Adaptive Dantzig density estimation

Erwan Le Pennec
LPMA, Univ. Paris Diderot / SELECT, INRIA Saclay - Univ. Paris Sud
with Karine Bertin (Univ. de Valparaiso, Chile)
and Vincent Rivoirard (Univ. Paris Sud - ENS Paris)

Oktober 2009

Theorem

Theorem 1. *Let $s \leq p/2$ and $l \geq s$ two integers such that either*

1. *we have $s + l \leq p$ and $l\phi_{\min}(s+l) > s\phi_{\max}(l)$ and set*
$$\kappa_{s,l} = \sqrt{\phi_{\min}(s+l)}\left(1 - \sqrt{\frac{s\phi_{\max}(l)}{l\phi_{\min}(s+l)}}\right) > 0, \quad \mu_{s,l} = \sqrt{\frac{\phi_{\max}(l)}{l}}.$$
2. *we have $\phi_{\min}(2s) > \theta_{s,2s}$ and set*
$$\kappa_{s,l} = \sqrt{\phi_{\min}(2s)}\left(1 - \frac{\theta_{s,2s}}{\phi_{\min}(2s)}\right) > 0, \quad \mu_{s,l} = \frac{\theta_{s,2s}}{\sqrt{s\phi_{\min}(2s)}}.$$

Then, with probability at least $1 - C_1(\varepsilon, \delta, \gamma')p^{1-\frac{\gamma'}{1+\varepsilon}}$, we have for any $\alpha > 0$,

$$\|\hat{f}^D - f_0\|_2^2 \leq \inf_{\lambda \in \mathbb{R}^p} \inf_{\substack{J_0 \subset \{1,\dots,p\} \\ |J_0| = s}} \left\{ \|f_\lambda - f_0\|_2^2 + \alpha \frac{\Lambda(\lambda, J_0^c)^2}{s}\left(1 + \frac{2\mu_{s,l}\sqrt{s}}{\kappa_{s,l}}\right)^2 \right.$$
$$\left. + 4s\left(\frac{1}{\alpha} + \frac{1}{\kappa_{s,l}^2}\right)\left(\|\eta_{\gamma'}\|_{\ell_\infty} + \|\eta_\gamma\|_{\ell_\infty}\right)^2 \right\}$$

where $\Lambda(\lambda, J_0^c) = \|\lambda_{J_0^c}\|_{\ell_1} + \frac{\left(\|\hat{\lambda}^D\|_{\ell_1} - \|\lambda\|_{\ell_1}\right)_+}{2}$.

Local Theorem

Theorem 2. *Let $J_0 \subset \{1,\dots,p\}$ be fixed. Suppose that for some constants $\kappa_{J_0} > 0$ and $\mu_{J_0} > 0$ depending on J_0, we have for any λ,*

$$\|f_\lambda\|_2 \geq \kappa_{J_0}\|\lambda_{J_0}\|_{\ell_2} - \mu_{J_0}\left(\|\lambda_{J_0^C}\|_{\ell_1} - \|\lambda_{J_0}\|_{\ell_1}\right)_+ . \qquad (LA(J_0,\kappa_{J_0},\mu_{J_0}))$$

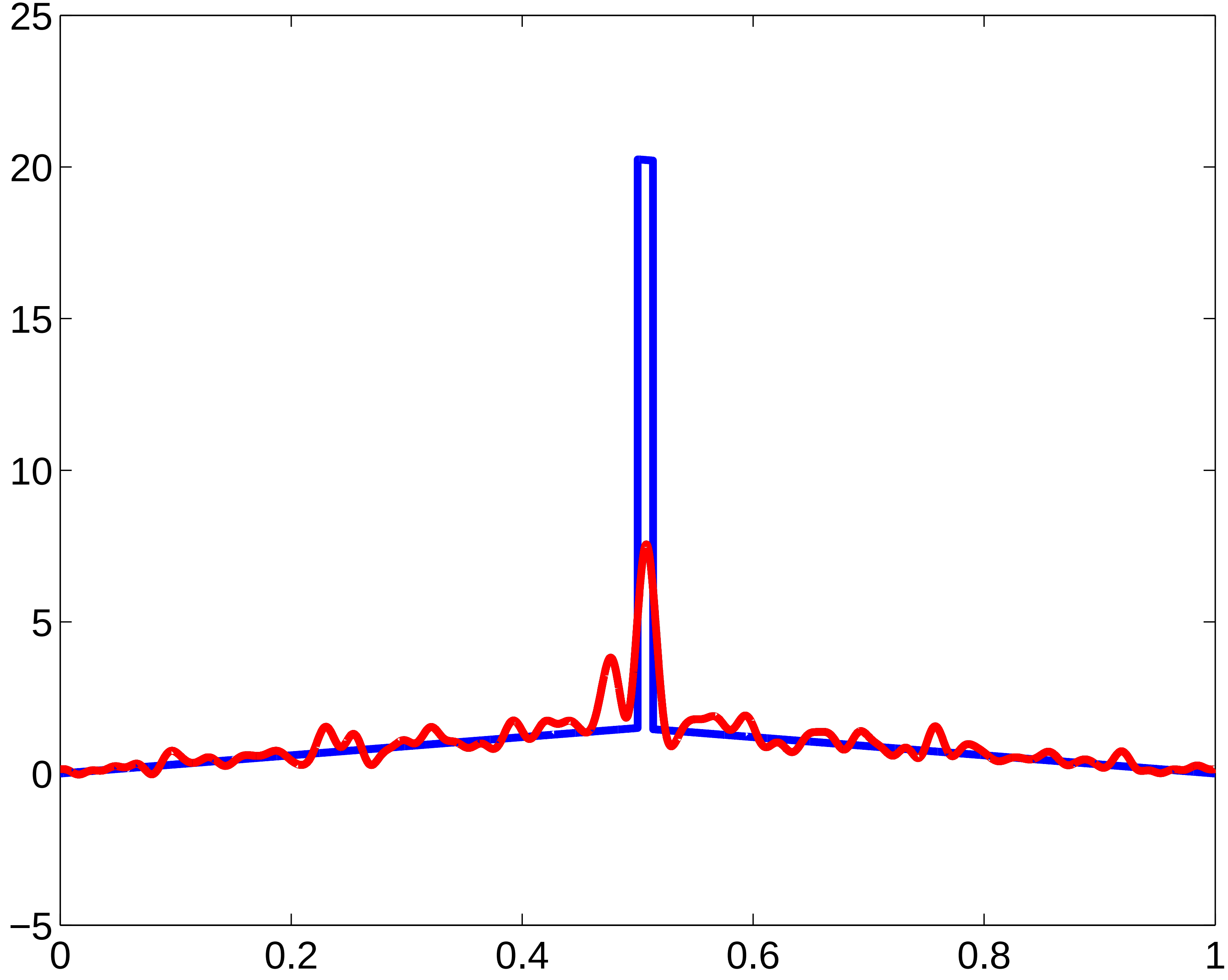
Then, with probability at least $1 - C_1(\varepsilon,\delta,\gamma')p^{1-\frac{\gamma'}{1+\varepsilon}}$, we have for any $\alpha > 0$,

$$\|\hat{f}^D - f_0\|_2^2 \leq \inf_{\lambda\in\mathbb{R}^p}\Bigg\{\|f_\lambda - f_0\|_2^2 + \alpha\frac{\Lambda(\lambda,J_0^c)^2}{|J_0|}\left(1 + \frac{2\mu_{J_0}\sqrt{|J_0|}}{\kappa_{J_0}}\right)^2 + 4|J_0|\left(\frac{1}{\alpha} + \frac{1}{\kappa_{J_0}^2}\right)\left(\|\eta_{\gamma'}\|_{\ell_\infty} + \|\eta_\gamma\|_{\ell_\infty}\right)^2\Bigg\},$$

with

$$\Lambda(\lambda,J_0^c) = \|\lambda_{J_0^C}\|_{\ell_1} + \frac{\left(\|\hat{\lambda}^D\|_{\ell_1} - \|\lambda\|_{\ell_1}\right)_+}{2}.$$

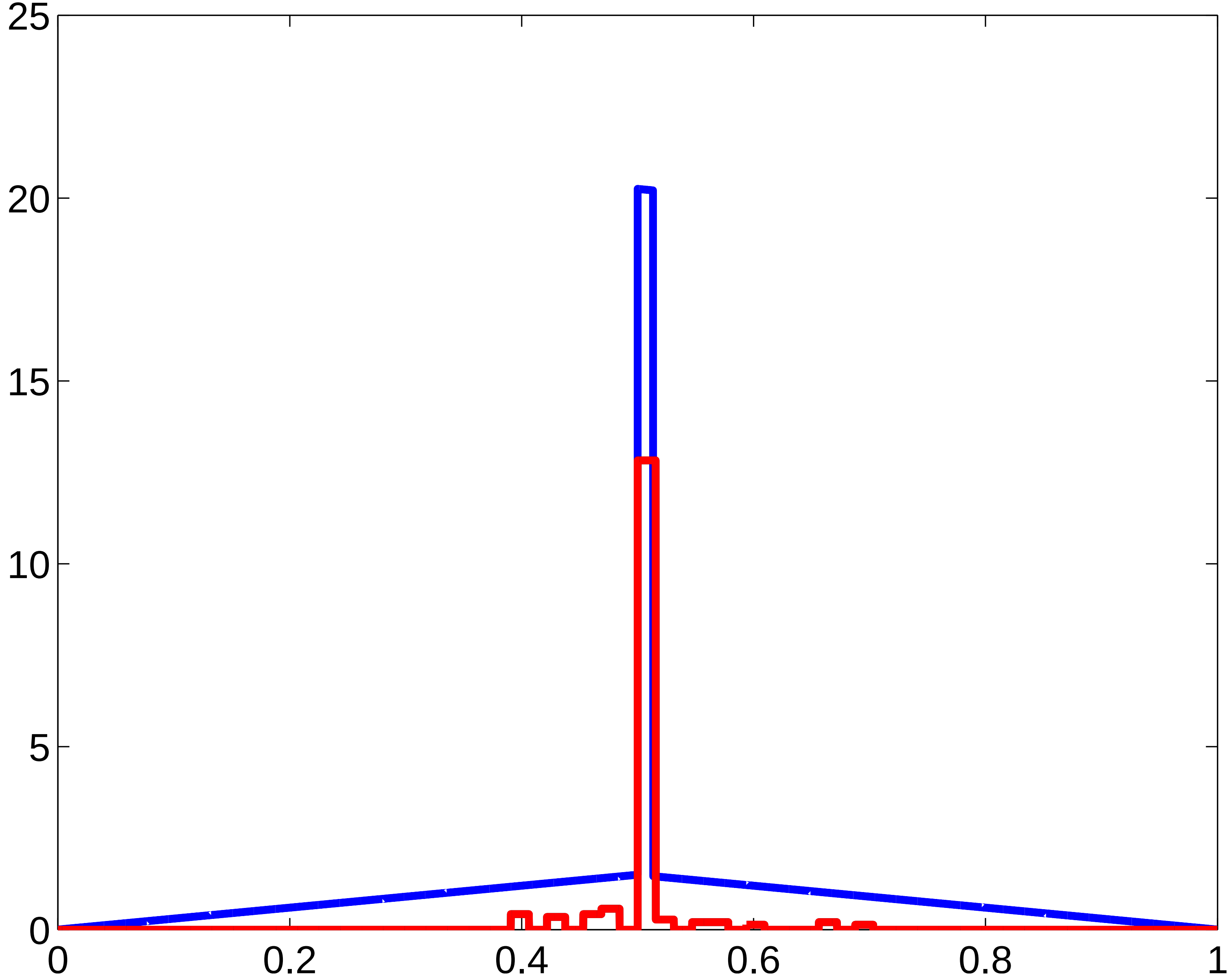
Fourier



$\|f - \hat{f}\|^2 = 2.16$

$\|\hat{\lambda}\|_0 = 67$

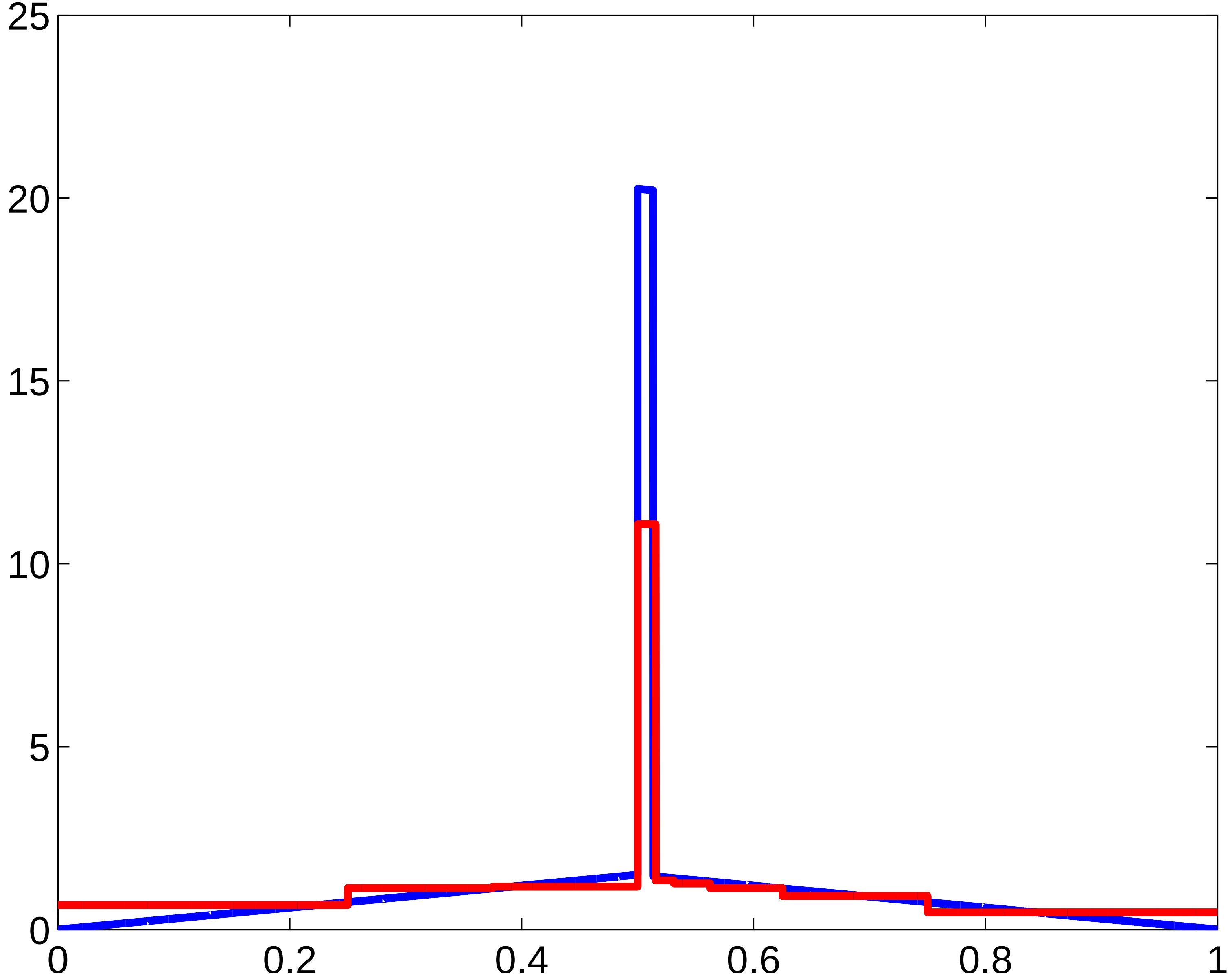
Boxes



$\|f - \hat{f}\|^2 = 1.51$

$\|\hat{\lambda}\|_0 = 7$

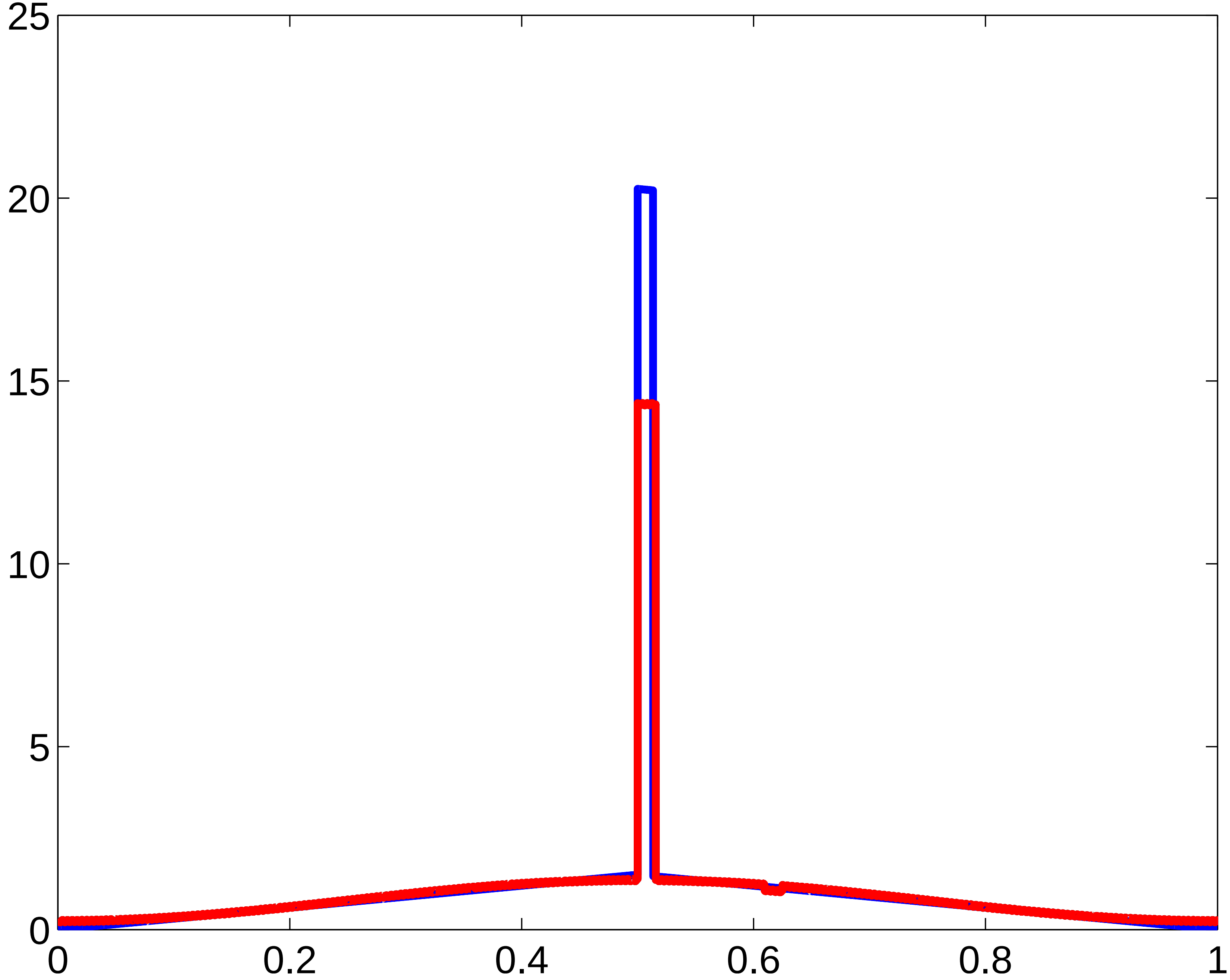
Haar



$$\|f - \hat{f}\|^2 = 1.09$$

$$\|\hat{\lambda}\|_0 = 8$$

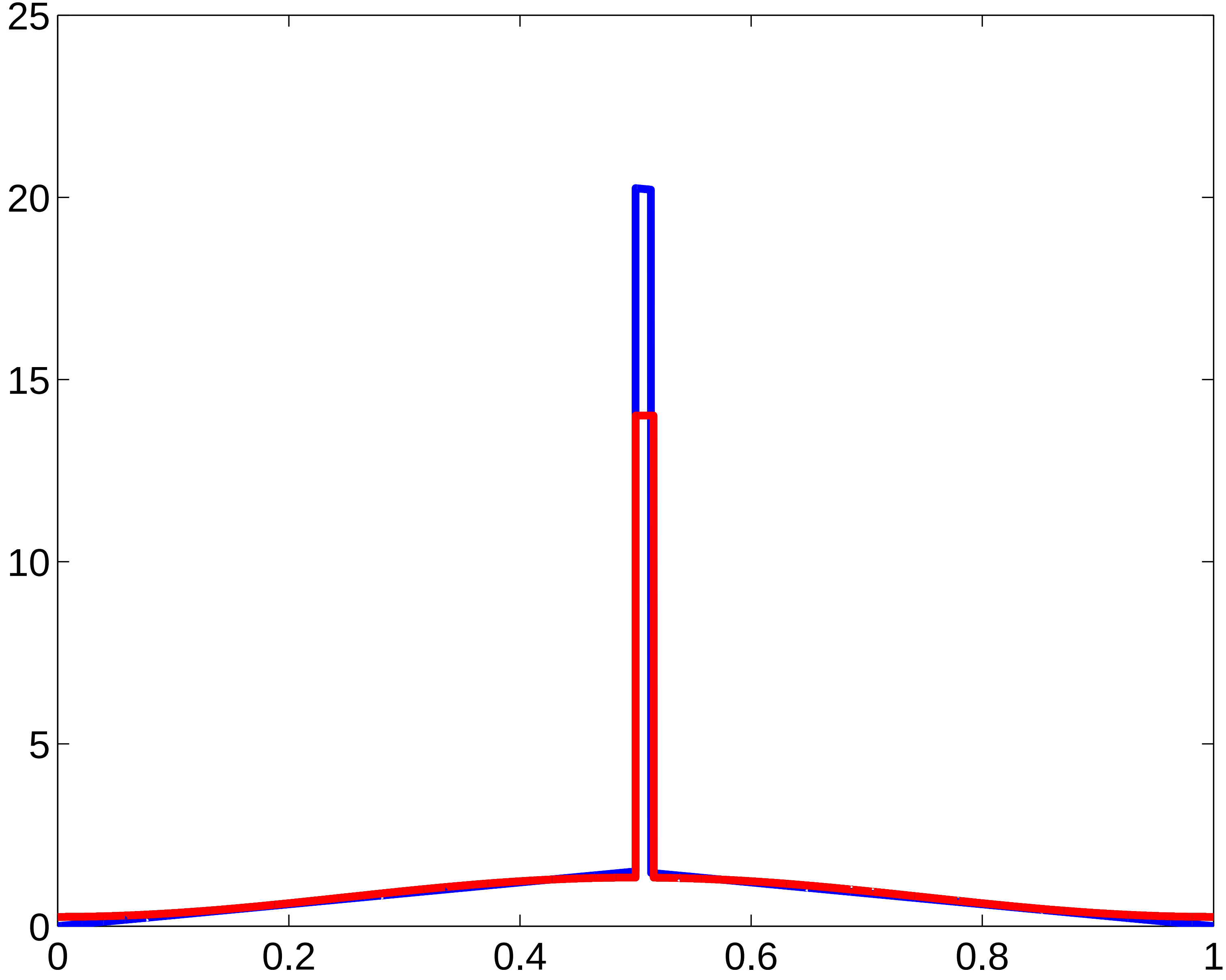
Fourier + Boxes



$$\|f - \hat{f}\|^2 = 0.74$$

$$\|\hat{\lambda}\|_0 = 5$$

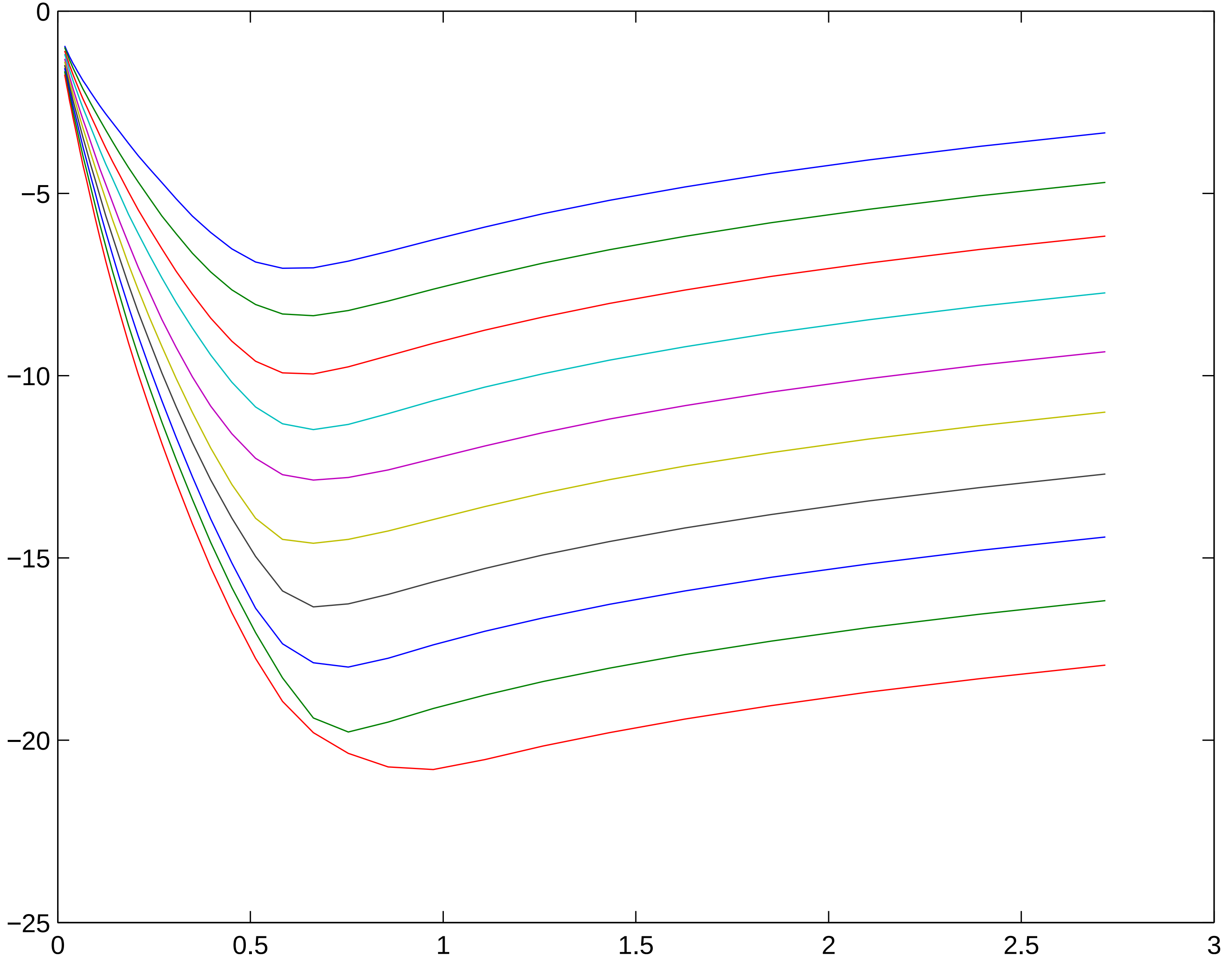
Fourier + Boxes + Haar



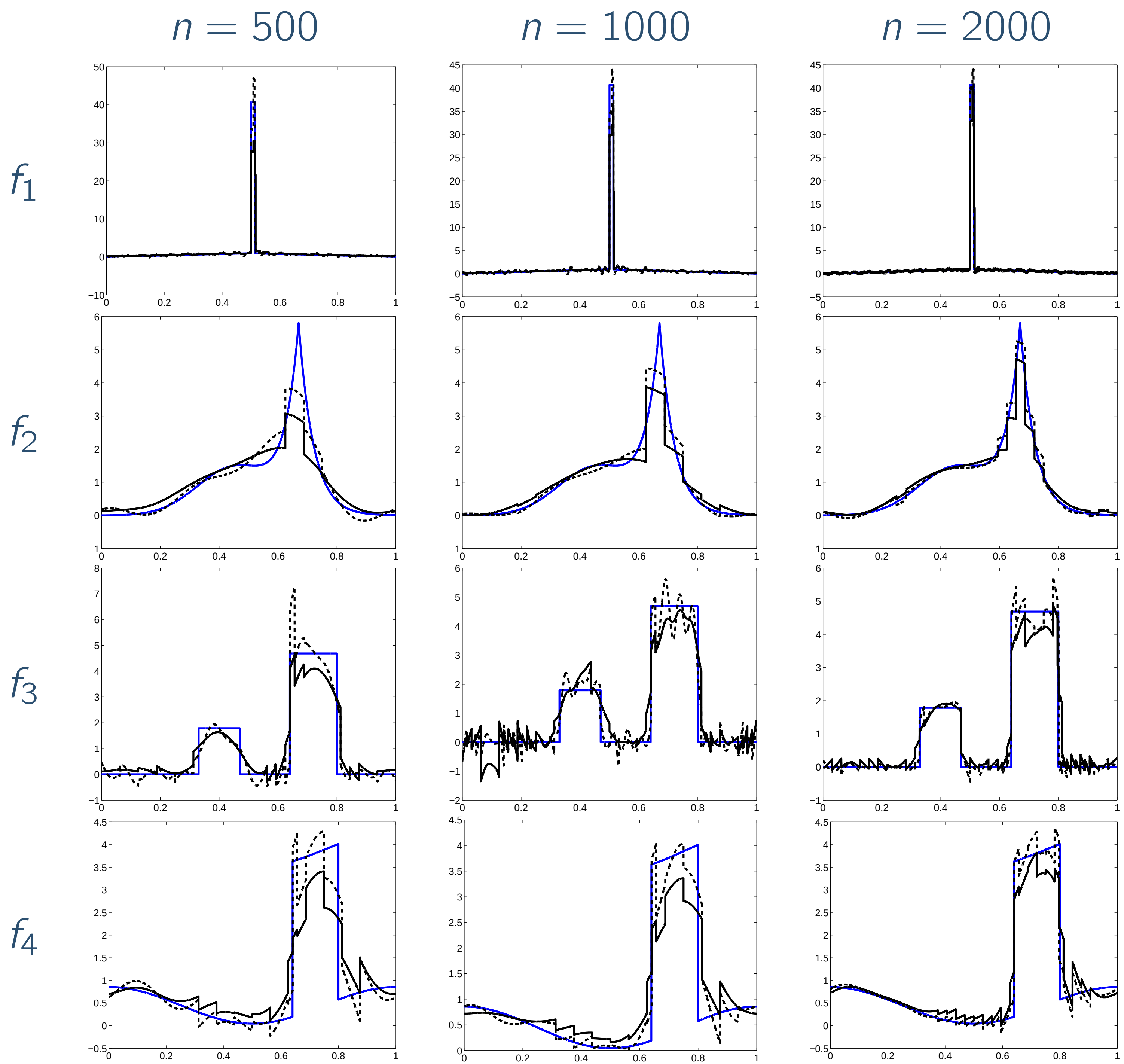
$\|f - \hat{f}\|^2 = 0.75$

$\|\hat{\lambda}\|_0 = 4$

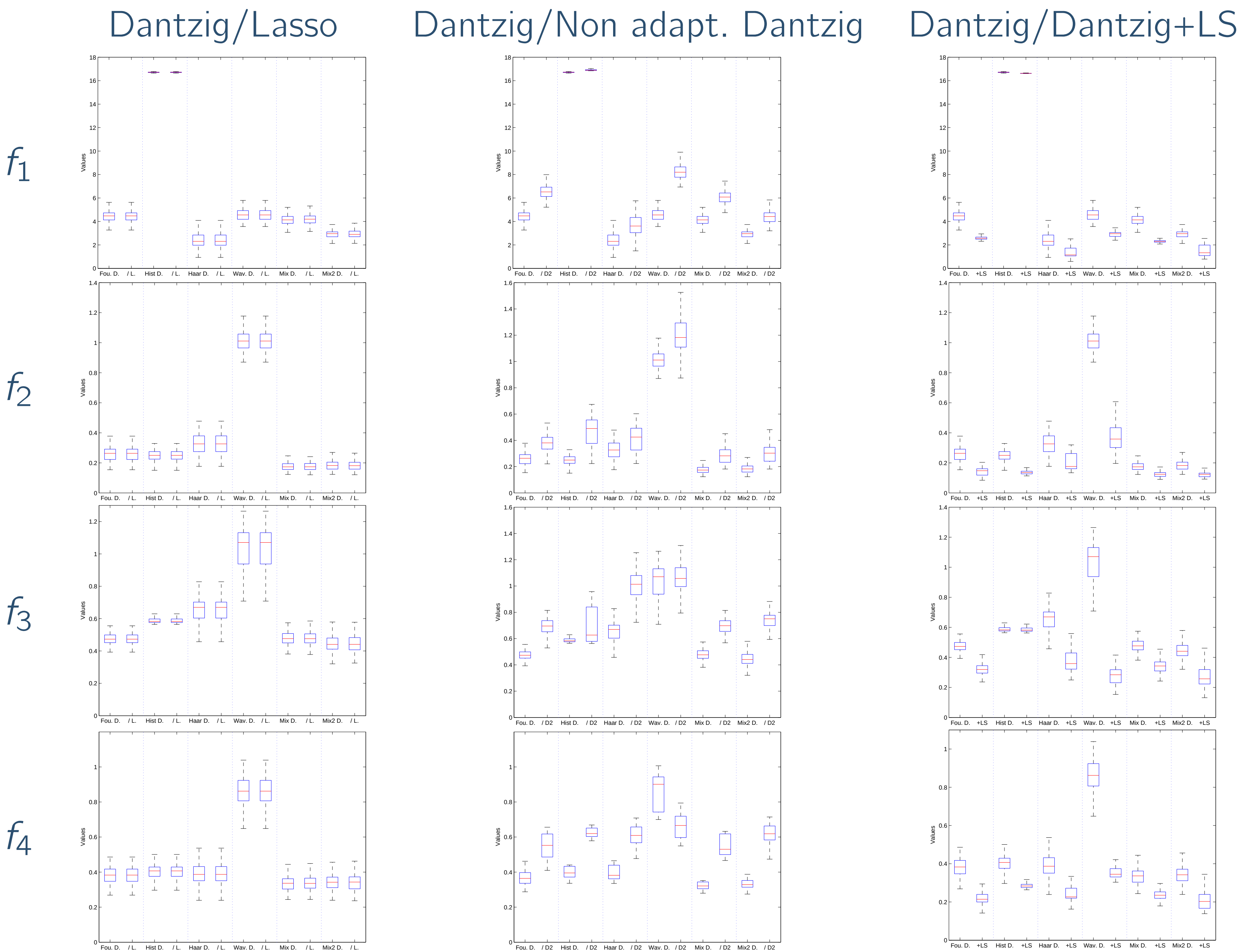
Minimal penalty



Dantzig and Dantzig+LS



Boxplots $n = 500$



Boxplots $n = 2000$

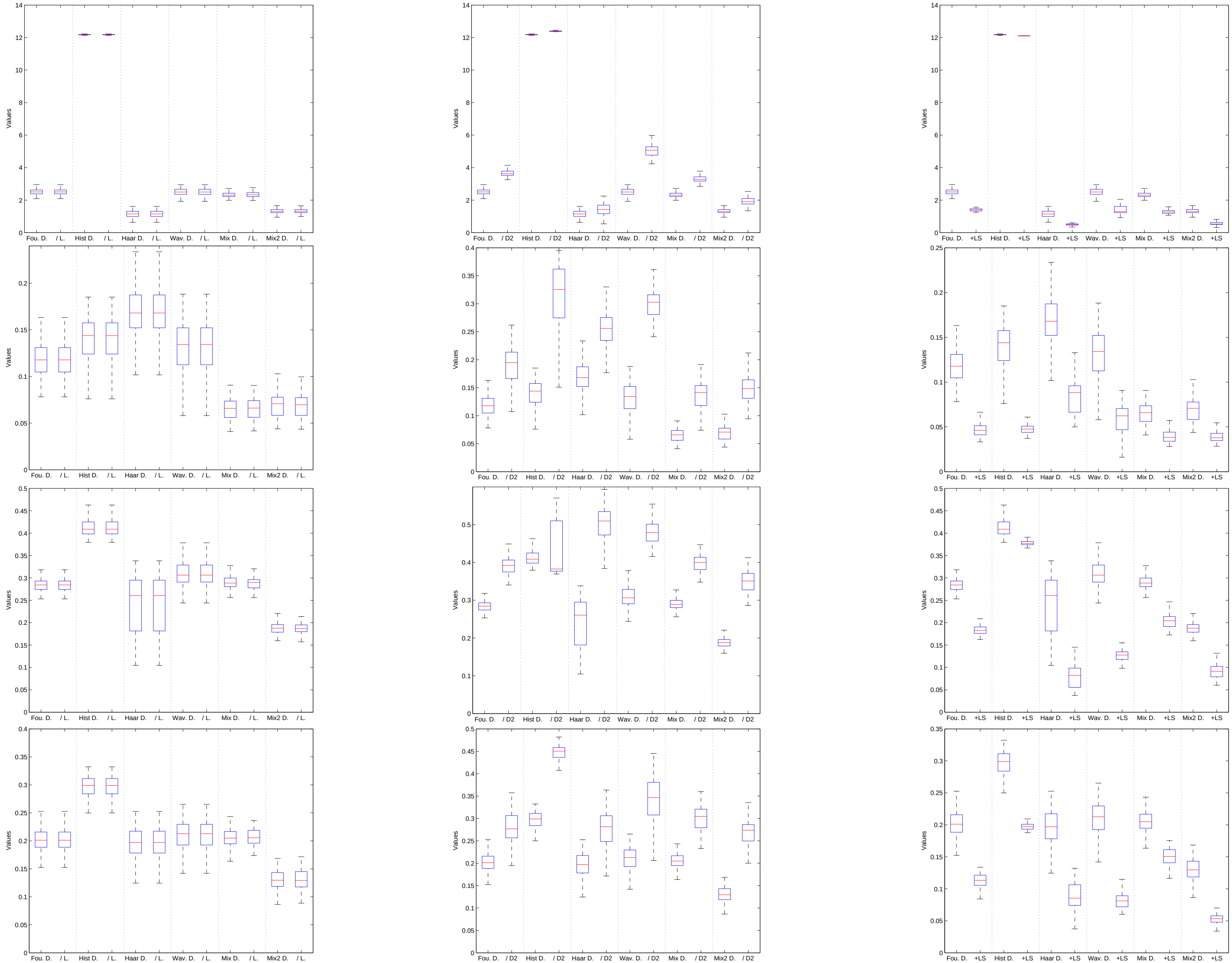
Dantzig/Lasso — Dantzig/Non adapt. Dantzig — Dantzig/Dantzig+LS

f_1

f_2

f_3

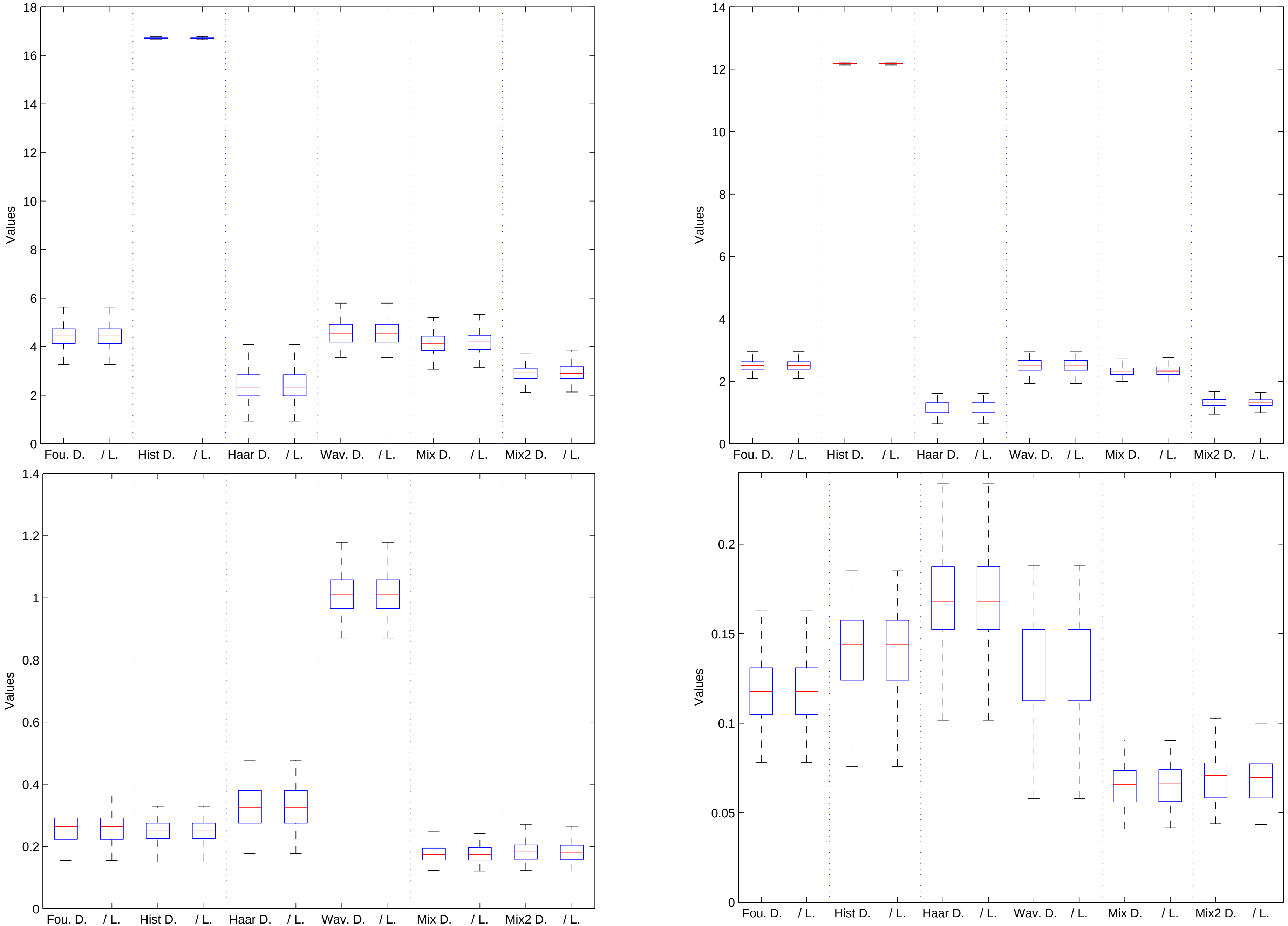
f_4



Dantzig / Lasso f_1/f_2

f_1

f_2

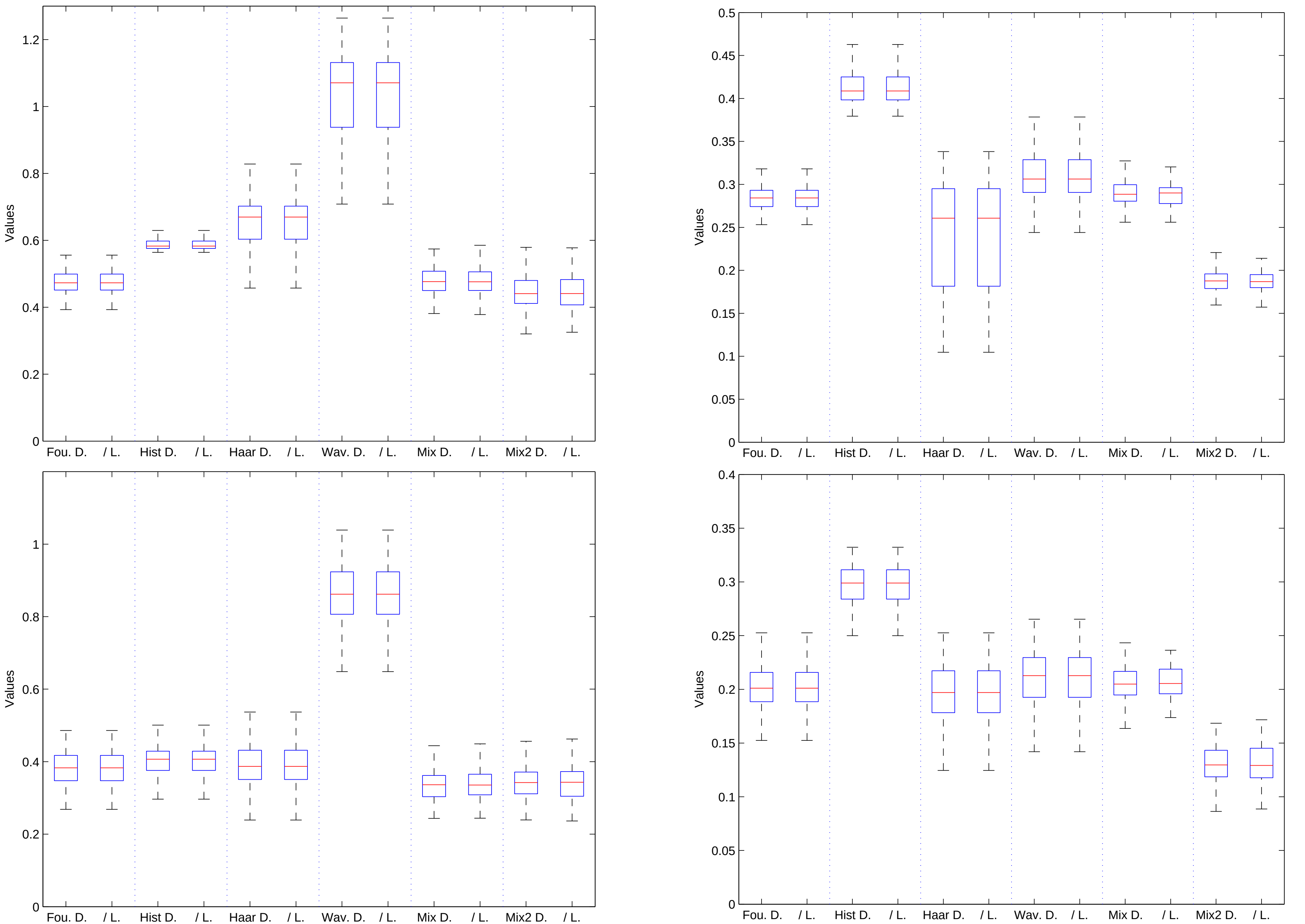


$n = 500$ $\quad$ $n = 2000$

Dantzig / Lasso f_3/f_4

f_3

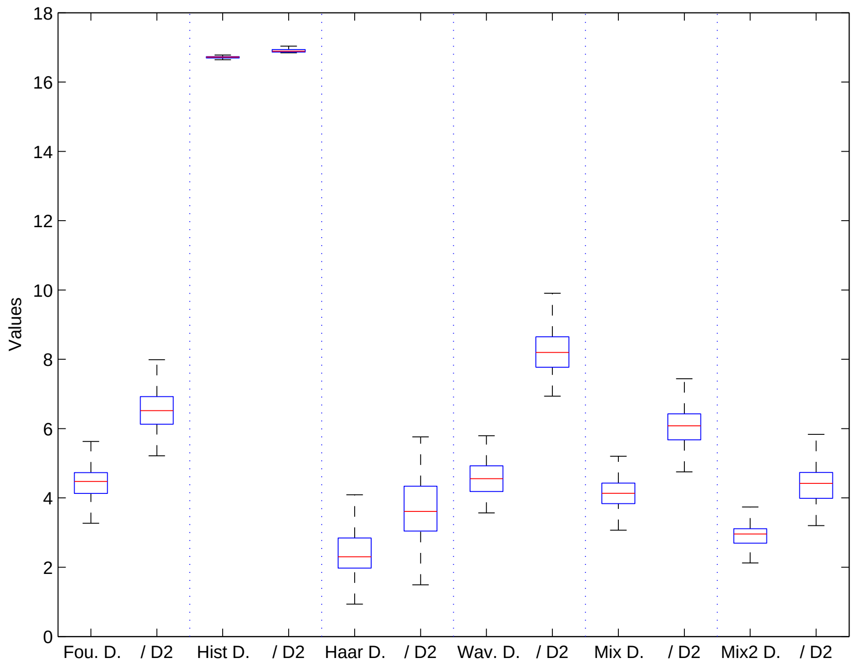
f_4

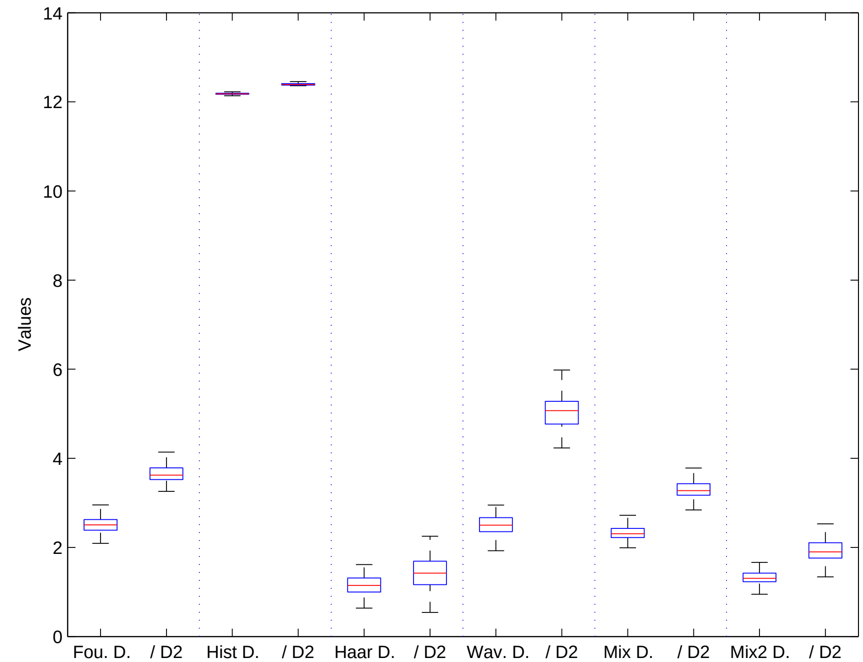


$n = 500$ $n = 2000$

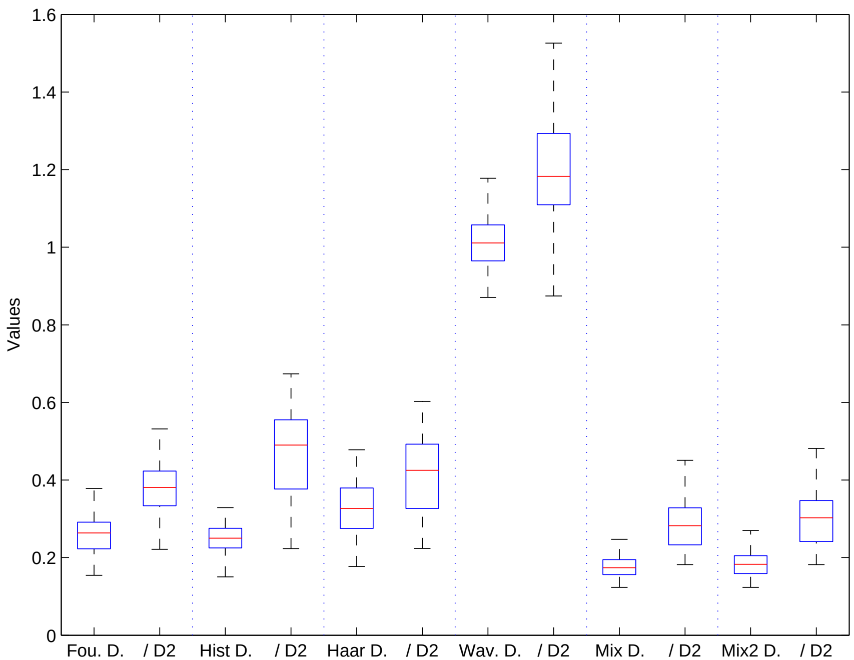
Dantzig / Non adaptive Dantzig f_1/f_2

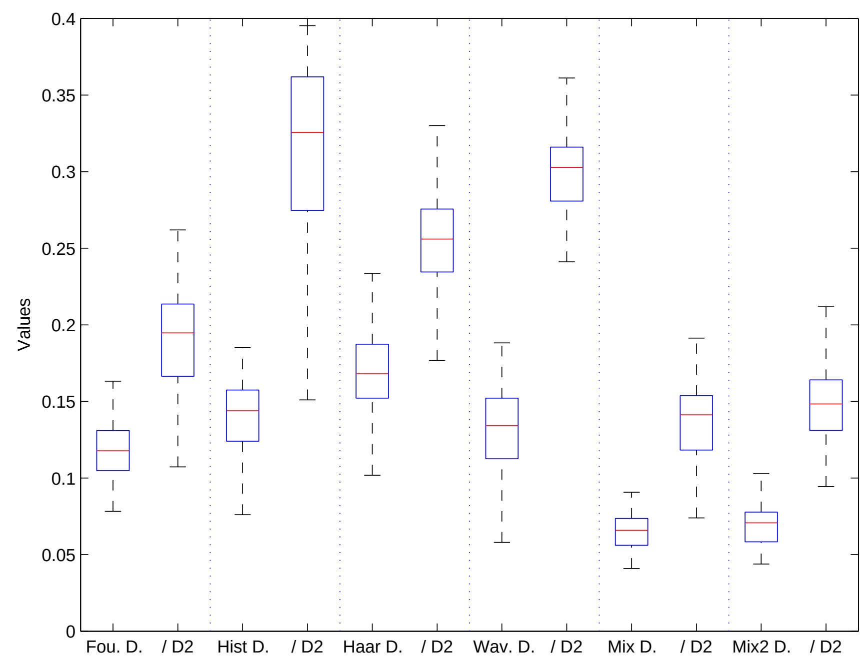
f_1





f_2

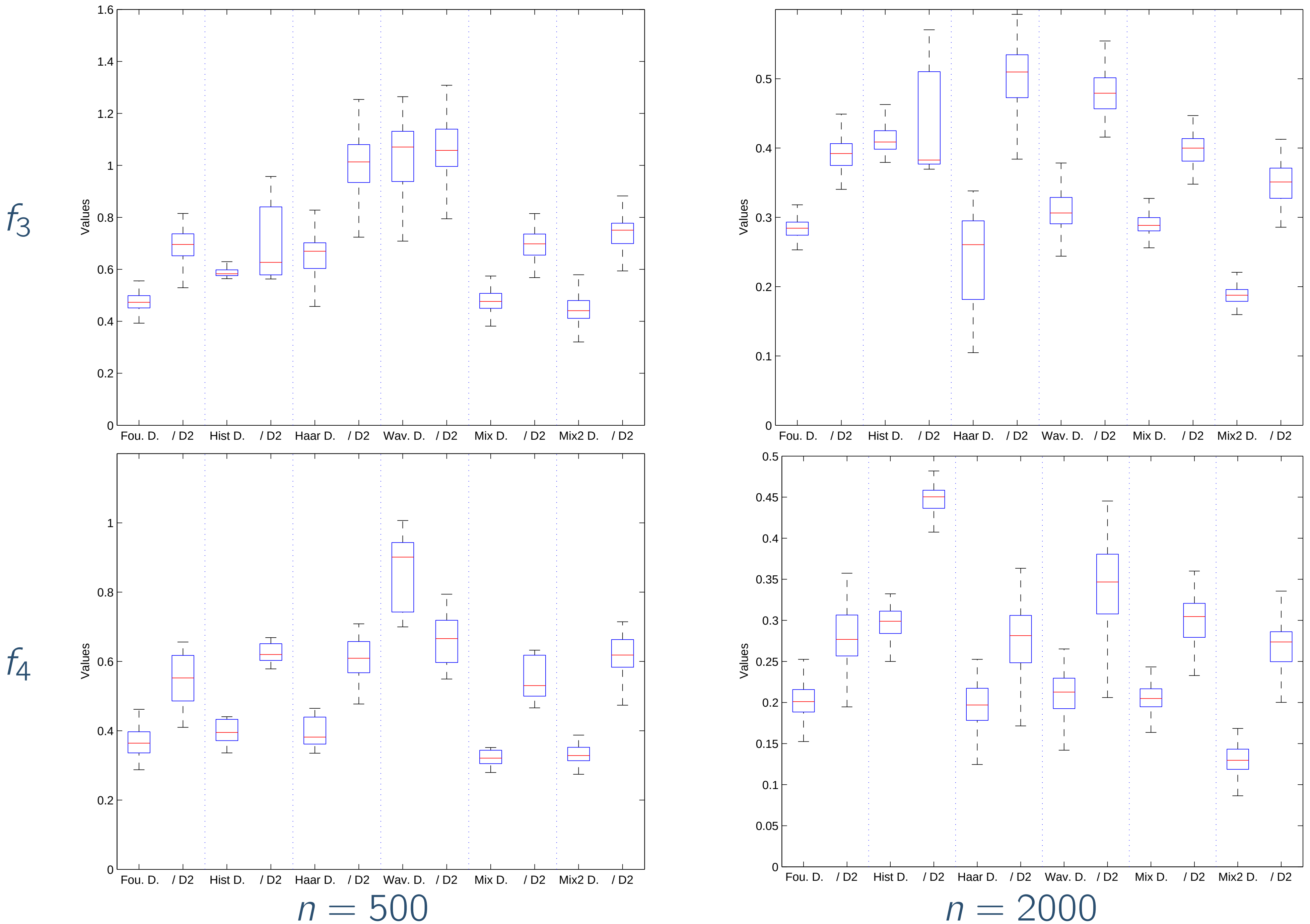




$n = 500$

$n = 2000$

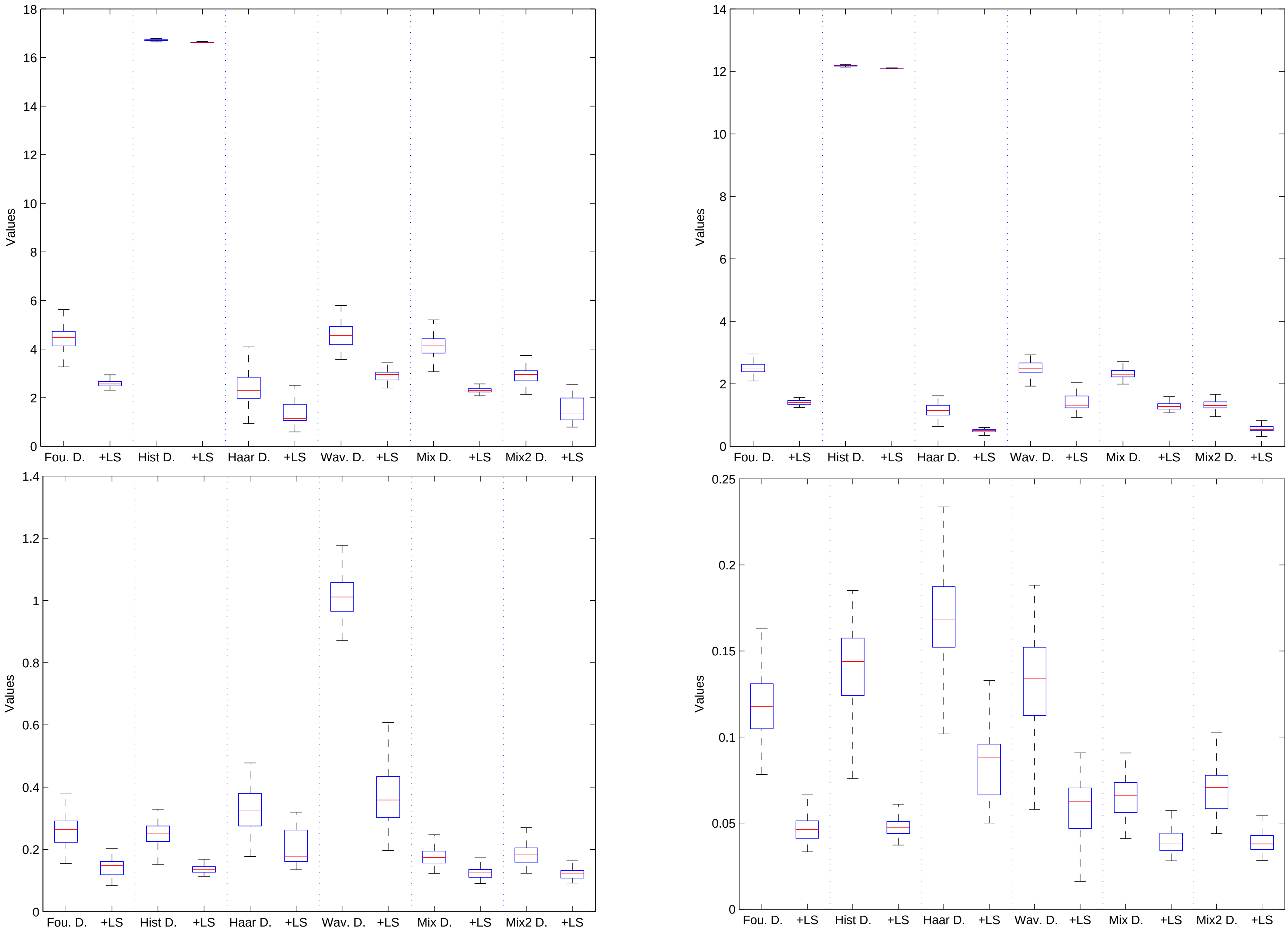
Dantzig / Non adaptive Dantzig f_3/f_4



Dantzig / Dantzig+LS f_1 / f_2

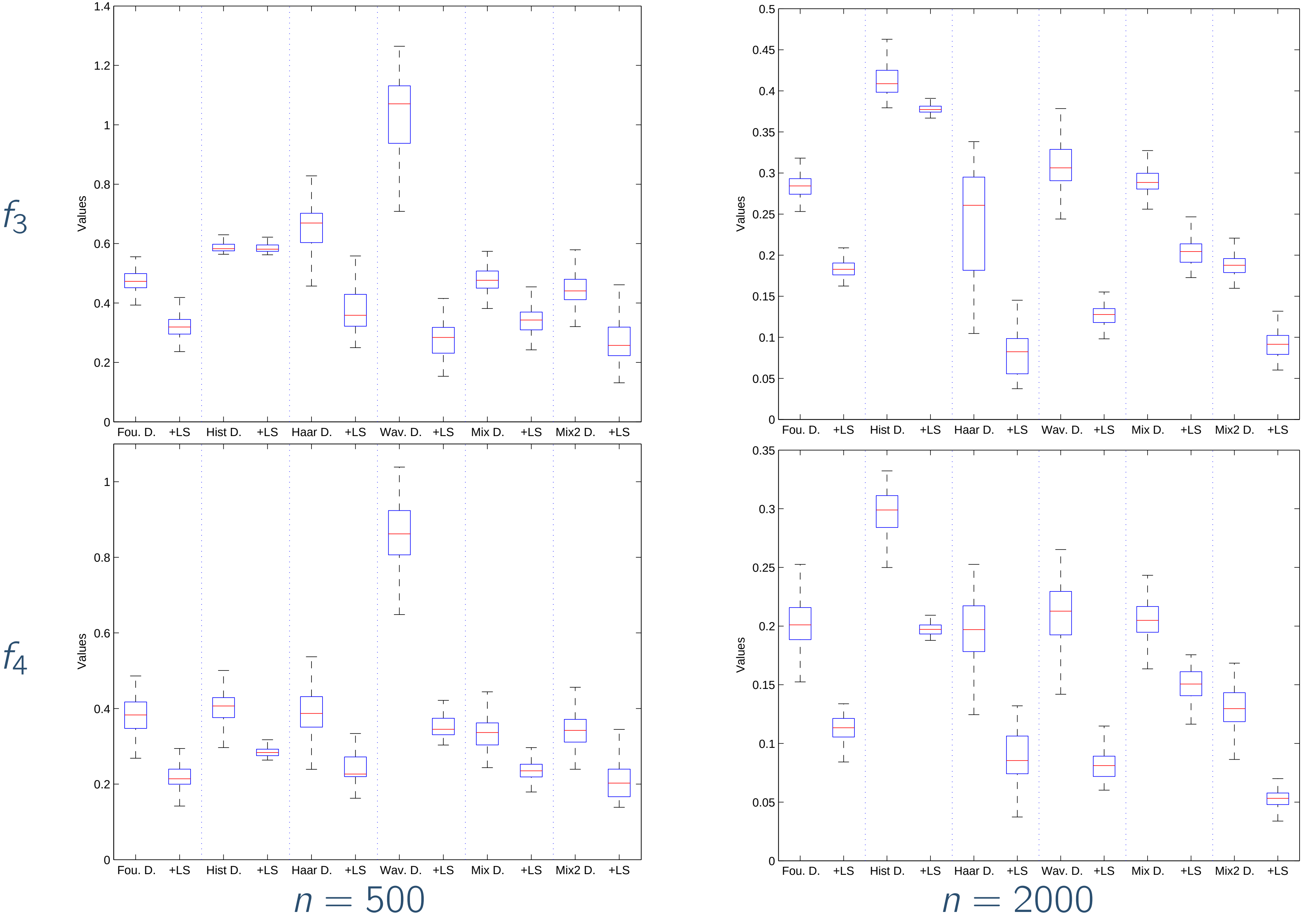
f_1

f_2

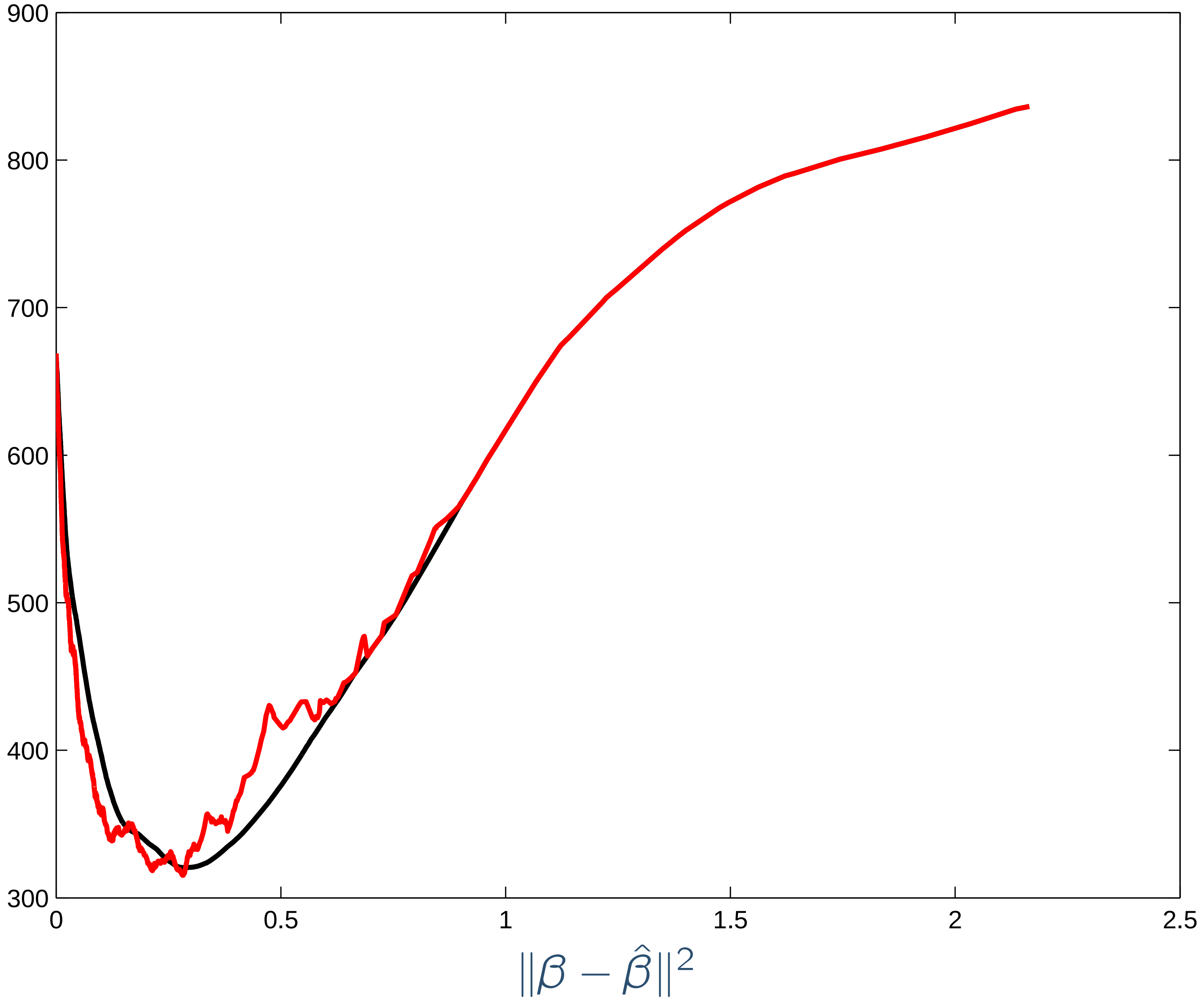


$n = 500$ $n = 2000$

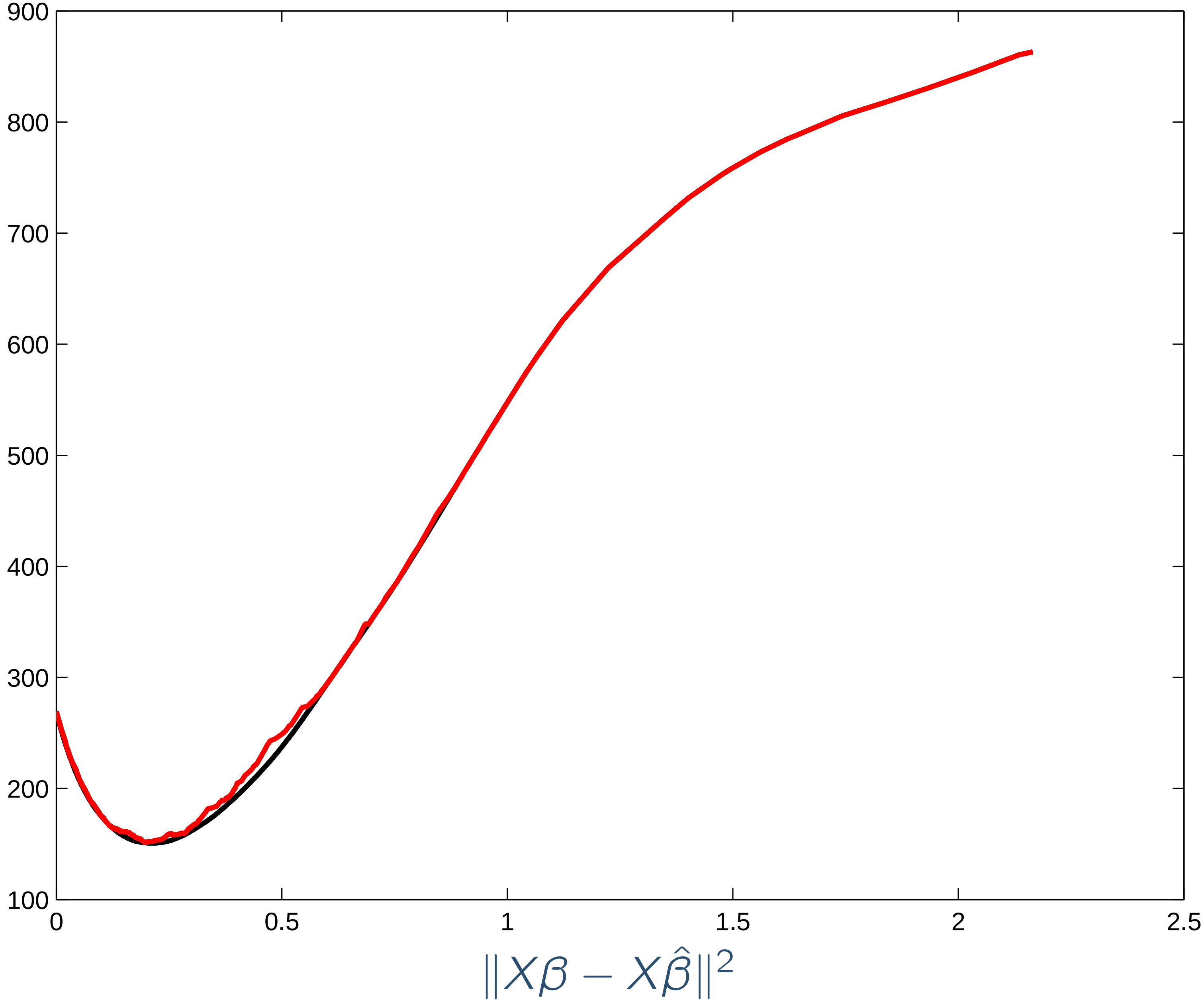
Dantzig / Dantzig+LS f_3/f_4



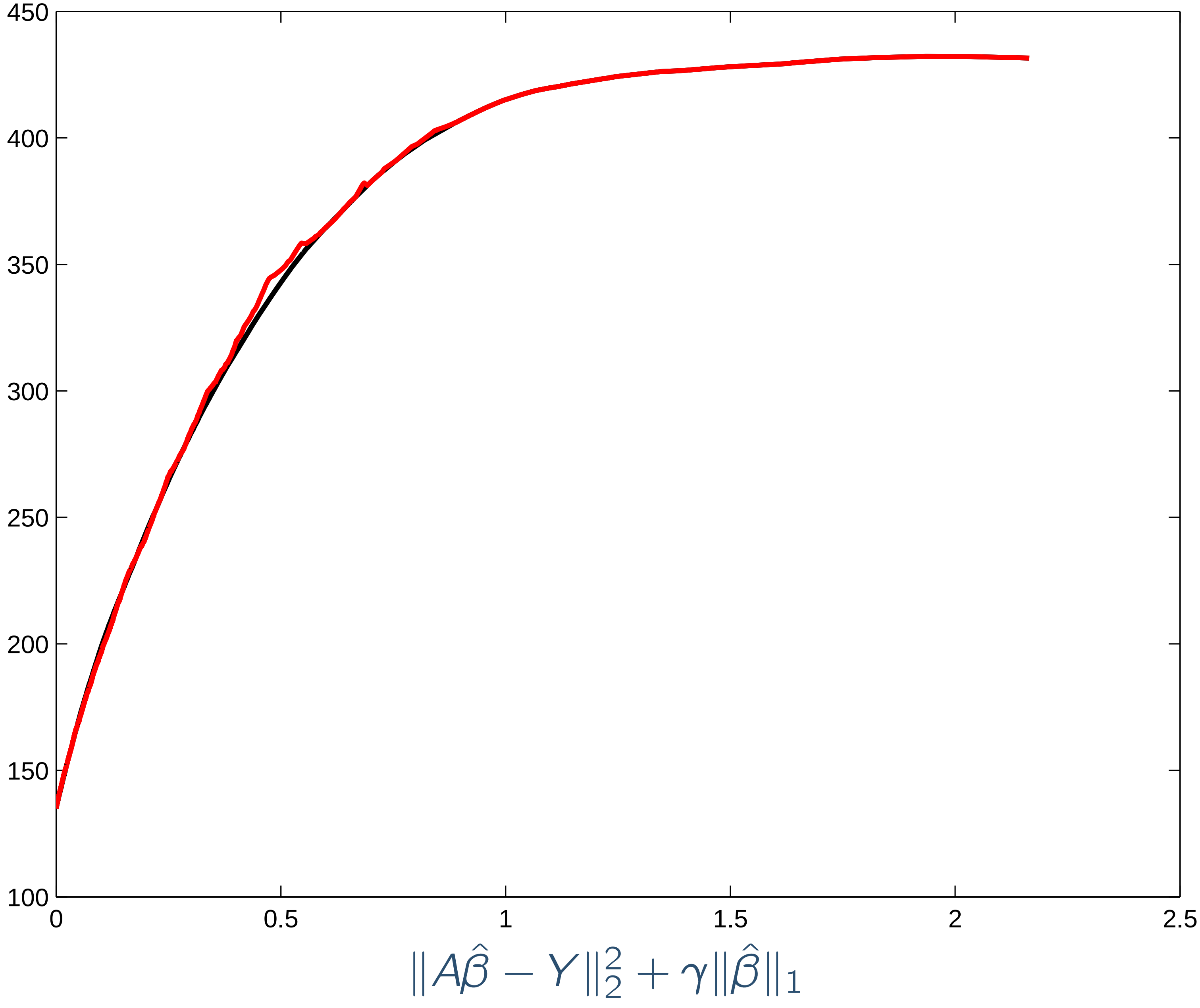
Lasso vs Dantzig



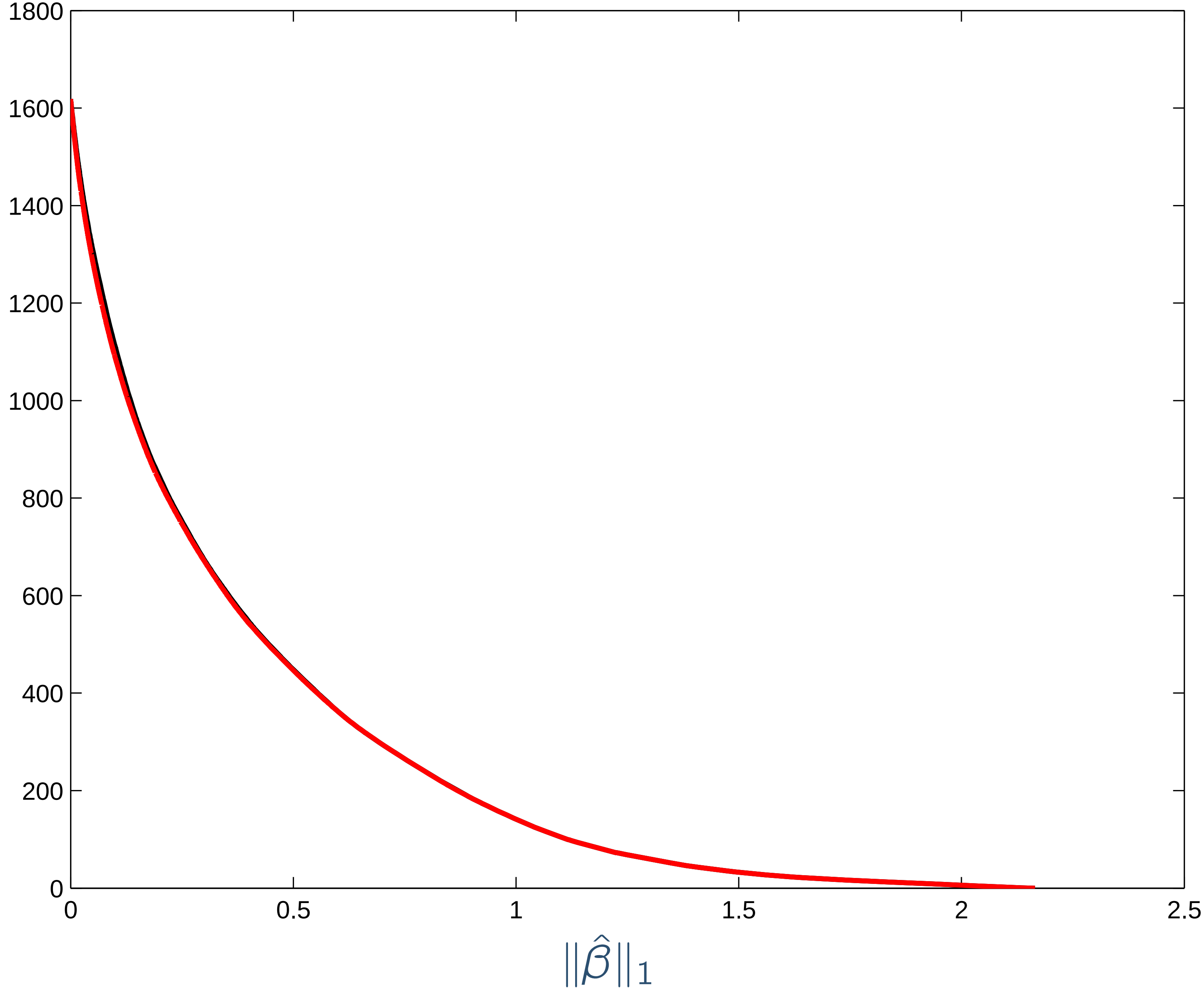
Lasso **vs** **Dantzig**



Lasso vs Dantzig



Lasso vs Dantzig



Lasso vs Dantzig