Wavelet regression in random design with heteroscedastic dependent errors

Rafał Kulik

Marseille, 16 December 2008

Session in memory of Marc Raimondo

Fixed-design regression under Long range Dependence

Fixed-design regression under Long range Dependence

We consider the fixed-design regression problem

$$Y_i = f(i/n) + \epsilon_i,$$

Fixed-design regression under Long range Dependence

We consider the fixed-design regression problem

$$Y_i = f(i/n) + \epsilon_i,$$

where $\epsilon_i, i \geq 1$ are LRD, Gaussian, with

$$\rho_m := \mathrm{E}\epsilon_0\epsilon_m \sim Lm^{-\alpha}, \alpha \in (0, 1]. \tag{1}$$

Fixed-design regression under Long range Dependence

We consider the fixed-design regression problem

$$Y_i = f(i/n) + \epsilon_i,$$

where $\epsilon_i, i \geq 1$ are LRD, Gaussian, with

$$\rho_m := \mathrm{E}\epsilon_0\epsilon_m \sim Lm^{-\alpha}, \alpha \in (0, 1]. \quad (1)$$

Then the equivalent continuous model is [1]

$$dY(t) = f(t)dt + \sigma n^{-\alpha/2} dB_{1-\alpha/2}(t).$$

[1] Brown, Low, Cai, 1990s

Estimator

Estimator

We expand

$$f(x) = \sum_{k=0}^{\infty} \alpha_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{\infty} \beta_{jk} \psi_{jk}(x),$$

where j_0 is a low resolution level, ϕ, ψ are scaling and wavelet functions, respectively; and $h_{jk}(x) := 2^{j/2} h(2^j x - k)$.

Estimator

We expand

$$f(x) = \sum_{k=0}^{\infty} \alpha_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{\infty} \beta_{jk} \psi_{jk}(x),$$

where j_0 is a low resolution level, ϕ, ψ are scaling and wavelet functions, respectively; and $h_{jk}(x) := 2^{j/2} h(2^j x - k)$. We estimate

$$\hat{f}_{j_1}(x) = \hat{\alpha}_{00} \phi_{00}(x) + \sum_{j=0}^{j_1} \sum_{k=0}^{2^j - 1} \hat{\beta}_{jk} 1\{|\hat{\beta}_{jk}| > \lambda\} \psi_{jk}(x), \quad (2)$$

where

$$\lambda = \tau 2^{-j(1-\alpha)/2} \frac{\sqrt{\log n}}{n^{\alpha/2}}.$$

Rates of convergence

Rates of convergence

Theorem 1. *Assume that* $p > 1$, $2^{j_1} \sim (n/\log(n))^{\alpha}$, $f \in B^s_{\pi,r}$ *with* $s \geq \frac{1}{\pi} \vee \frac{1}{2}$, *then there exists a constant* $C > 0$ *such that for all* $n \geq 0$,

$$\mathrm{E}\left\|f - \hat{f}_{j_1}\right\|_p^p \leq C\Big(\frac{(\log n)^{\frac{1}{\alpha}}}{n}\Big)^{\gamma},$$

$$\gamma = \frac{\alpha sp}{2(s+\frac{\alpha}{2})}, \quad \textit{if } s \geq \frac{\alpha}{2}(\frac{p}{\pi}-1), \quad s - \left(\frac{1}{\pi}-\frac{1}{p}\right)_+ > \frac{s}{2s+\alpha}, \tag{3}$$

$$\gamma = \frac{\alpha p(s-\frac{1}{\pi}+\frac{1}{p})}{2(s-\frac{1}{\pi}+\frac{\alpha}{2})}, \quad \textit{if } \frac{1}{\pi} < s < \frac{\alpha}{2}(\frac{p}{\pi}-1). \tag{4}$$

Rates of convergence

Theorem 1. *Assume that* $p > 1$, $2^{j_1} \sim (n/\log(n))^{\alpha}$, $f \in B^s_{\pi,r}$ *with* $s \geq \frac{1}{\pi} \vee \frac{1}{2}$, *then there exists a constant* $C > 0$ *such that for all* $n \geq 0$,

$$\mathrm{E}\left\|f - \hat{f}_{j_1}\right\|_p^p \leq C\left(\frac{(\log n)^{\frac{1}{\alpha}}}{n}\right)^{\gamma},$$

$$\gamma = \frac{\alpha sp}{2(s+\frac{\alpha}{2})}, \qquad \text{if } s \geq \frac{\alpha}{2}(\frac{p}{\pi}-1), \quad s - \left(\frac{1}{\pi}-\frac{1}{p}\right)_+ > \frac{s}{2s+\alpha}, \tag{3}$$

$$\gamma = \frac{\alpha p(s-\frac{1}{\pi}+\frac{1}{p})}{2(s-\frac{1}{\pi}+\frac{\alpha}{2})}, \qquad \text{if } \frac{1}{\pi} < s < \frac{\alpha}{2}(\frac{p}{\pi}-1). \tag{4}$$

2

[2] Wang (1996), Johnstone and Silverman (1997), Johnstone (1999), Kulik and Raimondo (2009)

Random-design regression

Random-design regression

Assume that

$$Y_i = f(X_i) + \sigma(X_i)\epsilon_i,$$

where $X_i \sim G$ are random.

Random-design regression

Assume that

$$Y_i = f(X_i) + \sigma(X_i)\epsilon_i,$$

where $X_i \sim G$ are random.

If ϵ_i are i.i.d., it turns out that this model is not equivalent to the continuous model, unless X_i are standard uniform.

Random-design regression

Assume that

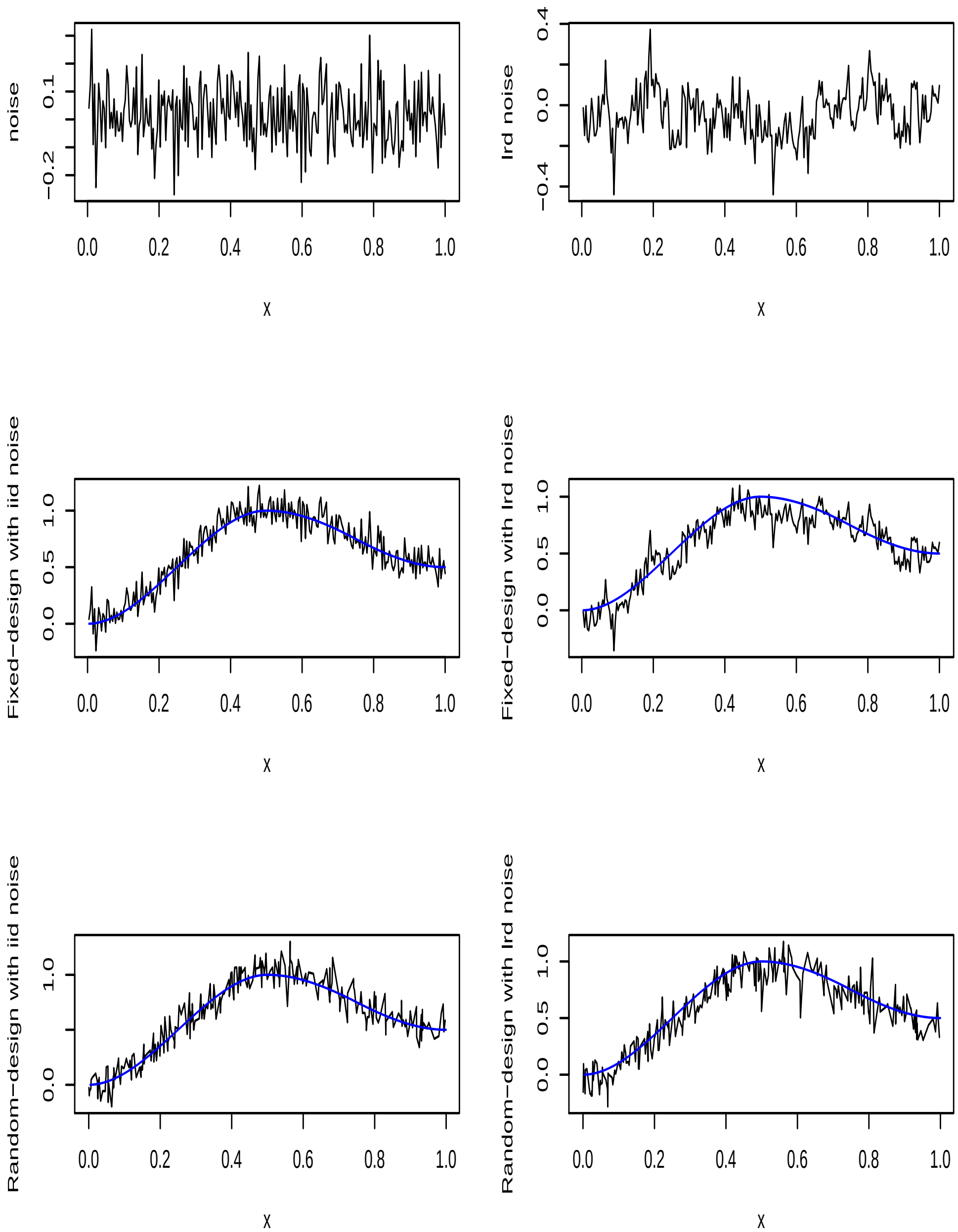
$$Y_i = f(X_i) + \sigma(X_i)\epsilon_i,$$

where $X_i \sim G$ are random.

If ϵ_i are i.i.d., it turns out that this model is not equivalent to the continuous model, unless X_i are standard uniform.

Moreover, if ϵ_i are LRD it cannot be equivalent to any continuous model.

Heuristic


noise
lrd noise
Fixed−design with iid noise
Fixed−design with lrd noise
Random−design with iid noise
Random−design with lrd noise
x
0.0
0.2
0.4
0.6
0.8
1.0

Estimator

Estimator

The estimator we are going to consider has the form as in (2) with:

$$\hat{\beta}_{jk} := \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(G(X_i))Y_i, \tag{5}$$

Estimator

The estimator we are going to consider has the form as in (2) with:

$$\hat{\beta}_{jk} := \frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(G(X_i))Y_i, \qquad (5)$$

[3]

[3] The idea originates from Kerkyacharian and Picard (2004) - warped wavelets

Estimator

The estimator we are going to consider has the form as in (2) with:

$$\hat{\beta}_{jk} := \frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(G(X_i))Y_i, \tag{5}$$

[3] where

$$2^{j_1} \sim \frac{n}{\log n},$$

[3] The idea originates from Kerkyacharian and Picard (2004) - warped wavelets

Estimator

The estimator we are going to consider has the form as in (2) with:

$$\hat{\beta}_{jk} := \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(G(X_i))Y_i, \tag{5}$$

[3] where

$$2^{j_1} \sim \frac{n}{\log n},$$

$$\lambda = \lambda_{n,j} = \max\left\{2^{-j/2}\mathrm{E}\left[\psi_{jk}(G(X_1))\sigma(X_1)\right]\frac{(\log n)^{1/2}}{n^{\alpha/2}}, \frac{\log n}{\sqrt{n}}\right\}.$$

[3] The idea originates from Kerkyacharian and Picard (2004) - warped wavelets

Estimator

The estimator we are going to consider has the form as in (2) with:

$$\hat{\beta}_{jk} := \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(G(X_i))Y_i, \qquad (5)$$

[3] where

$$2^{j_1} \sim \frac{n}{\log n},$$

$$\lambda = \lambda_{n,j} = \max\left\{2^{-j/2}\mathrm{E}\left[\psi_{jk}(G(X_1))\sigma(X_1)\right]\frac{(\log n)^{1/2}}{n^{\alpha/2}}, \frac{\log n}{\sqrt{n}}\right\}.$$

Note: In most *regular* cases, we have $\mathrm{E}[\cdot] = 0$.

[3] The idea originates from Kerkyacharian and Picard (2004) - warped wavelets

Rates of convergence

Rates of convergence

Theorem 2. *Assume that* $p > 1$, $2^{j_1} \sim n/\log(n)$, $f \circ G^{-1} \in B^s_{\pi,r}$ *with* $s \geq \frac{1}{\pi} \vee \frac{1}{2}$, *then there exists a constant* $C > 0$ *such that for all* $n \geq 0$

$$\mathrm{E}\|f - \hat{f}_n\|^p_{L^p(g)} \leq Cn^{-\gamma(s)}(\log n)^{\gamma(s)},$$

$$\gamma(s) = \begin{cases} \frac{ps}{2s+1}, & \text{if } s > \frac{p-\pi}{2\pi} \text{ and } \alpha > \alpha_D := \frac{2s}{2s+1}, \\ p\frac{s-\left(\frac{1}{\pi}-\frac{1}{p}\right)}{1+2\left(s-\frac{1}{\pi}\right)}, & \text{if } s < \frac{p-\pi}{2\pi} \text{ and } \alpha > \alpha_S := \frac{2(s-(1/\pi-1/p))}{2(s-1/\pi)+1}. \end{cases}$$

Otherwise, $\mathrm{E}\|f - \hat{f}_n\|^p_p \leq Cn^{-p\alpha/2}$.

Comments

Comments

- The result is stated in weighted norm $\|f-h\|_{L^p(g)}^p = \int |f-h|^p dG$.

Comments

- The result is stated in weighted norm $\|f-h\|_{L^p(g)}^p = \int |f-h|^p dG$.

- The estimator matches the minimax rates.

Comments

- The result is stated in weighted norm $\|f-h\|_{L^p(g)}^p = \int |f-h|^p dG$.
- The estimator matches the minimax rates.
- Trichotomy.

Comments

- The result is stated in weighted norm $\|f-h\|_{L^p(g)}^p = \int |f-h|^p dG$.

- The estimator matches the minimax rates.

- Trichotomy.

- One can replace G with G_n (empirical cdf) and get the same rates, however, we must stop at

$$2^{j_1} \sim \sqrt{\frac{n}{\log n}}.$$

Comments

- The result is stated in weighted norm $\|f-h\|_{L^p(g)}^p = \int |f-h|^p dG$.

- The estimator matches the minimax rates.

- Trichotomy.

- One can replace G with G_n (empirical cdf) and get the same rates, however, we must stop at

$$2^{j_1} \sim \sqrt{\frac{n}{\log n}}.$$

- If we estimate $f^* = f - \mathrm{E}[f(X_1)]$ we do not have LRD effect.

Proof

Proof

$$\hat{\beta}_{jk} - \beta_{jk} = \frac{1}{n}\sum_{i=1}^{n} \left(\psi_{jk}(G(X_i))Y_i - \mathrm{E}[\psi_{jk}(G(X_i))Y_i]\right)$$

$$= \quad \frac{1}{n}\sum_{i=1}^{n} \left(\psi_{jk}(G(X_i))f(X_i) - \mathrm{E}[\psi_{jk}(G(X_1))f(X_1)]\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i =: A_0 + A_1. \tag{6}$$

Proof

$$\hat{\beta}_{jk} - \beta_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left(\psi_{jk}(G(X_i))Y_i - \mathrm{E}[\psi_{jk}(G(X_i))Y_i]\right)$$

$$= \quad \frac{1}{n}\sum_{i=1}^{n}\left(\psi_{jk}(G(X_i))f(X_i) - \mathrm{E}[\psi_{jk}(G(X_1))f(X_1)]\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i =: A_0 + A_1. \tag{6}$$

Note that A_0 is the sum of iid random variables, whereas the dependence structure is included in A_1 only.

Proof

$$\hat{\beta}_{jk} - \beta_{jk} = \frac{1}{n}\sum_{i=1}^{n}(\psi_{jk}(G(X_i))Y_i - \mathrm{E}[\psi_{jk}(G(X_i))Y_i])$$

$$= \quad \frac{1}{n}\sum_{i=1}^{n}(\psi_{jk}(G(X_i))f(X_i) - \mathrm{E}[\psi_{jk}(G(X_1))f(X_1)])$$

$$+\frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i =: A_0 + A_1. \tag{6}$$

Note that A_0 is the sum of iid random variables, whereas the dependence structure is included in A_1 only.
The part A_1 is decomposed further. Let $\mathcal{F}_i = \sigma(X_i, \eta_i, X_{i-1}, \eta_{i-1}, \ldots)$.
Let $\epsilon_{i,i-1} = \epsilon_i - \eta_i$. Then,

$$\mathrm{E}[\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i|\mathcal{F}_{i-1}] = \epsilon_{i,i-1}\mathrm{E}[\psi_{jk}(G(X_1))\sigma(X_1)].$$

Proof - ctd.

Proof - ctd.

Write

$$A_1 = \frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i$$

$$= \quad \frac{1}{n}\sum_{i=1}^{n}(\psi_{jk}(G(X_i))\epsilon_i - \mathrm{E}[\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i|\mathcal{F}_{i-1}])$$

$$+\frac{1}{n}\mathrm{E}[\psi_{jk}(G(X_1))\sigma(X_1)]\sum_{i=1}^{n}\epsilon_{i,i-1} =: A_2 + A_3.$$

Proof - ctd.

Write

$$
\begin{aligned}
A_1 = \frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i \\
= \quad \frac{1}{n}\sum_{i=1}^{n}(\psi_{jk}(G(X_i))\epsilon_i - \mathrm{E}[\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i|\mathcal{F}_{i-1}]) \\
+\frac{1}{n}\mathrm{E}[\psi_{jk}(G(X_1))\sigma(X_1)]\sum_{i=1}^{n}\epsilon_{i,i-1} =: A_2 + A_3.
\end{aligned}
$$

The crucial feature of this decomposition is that we reduce the problem of studying the partial sums $\sum_{i=1}^{n}\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i$ to the partial sums $\sum_{i=1}^{n}\epsilon_{i,i-1}$ and the martingale A_2.

This leads to

$$\mathrm{E}\left[|\hat{\beta}_{jk} - \beta_{jk}|^p\right] \le \begin{cases} C\, n^{-p/2} & \text{if } 2^j > n^{1-\alpha}, \\ C\, 2^{-jp/2} n^{-p\alpha/2} & \text{if } 2^j < n^{1-\alpha} \end{cases}$$

and appropriate large deviations inequalities.

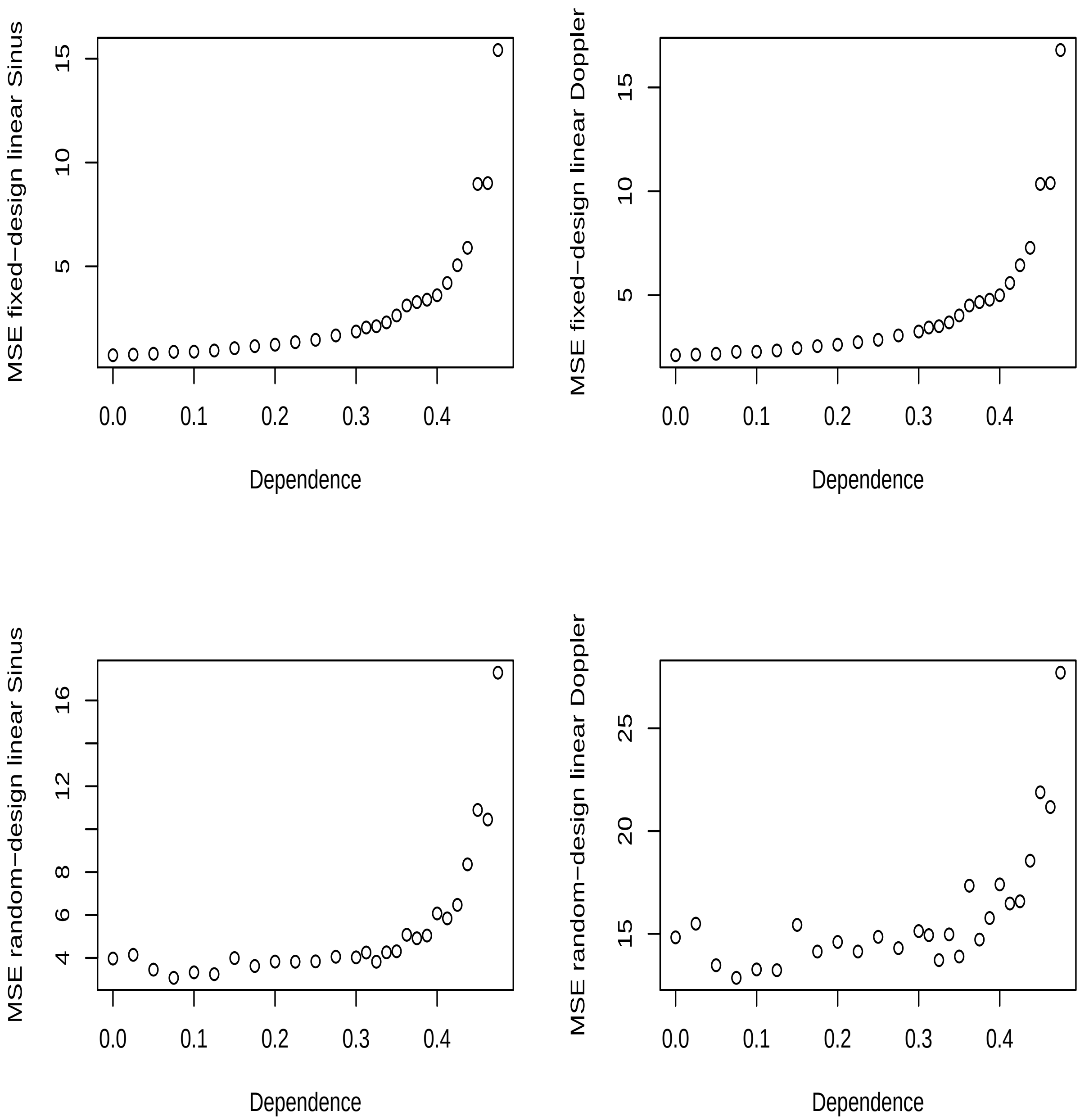
This leads to

$$E\left[|\hat{\beta}_{jk} - \beta_{jk}|^p\right] \le \begin{cases} C\, n^{-p/2} & \text{if } 2^j > n^{1-\alpha}, \\ C\, 2^{-jp/2} n^{-p\alpha/2} & \text{if } 2^j < n^{1-\alpha} \end{cases}$$

and appropriate large deviations inequalities.

In principle, low resolution levels are affected by long memory, high resolution levels are not.

Dichotomy


MSE fixed−design linear Sinus
MSE fixed−design linear Doppler
MSE random−design linear Sinus
MSE random−design linear Doppler
Dependence
0.0 0.1 0.2 0.3 0.4