Small Area Estimation under Fay–Herriot Models with Nonparametric Estimation of Heteroscedasticity. [a]

Domingo Morales González

d.morales@umh.es

Universidad Miguel Hernández de Elche

[a] In collaboration with W. González-Manteiga, M.J. Lombardía, I. Molina and L. Santamaría L.

The problem: Small area estimation of mean tourist expenditure in the autonomous community of Galicia with data from year 2004.

- Galicia is a region in the northwest of Spain which is partitioned into $D = 52$ counties, regarded here as small areas.
- The available data were:
  - the number of offered hotel vacancies x_d for each small area d, and
  - the expenditure in accommodation/day y_{dj} of each tourist j in each area d.
  - in total, 2496 tourists were interviewed.
- The direct estimate of the mean expense in accomodation/day in that small area is

$$y_d = \sum_{j=1}^{n_d} y_{dj}/n_d,$$

where n_d is the sample size of small area d.

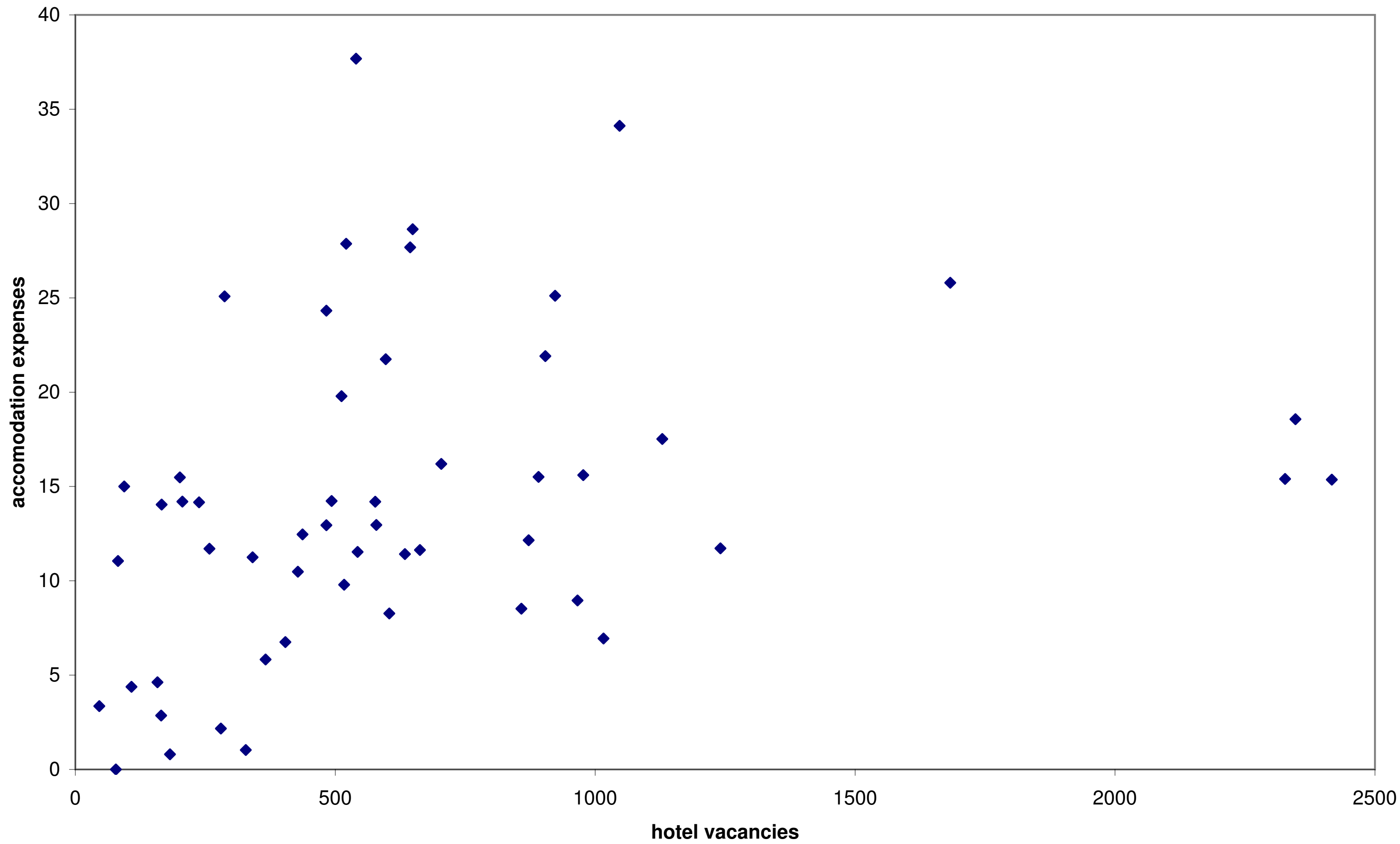

Figure 1a. Small area direct estimates of the mean tourist expense in accommodation (y_d) versus numbers of vacancies in hotels (x_d).

- The following Fay-Herriot model is considered

$$y_d = \beta_0 + \beta_1 x_d + u_d + w_d^{1/2} e_d, \quad d = 1, \ldots, D,$$

where $u_d \sim N(0, \sigma_u^2)$ and $e_d \sim N(0, \sigma_e^2)$, $d = 1, \ldots, D$, are all independent.

- First we fit the associated homoscedastic model obtained by replacing $w_d = 1$, $d = 1, \ldots, D$, and $\sigma_e^2 = S_y^2$ where

$$S_y^2 = \frac{1}{D} \sum_{d=1}^{D} S_{yd}^2, \quad S_{yd}^2 = \frac{1}{n_d(n_d - 1)} \sum_{j=1}^{n_d} (y_{dj} - \overline{y}_d)^2, \ d = 1, \ldots, D.$$

- S_{yd}^2 is the estimate of the sampling error of the direct estimator y_d, and S_y^2 the mean of S_{yd}^2, $d = 1, \ldots, D$.

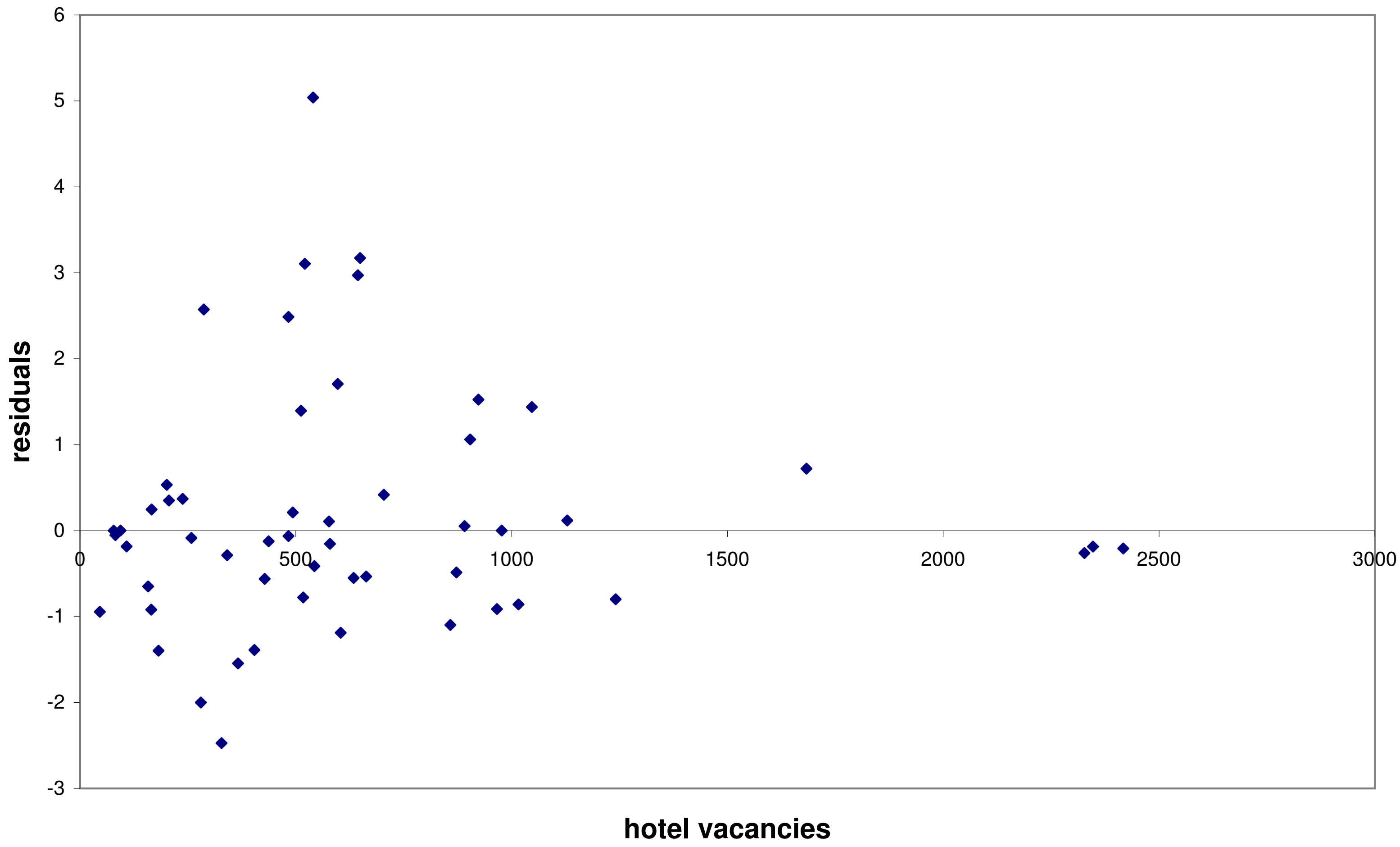

Figure 1b. Residuals $y_d - \mathbf{x}_d^t \hat{\beta}_0 - \hat{u}_{d0}$ versus numbers of hotel vacancies x_d in each area.

- In Figure 1 and 2 we observe heteroscedasticity ($w_{dj} \neq 1$).
- However, the parametric form of the error variance function cannot be clearly guessed from those plots.
- In this situation, it makes sense to consider a nonparametric approach for estimating the heteroscedasticity function.
- Thus, we consider that the available data $(\mathbf{x}_d^t, y_d)$, $d = 1, \ldots, D$, follow a Fay-Herriot model, where heteroscedasticity is introduced through the weights w_d, that is,

$$y_d = \mathbf{x}_d^t \boldsymbol{\beta} + u_d + w_d^{1/2} e_d, \quad u_d \stackrel{iid}{\sim} (0, \sigma_u^2), \quad e_d \stackrel{iid}{\sim} (0, \sigma_e^2). \quad (1)$$

  - y_d is the direct estimator of the mean of the d-th small area,
  - $\mathbf{x}_d$ is a vector that contains the values of p auxiliary variables (p constant),
  - $\boldsymbol{\beta}$ is the vector of (fixed) effects of the x-variables,

- Random effects and residuals are such that
  - u_d is a random effect associated to small area d with unknown constant variance σ_u^2, and
  - e_d is the random error that represents the sampling error of the direct estimator y_d, independent of u_d, and with variance σ_e^2.
- Let z_d denote either x_d if $p = 1$ or the mean response $\mathbf{x}_d^t\boldsymbol{\beta}$ if $p > 1$.
  - We assume that the heteroscedasticity weights are function of z_d; that is, $w_d = w(z_d)$, $d = 1, \ldots, D$, where the function $w(\cdot)$ is smooth in the sense of assumption (M4) below.
- In matrix notation model (1) can be written as

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u} + W^{1/2}\mathbf{e}, \quad \mathbf{u} \sim (\mathbf{0}, \sigma_u^2 I_D), \quad \mathbf{e} \sim (\mathbf{0}, \sigma_e^2 I_D),$$

The **main objective** is predicting a mixed effect $\tau = \boldsymbol{\ell}^t\boldsymbol{\beta} + \mathbf{m}^t\mathbf{u}$, where $\boldsymbol{\ell}$ and $\mathbf{m}$ are vectors with known elements.

- For instance, taking $\boldsymbol{\ell} = \mathbf{x}_d$ and $\mathbf{m}$ to be a vector of zeros except for a one in position d, we obtain the mean of the d-th area $\mu_d = \mathbf{x}_d^t\boldsymbol{\beta} + u_d$.

A predictor of τ can be obtained under three particular cases of model (1):

1) w_d known, $d = 1, \ldots, D$:

In this situation, a predictor of τ is typically obtained by a two-stage procedure.

- In the first stage σ_u^2 is considered to be known.
  - Then the best linear unbiased predictor (BLUP) of τ is given by $\tilde{\tau} = \boldsymbol{\ell}^t\tilde{\boldsymbol{\beta}} + \mathbf{m}^t\tilde{\mathbf{u}} = \tilde{\tau}(\sigma_u^2)$, where

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= \left(X^t\Sigma^{-1}X\right)^{-1}X^t\Sigma^{-1}\mathbf{y}\\ \tilde{\mathbf{u}} &= (\tilde{u}_1, \ldots, \tilde{u}_D)^t = \sigma_u^2\Sigma^{-1}(\mathbf{y} - X\tilde{\boldsymbol{\beta}}).\end{aligned}$$

- In the second stage σ_u^2 is regarded as unknown.
  Then, an unbiased estimator of σ_u^2 is (Prasad and Rao, 1990)

$$\hat{\sigma}_u^2 = \frac{\mathbf{y}^t P \mathbf{y} - \sigma_e^2 \,\mathrm{tr}(PW)}{D-p}, \quad \text{for } P = I_D - X(X^tX)^{-1}X^t. \tag{2}$$

- The predictor obtained by replacing σ_u^2 by $\hat{\sigma}_u^2$ in the BLUP, that is, $\hat{\tau} = \tilde{\tau}(\hat{\sigma}_u^2)$, is usually called empirical BLUP (EBLUP).
- If we denote $\hat{\Sigma} = \hat{\sigma}_u^2 I_D + \sigma_e^2 W$, the EBLUP can be written as $\hat{\tau} = \boldsymbol{\ell}^t \hat{\boldsymbol{\beta}} + \mathbf{m}^t \hat{\mathbf{u}}$, where

$$\hat{\boldsymbol{\beta}} = \left(X^t \hat{\Sigma}^{-1} X\right)^{-1} X^t \hat{\Sigma}^{-1} \mathbf{y}, \quad \hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_D)^t = \sigma_u^2 \hat{\Sigma}^{-1}(\mathbf{y} - X\hat{\boldsymbol{\beta}}). \tag{3}$$

2) $w_d = 1, d = 1, \ldots, D$: In this case the model is homoscedatic.

- The estimators $\tilde{\boldsymbol{\beta}}$ and $\tilde{\sigma}_u^2$ defined in Case 1) reduce to

$$\hat{\boldsymbol{\beta}}_0 = (X^t X)^{-1} X^t \mathbf{y} \quad \text{and} \quad \hat{\sigma}_{u0}^2 = \frac{(\mathbf{y} - X\hat{\boldsymbol{\beta}}_0)^t (\mathbf{y} - X\hat{\boldsymbol{\beta}}_0)}{D - p} - \sigma_e^2.$$

- The EBLUP of τ obtained from them is

$$\begin{aligned} \hat{\tau}_0 &= \boldsymbol{\ell}^t \hat{\boldsymbol{\beta}}_0 + \mathbf{m}^t \hat{\mathbf{u}}_0, \\ \hat{\mathbf{u}}_0 &= (\hat{u}_{10}, \ldots, \hat{u}_{D0})^t = \hat{\sigma}_{u0}^2 (\hat{\sigma}_{u0}^2 + \sigma_e^2)^{-1} (\mathbf{y} - X\hat{\boldsymbol{\beta}}_0). \end{aligned}$$

3) $w_d = w(z_d)$, $d = 1, \ldots, D$ *where the function* $w(\cdot)$ *is unknown:*

In this case the BLUP of τ is function of the unknown parameters

$$\sigma_u^2 \quad \text{and} \quad W = \text{diag}_{1 \le d \le D}(w_d);$$

that is,

$$\tilde{\tau} = \boldsymbol{\ell}^t \tilde{\boldsymbol{\beta}} + \mathbf{m}^t \tilde{\mathbf{u}} = \tilde{\tau}(\sigma_u^2, W)$$

and we need estimators of σ_u^2 and W to derive the EBLUP of τ.

- Carroll (1982) proposed a kernel-based estimator of the error variance function $\sigma^2(z)$ under the following model

$$y_d = \mathbf{x}_d^t \boldsymbol{\beta} + \sigma_d \, \varepsilon_d, \quad \varepsilon_d \overset{iid}{\sim} (0,1), \tag{4}$$

  where σ_d denotes the squared root of the variance $\sigma_d^2 = \sigma^2(z_d)$.
- Model (1) can be restated in the form of (4), where the error variance function is related to the heteroscedasticity function $w(\cdot)$ of model (1) through $\sigma^2(z) = \sigma_u^2 + w(z)\,\sigma_e^2$.
- Hence the estimator proposed by Carroll (1982), together with an estimator of σ_u^2, yield an estimator of $w(z)$.

Following this idea we propose the algorithm below for estimating the unknown parameters involved in model (1); namely, $\boldsymbol{\beta}$, σ_u^2 and the heteroscedasticity function $w(z)$.

Step 1. Calculate initial parameter estimates $\hat{\boldsymbol{\beta}}_0$ and $\hat{\sigma}^2_{u0}$ via an erroneous homoscedastic model obtained by replacing $w_d = 1$ in model (1).

Proposition. Under regularity assumptions, the estimators $\hat{\boldsymbol{\beta}}_0$ and $\hat{\sigma}^2_{u0}$ are consistent for their respective true values $\boldsymbol{\beta}$ and σ^2_u under the correct model (1).

Step 2. Using $\hat{\boldsymbol{\beta}}_0$ and $\hat{\sigma}^2_{u0}$, estimate the heteroscedasticity function $w(z)$ as

$$\hat{w}_h(z) = (\hat{\sigma}^2_h(z) - \hat{\sigma}^2_{u0})/\sigma^2_e,$$

where $\hat{\sigma}^2_h(z)$ is the estimator proposed by Carroll (1982), namely

$$\hat{\sigma}^2_h(z) = \frac{\sum_{d=1}^{D} k\left\{(z - z_{d0})/h\right\}(y_d - \mathbf{x}_d^t\hat{\boldsymbol{\beta}}_0)^2}{\sum_{d=1}^{D} k\left\{(z - z_{d0})/h\right\}}.$$

where z_{d0} denotes x_d if $p = 1$ or $\mathbf{x}_d^t\hat{\boldsymbol{\beta}}_0$ if $p > 1$, $k(\cdot)$ is a kernel function and h is a bandwidth.

- $\hat{\sigma}^2_{d,h} \triangleq \hat{\sigma}^2_h(z_d)$ denotes the estimator of the error variance σ^2_d of equation (4),
- $\hat{w}_{d,h} \triangleq \hat{w}_h(z_d)$ denotes the estimator of the heteroscedasticity weight w_d in model (1), for $d = 1, \ldots, D$.

Step 3. With the estimated matrix $\hat{W}_h = \mathrm{diag}_{1\leq d\leq D}(\hat{w}_{d,h})$ obtained from Step 2, reestimate model parameters $\boldsymbol{\beta}$ and σ^2_u.

- σ^2_u is estimated by the method of moments,

$$\hat{\sigma}^2_{u,h} = \frac{\mathbf{y}^t P \mathbf{y} - \sigma^2_e \,\mathrm{tr}(P\hat{W}_h)}{D-p}, \quad \text{for } P = I_D - X(X^tX)^{-1}X^t.$$

- Then $\boldsymbol{\beta}$ can be reestimated by the following formula

$$\hat{\boldsymbol{\beta}}_h = (X^t\hat{\Sigma}_h^{-1}X)^{-1}X^t\hat{\Sigma}_h^{-1}\mathbf{y}, \quad \text{for } \hat{\Sigma}_h = \hat{\sigma}^2_{u,h}I_D + \sigma^2_e\hat{W}_h.$$

Proposition. Under regularity conditions, the initial estimators $\hat{\boldsymbol{\beta}}_0$ and $\hat{\sigma}^2_{u0}$ defined in Step 1 satisfy

$$(a)|\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}\| = O_p(D^{-1/2}); \qquad (b)|\hat{\sigma}^2_{u0} - \sigma^2_u| = O_p(D^{-1/2}).$$

Proposition. Under regularity conditions

$$\sup_{z\in\mathcal{Z}} |\hat{w}_h(z) - w(z)| = o_p(D^{-1/4}).$$

Proposition. Under regularity conditions, the estimators $\hat{\sigma}^2_{d,h}$ and $\hat{w}_{d,h}$ obtained in Step 3 satisfy

$$(a)\left|\frac{1}{D}\sum_{d=1}^{D}\left(\hat{\sigma}^2_{d,h} - \sigma^2_d\right)\right| = O_p(D^{-1/2});\ (b)\left|\frac{1}{D}\sum_{d=1}^{D}\left(\hat{w}_{d,h} - w_d\right)\right| = O_p(D^{-1/2}).$$

Proposition. Under Reg. Ass. bandwidths minimizing AMSE and AMISE are:

(i) The local optimal bandwidth

$$h_d = \left[\frac{\sigma^4(z_d) \int k^2(t)dt}{jDk_j^2 \left[(\sigma^2)^{(j)}(z_d) \right]^2} \right]^{1/(2j+1)}, \quad d = 1, \ldots, D;$$

(ii) The global optimal bandwidth

$$h = \left[\frac{\int \sigma^4(z)dz \int k^2(t)dt}{jD\, k_j^2 \int \left[(\sigma^2)^{(j)}(z) \right]^2 dz} \right]^{1/(2j+1)}.$$

An easy-to-obtain bandwidth is

$$h^\star = \text{argmin}_{h \geq 0} \left\{ \sum_{d=1}^{D} (y_d - \hat{\mu}_{d,h})^2 \right\}.$$

- For illustration, we give the spelled out formula of the naive estimator of the mean squared error for one explanatory variable; that is, for $\mathbf{x}_d = x_d$ and $\boldsymbol{\beta} = \beta$, and for the parameter $\tau = \mu_d = x_d\beta + u_d$.
- Suppose that $W = \mathrm{diag}_{1\le d\le D}(w_d)$ is known but σ_u^2 unknown.
- Under this setup, $\hat{\sigma}_u^2$ given in (2) is an unbiased estimator of σ_u^2, with variance

$$Var(\hat{\sigma}_u^2) = \frac{2}{D}\left[\sigma_u^4 + 2\sigma_u^2\sigma_e^2 + \frac{\sigma_e^4}{D}\sum_{d=1}^{D} w_d^2\right] + o(D^{-1}),$$

see Prasad and Rao (1990), p. 167.

- The EBLUP of μ_d is $\hat{\mu}_d = x_d\hat{\beta} + \hat{u}_d$, where $\hat{\beta}$ and $\hat{u}_d$ are given in (3).

The Prasad-Rao approximation of the mean squared error of $\hat{\mu}_d$ is

$$MSE(\hat{\mu}_d) \approx g_{1d}(\sigma_u^2, W) + g_{2d}(\sigma_u^2, W) + g_{3d}(\sigma_u^2, W), \tag{5}$$

$$\begin{aligned} g_{1d}(\sigma_u^2, W) &= \frac{\sigma_u^2 \sigma_e^2 \, w_d}{\sigma_u^2 + \sigma_e^2 \, w_d}, \\ g_{2d}(\sigma_u^2, W) &= \frac{\sigma_e^4 \, w_d^2 \, x_d^2}{(\sigma_u^2 + \sigma_e^2 \, w_d)^2} \left[\sum_{d=1}^{D} x_d^2 (\sigma_u^2 + \sigma_e^2 \, w_d)^{-1} \right]^{-1}, \\ g_{3d}(\sigma_u^2, W) &= \frac{\sigma_e^4 \, w_d^2 \, Var(\hat{\sigma}_u^2)}{(\sigma_u^2 + \sigma_e^2 \, w_d)^3}. \end{aligned}$$

Thus, (5) can be used as a naive approximation of the mean squared error of the predictor $\hat{\mu}_{d,h} = x_d \hat{\beta}_h + \hat{u}_{d,h}$. An estimator of $MSE(\hat{\mu}_{d,h})$ is

$$mse(\hat{\mu}_{d,h}) = g_{1d}(\hat{\sigma}_{u,h}^2, \hat{W}_h) + g_{2d}(\hat{\sigma}_{u,h}^2, \hat{W}_h) + 2 g_{3d}(\hat{\sigma}_{u,h}^2, \hat{W}_h).$$

Alternatively, a bootstrap procedure is:

Step 1. Take a "pilot" bandwidth g with $g > h$. Fit the model to the initial sample $\mathbf{y}$ by using the introduced algorithm. Obtain model parameter estimates $\hat{W}_g = \hat{W}_g(\mathbf{y})$, $\hat{\sigma}^2_{u,g} = \hat{\sigma}^2_{u,g}(\mathbf{y})$ and $\hat{\boldsymbol{\beta}}_g = \hat{\boldsymbol{\beta}}_g(\mathbf{y})$.

Step 2. Generate a vector $\mathbf{T}_1$ with D independent copies of a variable T_1 with $E(T_1) = 0$ and $E(T_1^2) = 1$. Construct the vector $\mathbf{u}^* = \hat{\sigma}_{u,g}\mathbf{T}_1$, with mean $\mathbf{0}_D$ and covariance matrix $\hat{\sigma}^2_{u,g} I_D$.

Step 3. Generate a vector $\mathbf{T}_2$ with D independent copies of a random variable T_2, independent of T_1, with $E(T_2) = 0$ and $E(T_2^2) = E(T_2^3) = 1$. Construct the vector $\mathbf{e}^* = \sigma_e \mathbf{T}_2$, with mean $\mathbf{0}_D$ and covariance matrix $\sigma_e^2 I_D$.

Step 4. With the elements of the incidence matrix X known, generate bootstrap data $\mathbf{y}^*$ from the model

$$\mathbf{y}^* = X\hat{\boldsymbol{\beta}}_g + \mathbf{u}^* + \hat{W}_g^{1/2}\mathbf{e}^*. \tag{6}$$

Let E_*, Var_* and MSE_* denote respectively expectation, variance and mean squared error under the bootstrap model (6) given the initial data $\mathbf{y}$. It holds that

$$E_*(\mathbf{y}^*) = X\hat{\boldsymbol{\beta}}_g, \quad \text{and} \quad Var_*(\mathbf{y}^*) = \hat{\sigma}^2_{u,g} I_D + \sigma_e^2 \hat{W}_g,$$

Step 5. Using the bootstrap data $\mathbf{y}^*$ and the introduced algorithm with bandwidth parameter $h < g$, calculate bootstrap parameter estimates

$$\begin{aligned} \hat{W}_h^* &= \hat{W}_h(\mathbf{y}^*),\ \hat{\sigma}^{2*}_{u,h} = \hat{\sigma}^2_{u,h}(\mathbf{y}^*),\ \hat{\boldsymbol{\beta}}_h^* = \hat{\boldsymbol{\beta}}_h(\mathbf{y}^*), \\ \hat{\mathbf{u}}_h^* &= \hat{\mathbf{u}}_h(\mathbf{y}^*) = (\hat{u}^*_{1,h}, \ldots, \hat{u}^*_{D,h})^t. \end{aligned}$$

From them, construct the bootstrap EBLUP

$$\hat{\tau}_h^* = \boldsymbol{\ell}^t \hat{\boldsymbol{\beta}}_h^* + \mathbf{m}^t \hat{\mathbf{u}}_h^*.$$

- The **bootstrap estimate** of $MSE(\hat{\tau}_h)$ is $MSE_*(\hat{\tau}_h^*) = E_*(\hat{\tau}_h^* - \tau^*)^2$.
- This quantity can be approximated by Monte Carlo, repeating Steps 2-5 for $b = 1, \ldots, B$.
- Let $\mathbf{u}^{*(b)}$ be the generated vector of random effects, $\tau^{*(b)} = \boldsymbol{\ell}^t \hat{\boldsymbol{\beta}}_g + \mathbf{m}^t \mathbf{u}^{*(b)}$ the true value of the parameter and $\hat{\tau}_h^{*(b)}$ the estimate of $\tau^{*(b)}$ in the b-th replication.
- The **Monte Carlo approximation** of $MSE_*(\hat{\tau}_h^*)$ is

$$mse_{*1}(\hat{\tau}_h) = B^{-1} \sum_{b=1}^{B} \left(\hat{\tau}_h^{*(b)} - \tau^{*(b)} \right)^2 . \tag{7}$$

A **bias corrected bootstrap estimator** is

$$
\begin{aligned}
mse_{*2}(\hat{\mu}_{d,h}) &= \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mu}_{d,h}^{*(b)} - \tilde{\mu}_{d}^{*(b)}\right)^2 + 2\left[g_{1d}(\hat{\sigma}_{u,h}^2, \hat{W}_h) + g_{2d}(\hat{\sigma}_{u,h}^2, \hat{W}_h)\right] \\
&- \frac{1}{B}\sum_{b=1}^{B}\left[g_{1d}(\hat{\sigma}_{u,h}^{2*(b)}, \hat{W}_h^{*(b)}) + g_{2d}(\hat{\sigma}_{u,h}^{2*(b)}, \hat{W}_h^{*(b)})\right],
\end{aligned}
$$

where $\tilde{\mu}_d^{*(b)}$ and $\hat{\mu}_{d,h}^{*(b)}$ are respectively the BLUP and EBLUP of the bootstrap mean $\mu_d^{*(b)} = x_d\hat{\beta}_g + u_d^{*(b)}$ calculated with the b-th bootstrap sample, that is,

$$
\begin{aligned}
\tilde{\mu}_d^{*(b)} &= \tilde{\mu}_d(\mathbf{y}^{*(b)}) = \mathbf{x}_d^t\tilde{\boldsymbol{\beta}}(\mathbf{y}^{*(b)}) + \tilde{u}_d(\mathbf{y}^{*(b)}), \\
\hat{\mu}_{d,h}^{*(b)} &= \hat{\mu}_{d,h}(\mathbf{y}^{*(b)}) = \mathbf{x}_d^t\hat{\boldsymbol{\beta}}_h(\mathbf{y}^{*(b)}) + \hat{u}_{d,h}(\mathbf{y}^{*(b)}).
\end{aligned}
$$

A first simulation experiment was planned for analyzing the performance of

- the kernel-based estimator of the variance function $\hat{\sigma}_h^2(\cdot)$, and
- the estimators of the small area means $\hat{\mu}_{d,h}$.

We were also interested in comparing the results obtained with

- the local optimal bandwidth, and
- the simple global bandwidth $h_0 = \mathrm{range}(x_d)/2$.

For this, $I = 10^5$ samples $\mathbf{y}^{(i)}$, $i = 1, \ldots, I$, were generated as follows:

- The values of the explanatory variable were taken as $\mathbf{x}_d = x_d = -0.5 + d/D$, $d = 1, \ldots, D$.

- Three families of standard deviations $\nu_d = \sigma_e^2 w_d$ were considered (Carroll, 1982),

  A) $\nu_d^{1/2} = [a_{11} + a_{12}(\alpha_{10} + \alpha_{11} x_d)^2]^{1/2}$, with $a_{11} = 0.15$, $a_{12} = 0.03$, $\alpha_{10} = 5$ and $\alpha_{11} = 6$;

  B) $\nu_d^{1/2} = a_{21} \exp\{a_{22} |\alpha_{20} + \alpha_{21} x_d|\}$, with $a_{21} = 0.15$, $a_{22} = 0.04$, $\alpha_{20} = 50$ and $\alpha_{21} = 60$;

  C) $\nu_d^{1/2} = a_{31} \exp\{a_{32}(\alpha_{30} + \alpha_{31} x_d)^2\}$, with $a_{31} = 0.15$, $a_{32} = 1/1600$, $\alpha_{30} = 50$ and $\alpha_{31} = 60$.

- The values of the response variable y_d were generated from model (1) with $\boldsymbol{\beta} = \beta = 60$, $\sigma_u^2 = 9$ and one of the families A–C of error variances.
- The simulations were reproduced for sample sizes $D = 100$ and $D = 400$ in order to observe the asymptotic behavior.

- For each sample i, we computed the estimated error variances and the predictors of the small area means $\mu_d = x_d\beta + u_d$ under the following four scenarios:
  - *Scenario 1.* The true error variances ν_d are known;
  - *Scenario 2.* The model is erroneously regarded as homoscedastic, and therefore a homoscedastic model is fitted. We have taken as constant error variance, the mean of the true variances $\sigma_e^2 = \sum_{d=1}^{D} \nu_d / D$.
  - *Scenario 3.* The heteroscedasticity function $w(x) = \nu(x)/\sigma_e^2$ is unknown and it is estimated with the local optimal bandwidth and the kernel function

$$k(v) = \begin{cases} 3(1-|v|)^2/2 & \text{if } |v| < 1; \\ 0 & \text{if } |v| \geq 1. \end{cases} \tag{8}$$

  - *Scenario 4.* The same as Scenario 3, but with global bandwidth $h_0 = \text{range}(x_d)/2 = 1/2$.

- For each small area d, let us denote $\hat{\nu}_d^{k(i)}$ and $\hat{\mu}_d^{k(i)}$ respectively to the error variance and the small area estimator obtained with Monte Carlo sample i under Scenario k, $k = 1, 2, 3, 4$.
- The output of the simulation study is the empirical expectation of $\hat{\nu}_d^{k(i)}$ and the empirical mean squared error of $\hat{\mu}_d^{k(i)}$, that is,

$$\bar{\hat{\nu}}_d^k = \frac{1}{I}\sum_{i=1}^{I} \hat{\nu}_d^{k(i)}, \quad EMSE(\hat{\mu}_d^k) = \frac{1}{I}\sum_{i=1}^{I} \left(\hat{\mu}_d^{k(i)} - \mu_d^{(i)}\right)^2, \quad k = 1, 2, 3, 4.$$

- In Figures 2-4 we plot the values of $EMSE(\hat{\mu}_d^k)$ Under the 4 scenarios.
- In Figures 5-7 we plot the true error variances ν_d (labelled with *var1*), the (constant) error variance σ_e^2 (labelled with *var2*) and the empirical expectations of the error variance estimators $\bar{\hat{\nu}}_d^k$, $k = 3, 4$, under Scenarios 3 and 4 respectively (labelled with *var3* and *var4*).

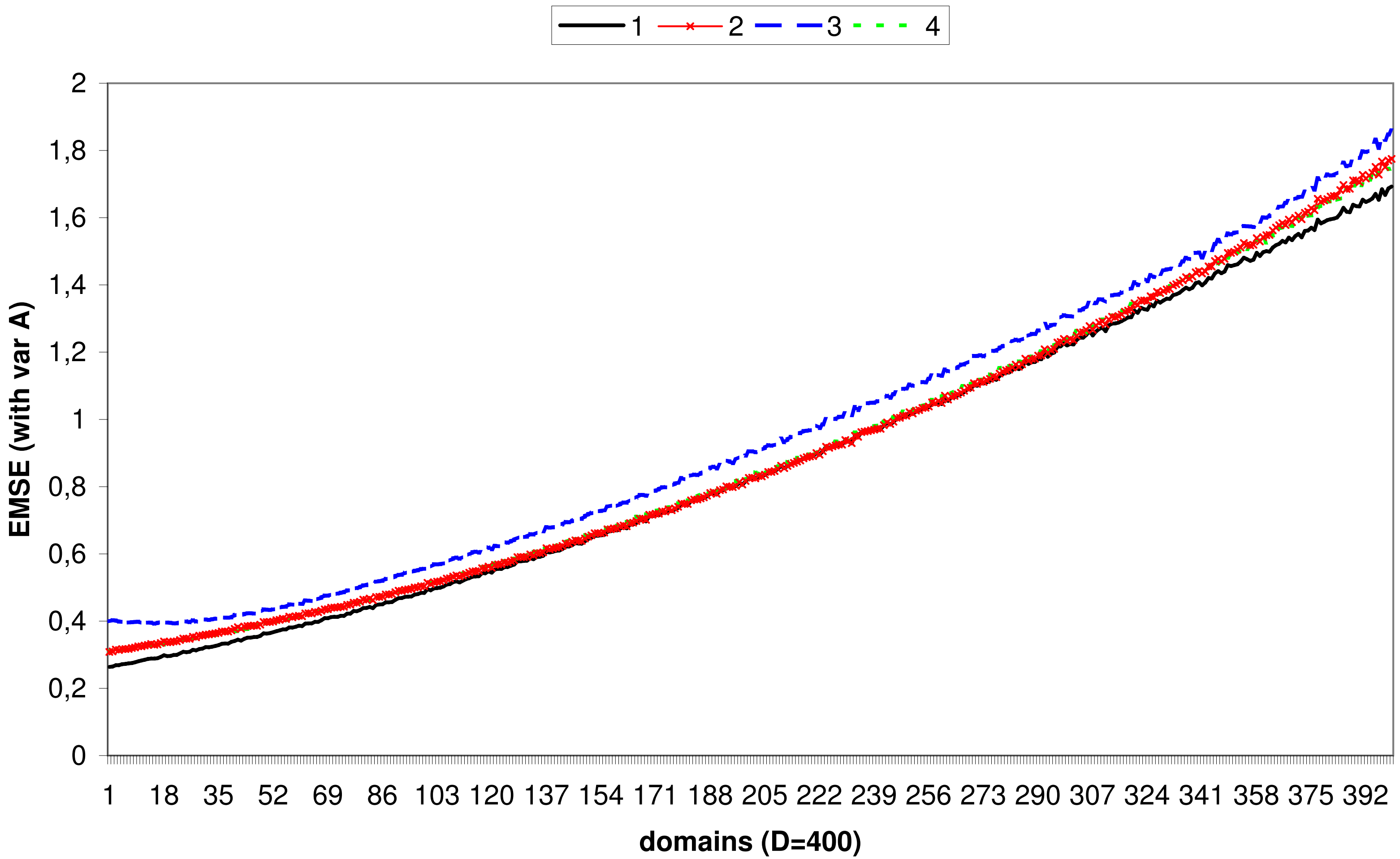

Figure 2. $EMSE(\hat{\mu}_d^k)$, $k = 1, 2, 3, 4$, for family A and for 400.

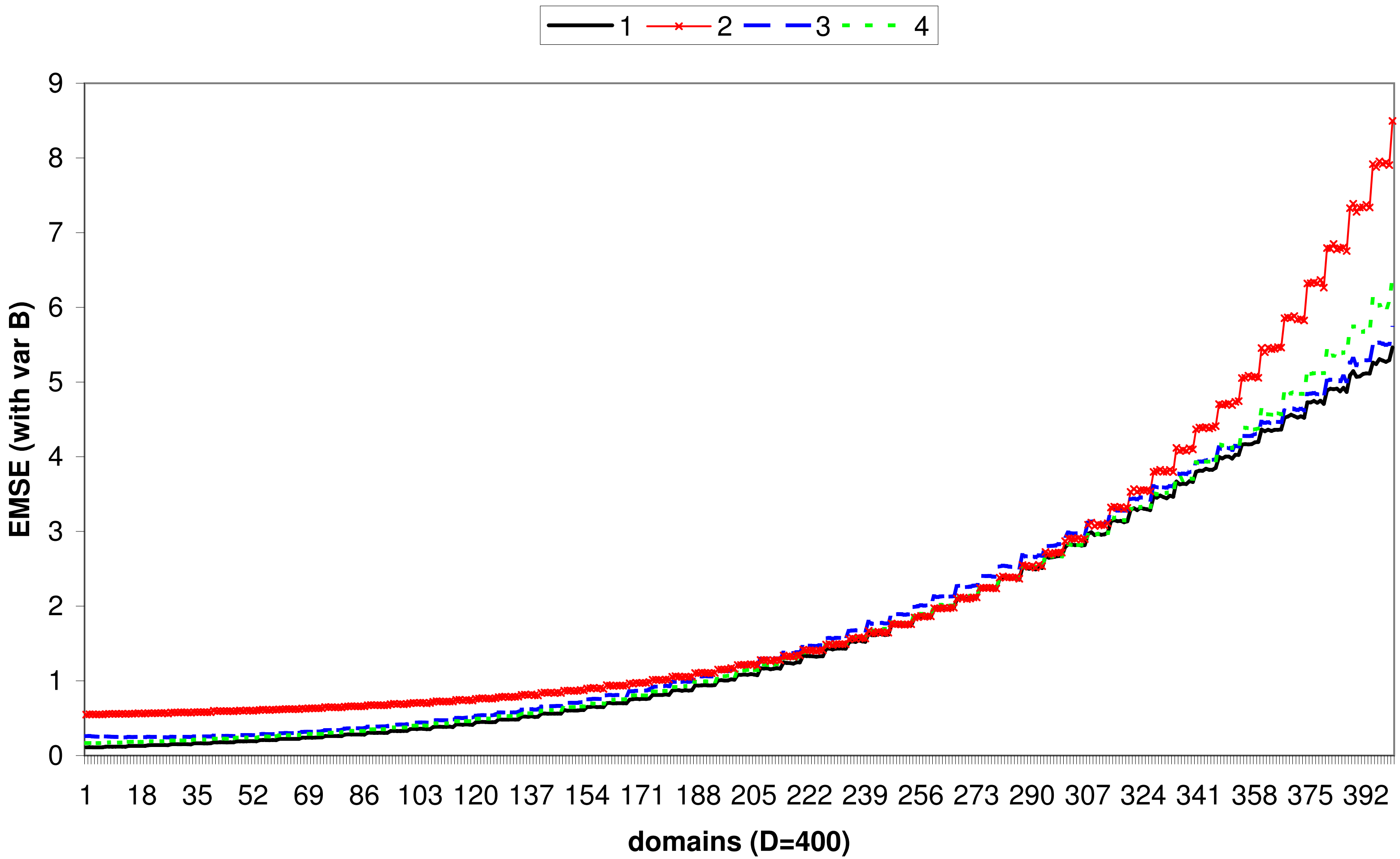

Figure 3. $EMSE(\hat{\mu}_d^k)$, $k = 1, 2, 3, 4$, for family B and for 400.

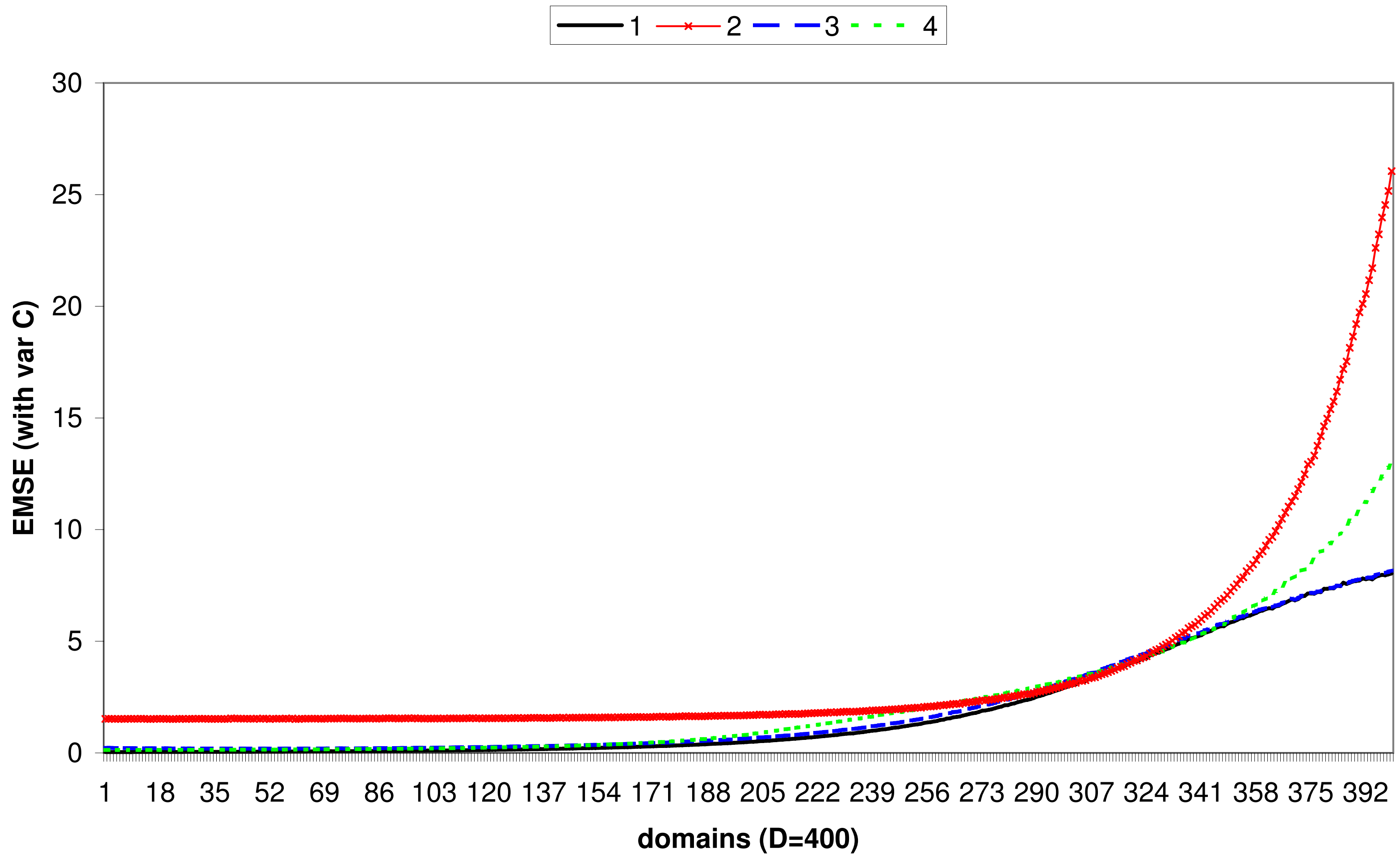

Figure 4. $EMSE(\hat{\mu}_d^k)$, $k = 1, 2, 3, 4$, for family C and for 400.

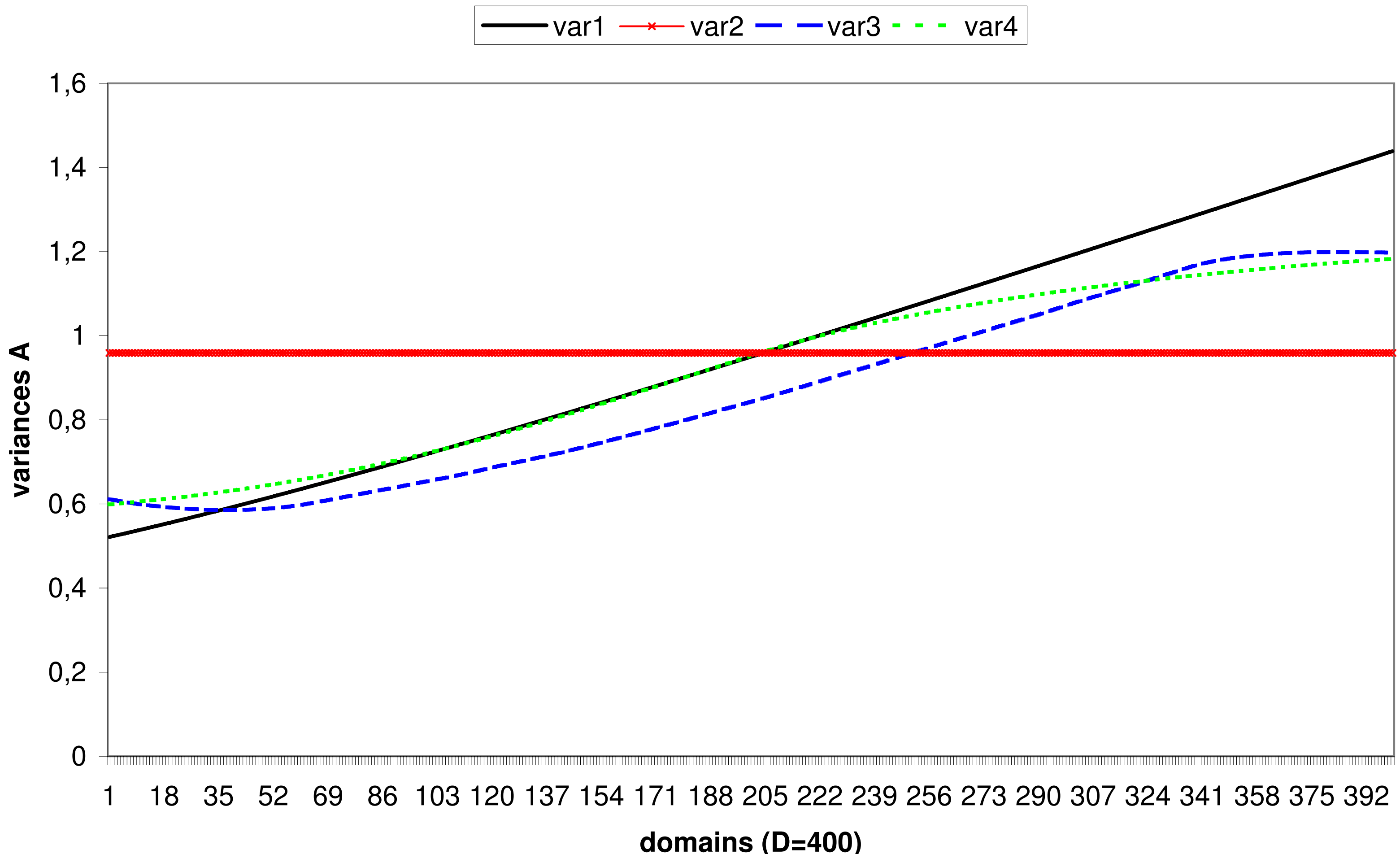

Figure 5. $\hat{\bar{\nu}}_d^k$, $k = 1, 2, 3, 4$ for family A and for $D = 400$.

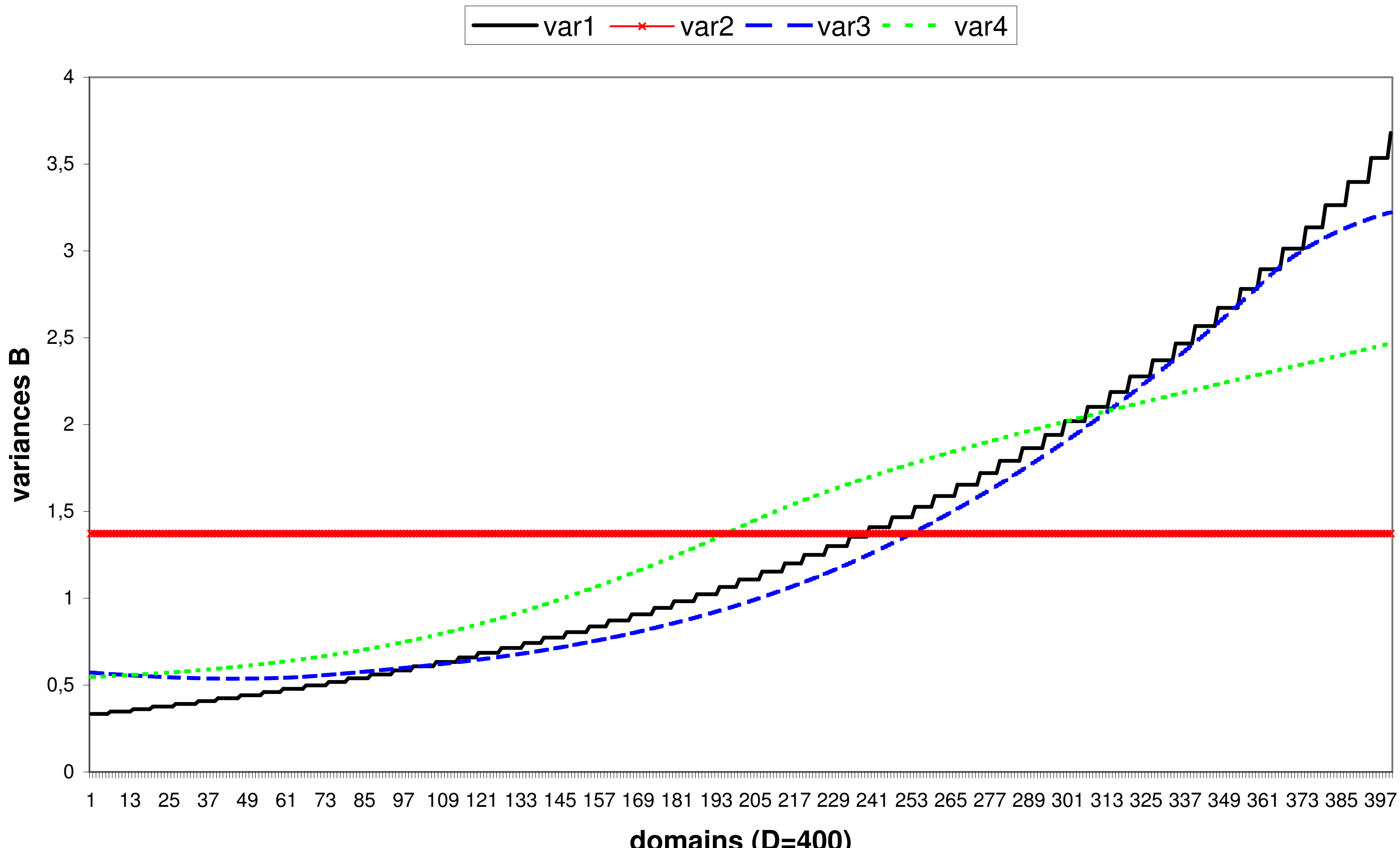

Figure 6. $\hat{\bar{\nu}}_d^k$, $k = 1, 2, 3, 4$ for family B and for $D = 400$.

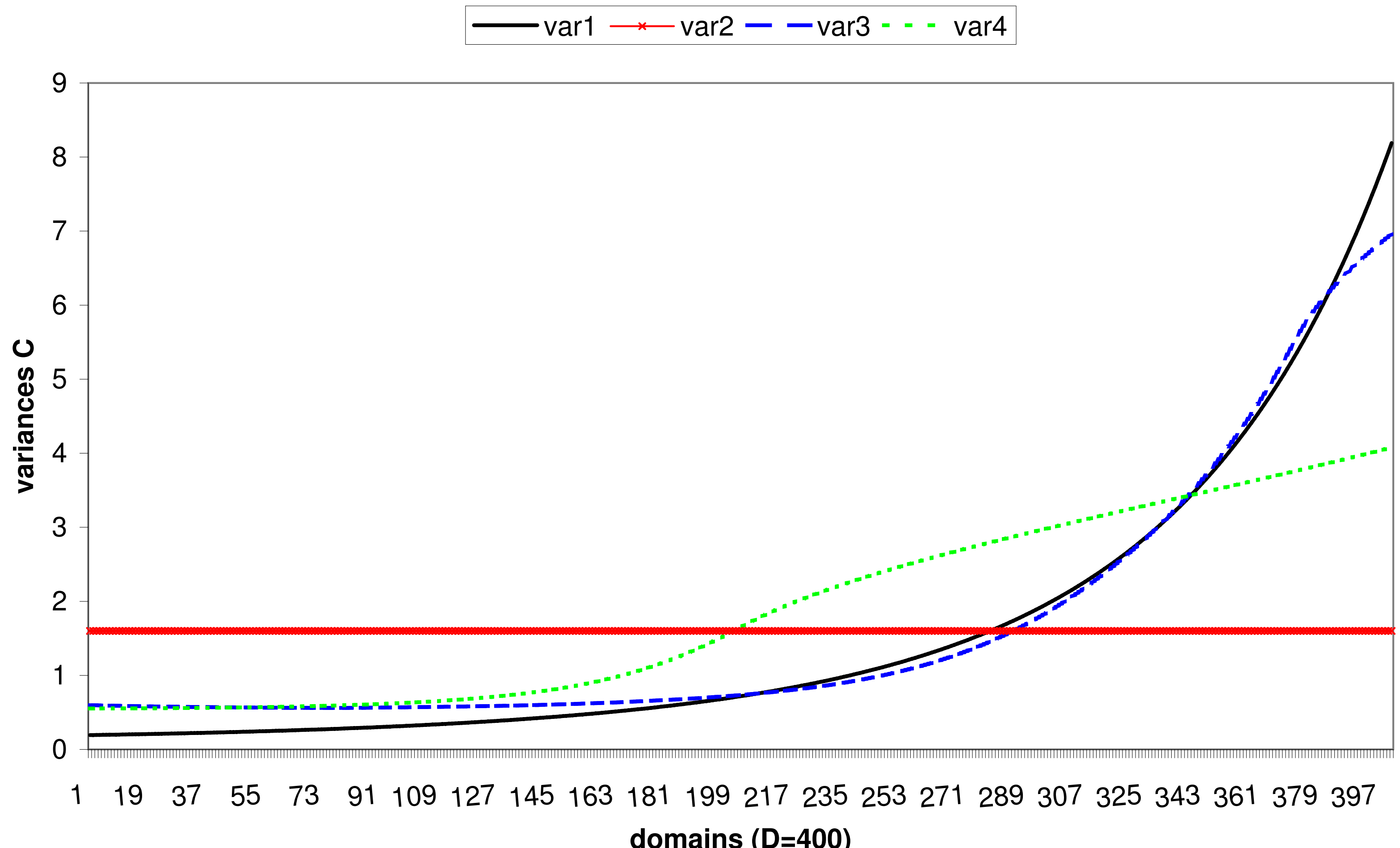

Figure 7. $\hat{\tilde{\nu}}_d^k$, $k = 1, 2, 3, 4$ for family C and for $D = 400$.

- The **second simulation experiment** was designed for comparing under scenario 3 the two bootstrap-based estimators of $MSE(\hat{\mu}_{d,h})$ and the naive analytical estimator derived from Prasad-Rao's formula, with the empirical values $EMSE(\hat{\mu}_{d,h})$.
- In this experiment, $I = 10^3$ Monte Carlo samples were generated from model (1) with $\beta = 60$ and $\sigma_u^2 = 9$ as before.
- With each Monte Carlo sample, model parameter estimates were obtained by using the introduced algorithm with the local optimal bandwidth h_d and the kernel given in simulation 1.
- From them, the bootstrap model was constructed and $B = 10^3$ bootstrap samples were generated.
- Following the recommendation of Härdle and Marron (1991), the bootstrap model was constructed using a local "pilot" bandwidth g_d greater than h_d; concretely, we took $g_d = 2h_d$, $d = 1, \ldots, D$.
- Then the model was fitted again to each bootstrap sample, and finally $MSE(\hat{\mu}_{d,h})$ is estimated as a mean over bootstrap replicates.

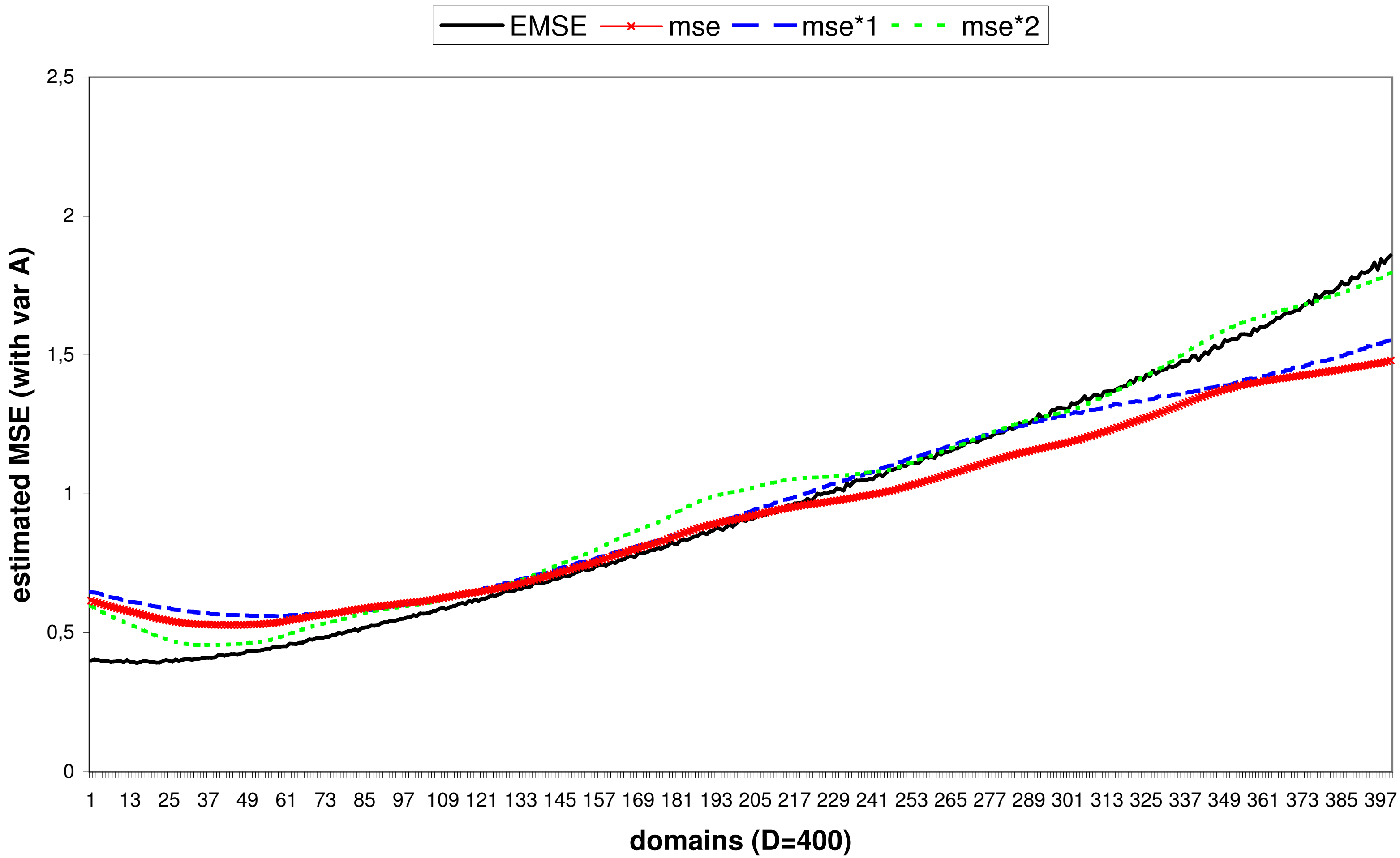

Figure 8. $EMSE_d$, mse_d, mse_d^{*1} and mse_d^{*2}, under family A, for $D = 400$.

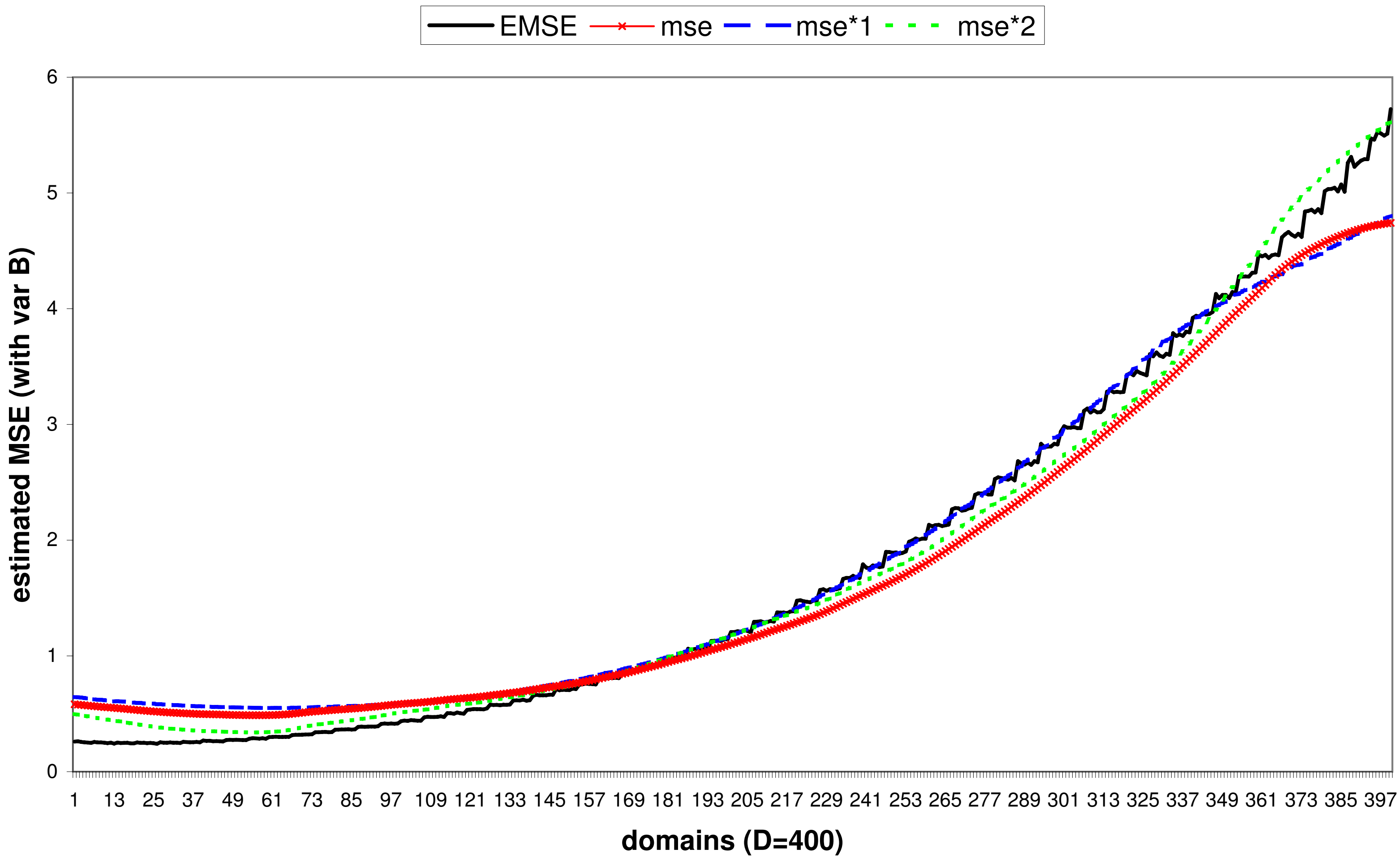

Figure 9. $EMSE_d$, mse_d, mse_d^{*1} and mse_d^{*2}, under family B, for $D = 400$.

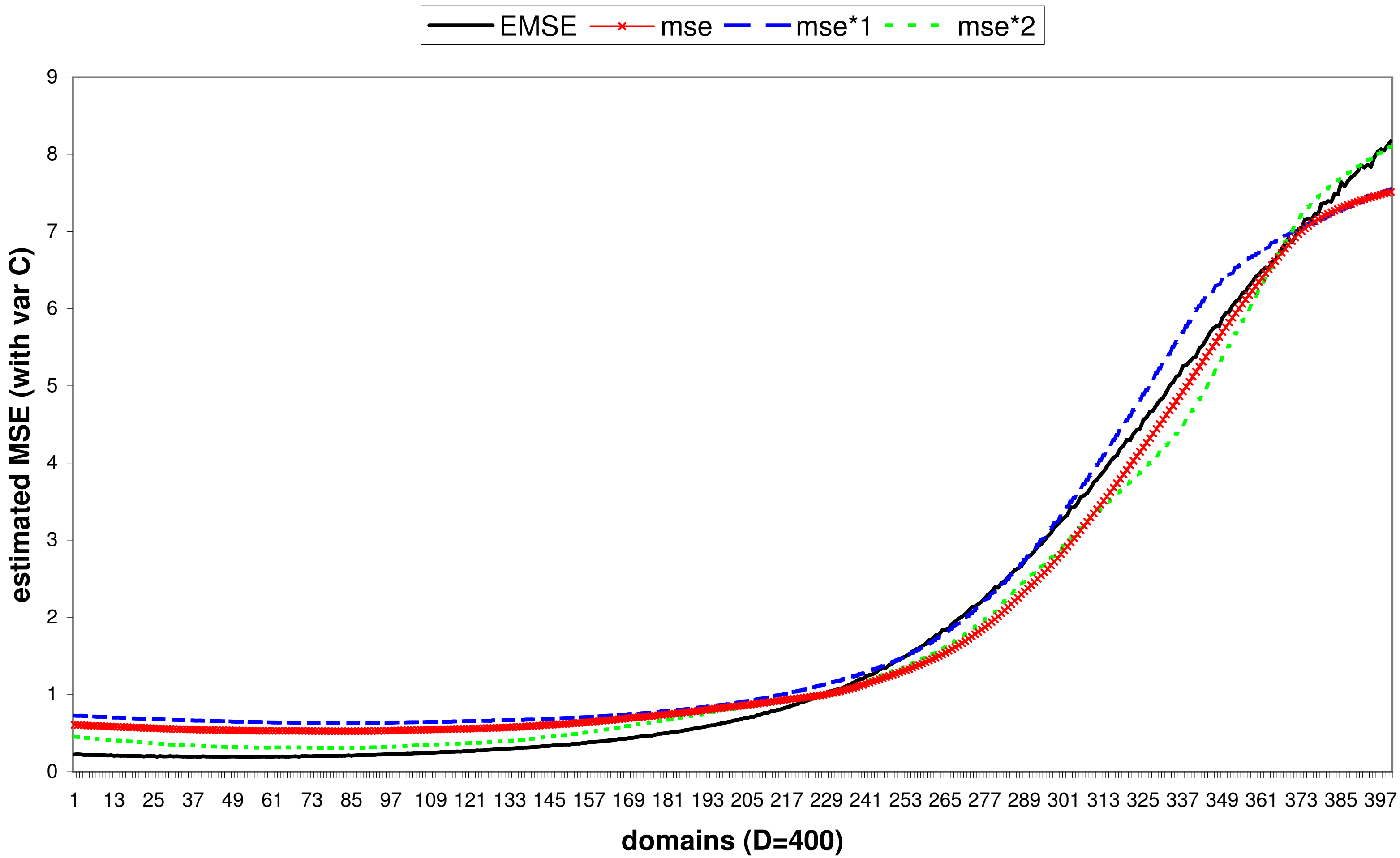

Figure 10. $EMSE_d$, mse_d, mse_d^{*1} and mse_d^{*2}, under family C, for $D = 400$.

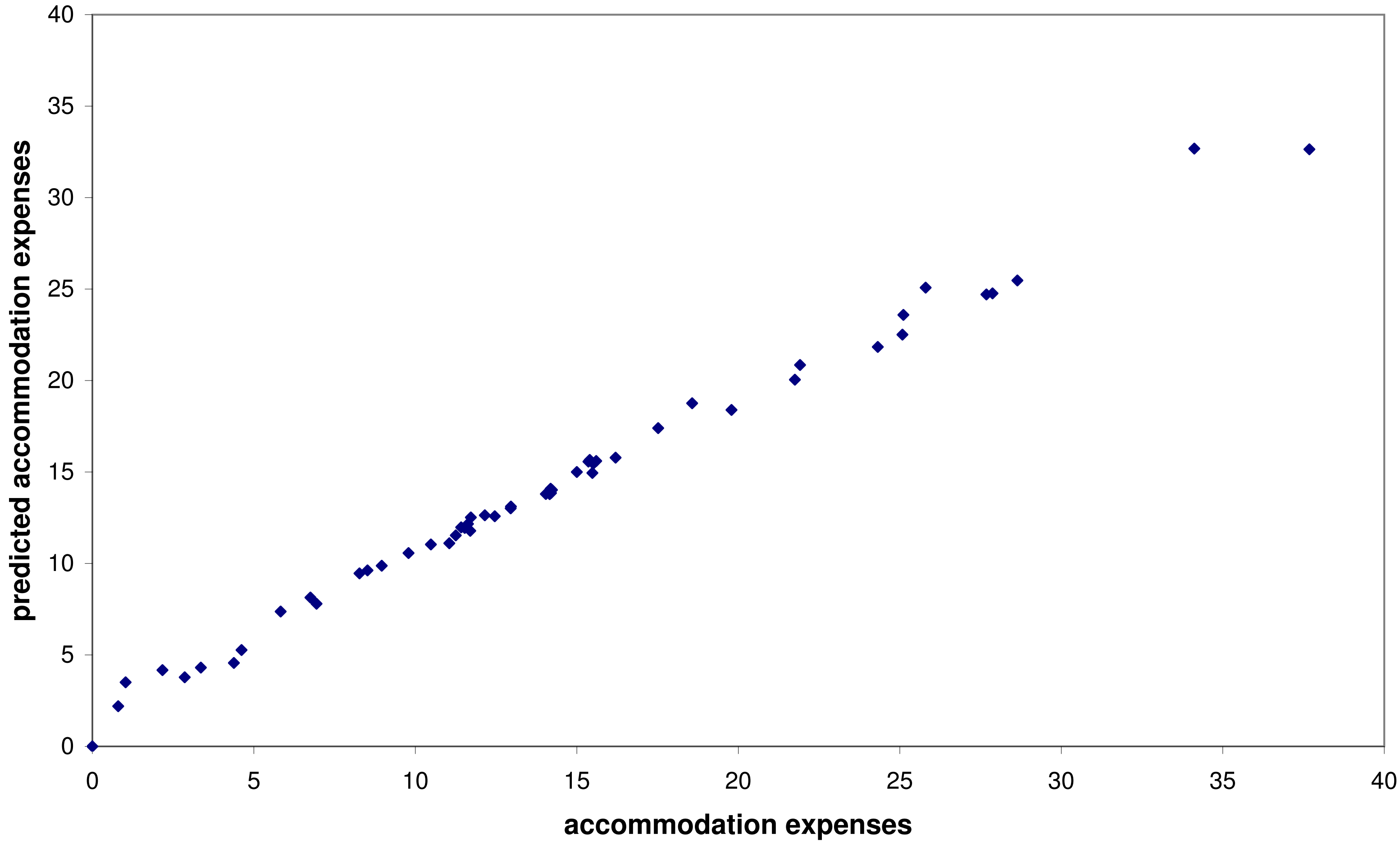

Figure 11. Predicted accommodation expenses $\hat{\mu}_{d,h^\star}$ versus accommodation expenses y_d.

d	x_d	y_d	S^2_{yd}	$\hat{\mu}_{d,h^\star}$	mse_d	d	x_d	y_d	S^2_{yd}	$\hat{\mu}_{d,h^\star}$	mse_d
1	287	25.08	12.29	22.51	11.18	28	543	11.53	8.04	11.94	11.52
2	966	8.96	1.87	9.87	7.66	29	1047	34.12	61.58	32.68	4.34
3	94	15.00	17.31	15.00	0.01	30	649	28.64	19.72	25.47	12.06
4	663	11.63	4.30	12.16	12.04	31	201	15.48	15.86	14.95	8.19
5	437	12.46	3.71	12.58	12.59	32	428	10.48	10.68	11.04	12.58
6	2347	18.57	2.97	18.76	2.86	33	328	1.03	0.06	3.50	12.01
7	891	15.51	7.05	15.46	8.94	34	597	21.75	18.83	20.04	11.88
8	872	12.15	2.46	12.64	9.16	35	182	0.80	0.64	2.20	7.01
9	858	8.52	3.49	9.62	9.35	36	644	27.68	11.26	24.71	12.07
10	483	12.95	3.05	13.01	12.47	37	521	27.87	23.48	24.77	11.93
11	704	16.20	6.89	15.78	12.11	38	82	11.05	13.64	11.10	11.19
12	404	6.75	6.60	8.14	12.67	39	108	4.38	9.41	4.56	1.42
13	517	9.79	5.06	10.57	11.99	40	258	11.70	6.41	11.79	10.42
14	1241	11.72	2.13	12.52	8.62	41	604	8.27	6.19	9.46	11.92
15	483	24.32	9.63	21.83	12.47	42	579	12.96	9.54	13.11	11.70
16	166	14.04	7.38	13.79	5.82	43	280	2.17	1.13	4.17	11.01
17	366	5.83	4.48	7.37	12.57	44	1683	25.80	6.40	25.08	5.84
18	165	2.86	1.85	3.78	5.72	45	904	21.91	6.47	20.85	8.80

Table B.1. Area-level data, sampling errors of direct estimators, predicted values and estimated mean squared errors.

Linear mixed models

- We study the estimation of the error variances in heteroscedastic Fay-Herriot models.
- We propose a non-parametric estimator of the error variance, giving its order of consistency and showing empirically its behavior.
- We give local and global optimal bandwidths for the non-parametric variance estimator by minimizing the corresponding asymptotic mean squared error and asymptotic mean integrated squared error.
- We address the problem of bandwidth selection in practice.

In **small-area estimation**, a common target is predicting mixed effects, for instance the means of the areas.

- We propose a fitting algorithm for obtaining an EBLUP under a Fay-Herriot model with unknown heteroscedasticity weights.
- We provide three estimators of the mean squared error.
  - an analytical approximation derived from the results of Prasad and Rao (1990), and
  - two more estimators based on bootstrapping.
- In the simulation study we have seen that the bias-corrected bootstrap estimator is often less biased than the other two.

Finally, we apply the introduced methodology to **small area estimation of tourist expenditure in Galicia in 2004**.

- We show that in comparison with direct estimators, the new model-based estimators preserve the bias small, and at the same time reduce the variance.

Carroll, R.J. (1982). Adapting for heteroscedasticity in linear models, *The Annals of Statistics*, 10, 1224–1233.

Fay, R. E. and Herriot, R. A. (1979). Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**, 269–277.

González-Manteiga, W., Lombardía, M.J., Molina, I., Morales, D. and Santamaría, L. (2007). Estimation of the Mean Squared Error of Predictors of Small Area Linear Parameters Under a Logistic Mixed Model. *Computational Statistics & Data Analysis*, **51**, N. 5, 2720-2733.

Jiang, J. and Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation. *Test*, **15**, 111-207.

Härdle, W. and Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*, **19**, 778–796.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under selection model. *Biometrics*, **31**, 423–427.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163–171.

Rao, J. N. K. (2003). *Small area estimation*, Wiley.