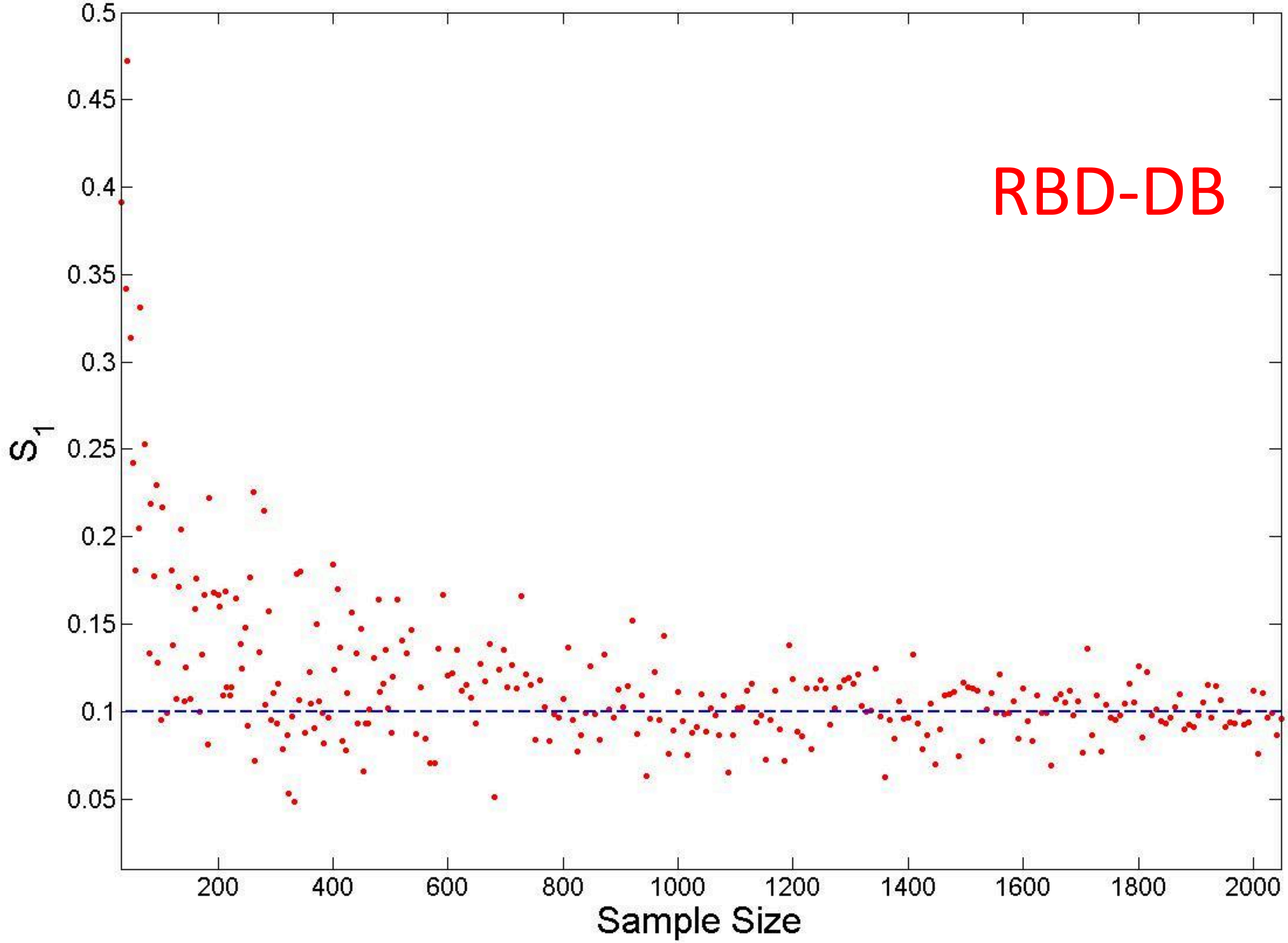
Matlab function:

```
model=@(x)sum(x,2)-sum(x<.5,2)/2;
```



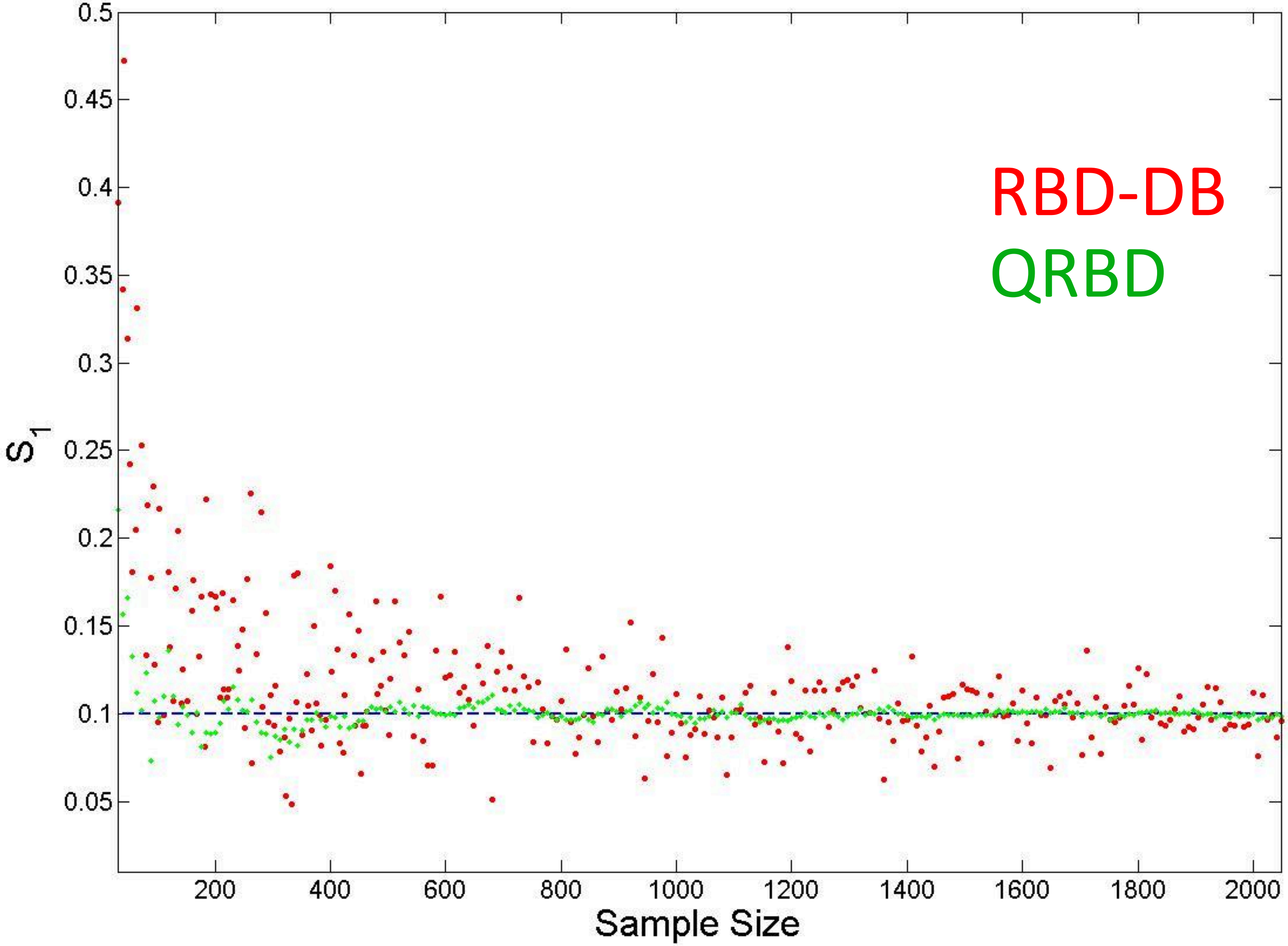
But in ten dimensions

X_1



RBD-DB
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
S_1
200
400
600
800
1000
1200
1400
1600
1800
2000
Sample Size

X_1



0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
S_1
200
400
600
800
1000
1200
1400
1600
1800
2000
Sample Size
RBD-DB
QRBD

Conclusions

QRBD is considerably superior to RBD and RBD-(de biased)

Much less random error and much better convergence

QRBD does not need to be debiased further

Plischke, Tarantola , Mara
Advanced Sensitivity Methods using Random Balance Design
In preparation

Thank you!