UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Center For International Economics

Working Paper Series

Working Paper No. 2022-04

An iterative plug-in algorithm for P-Spline regression

Sebastian Letmathe and Yuanhua Feng

September 2022

Center For International Economics

An iterative plug-in algorithm for P-Spline regression

Sebastian Letmathe and Yuanhua Feng

Faculty of Business Administration and Economics, Paderborn University

Abstract

This paper proposes a new IPI- (iterative plug-in) rule for optimal smoothing for penalised splines with truncated polynomials. The IPI is based on a closed-form approximation to the optimal smoothing parameter. In contrast to a DPI- (direct plug-in) approach the current algorithm is fully automatic and self-contained. Our proposal is a fixpoint-search procedure and the resulting smoothing parameter is (theoretically) independent of the initial value. Like the DPI, the IPI-rule can be employed as a refining stage in order to improve the quality of other selection methods, e.g. Mallow's Cp, Cross Validation or Residual Maximum Likelihood. Some numerical features of P-Splines as well as the performance of the IPI-algorithm are examined in detail through a simulation study. Our results reveal that our proposal works very well. Practical relevance of the IPI is illustrated by different data examples.

Key Words: P-Splines, smoothing parameter, iterative plug-in, simulation

JEL Codes: C14, C51

1 Introduction

Non-parametric smoothing methods, especially penalised spline (P-spline) smoothing, have gained more attention during the last decades due to advancing technology as well as the increasing complexity and scale of Big Data. P-spline regression offers an appropriate alternative to parametric and to more common non-parametric methods like kernel regression (Nadaraya, 1964; Watson, 1964) or local polynomial regression (Cleveland, 1979). So far, the application of P-splines has mainly occurred in the field of natural sciences and has rarely been applied in the context of empirical economic and financial research. P-spline estimation was introduced by Parker and Rice (1985) as well as by O'sullivan et al. (1986), who had the idea to use a set of basis functions in combination with a penalty controlling for model complexity. Eilers and Marx (1996) followed their approach and illustrated that penalised spline regression is an applicable and flexible method. It can be considered as a compromise between regression splines without a penalty and fewer knots than the sample size and smoothing splines with knots or basis functions being equal to the number of observations (Schwarz and Krivobokova, 2016). Analogue to kernel regression and local polynomial regression the smoothness strongly depends on the smoothing parameter, which controls the trade-off between integrity of the data and complexity of the model. Consequently, the main challenge in non-parametric regression is the selection of an optimal smoothing parameter. Well known criteria for determining this parameter are for instance Mallow's Cp (M_{C_p}) (Mallows, 1973), Akaike information criterion (Akaike, 1974), Cross Validation (CV) (Mosier, 1951) or Generalised Cross Validation (GCV) (Wahba, 1977; Craven and Wahba, 1978). For an illustration of the application of some of those criteria in P-spline regression please see Wager et al. (2007), Kauermann (2005) and Eilers et al. (2015). Theoretical results of the P-spline estimator were presented by Aerts et al. (2002), Li and Ruppert (2008), Claeskens et al. (2009) and Wang et al. (2011). Claeskens et al. (2009) investigated both asymptotic scenarios close to regression splines and smoothing splines. Based on their findings the authors recommend to reduce the amount of knots in order to obtain a smaller mean squared error. Krivobokova (2013) conducted a comparative simulation study of the asymptotic properties of two penalised spline estimators, which are based on M_{C_p} and maximum likelihood (ML). Their results show, that these estimators usually have a relatively large variance. The asymptotic behaviour of the P-spline estimator is still very unexplored and the development of a fast, simple and reliable method with a small amount of user intervention to select the appropriate smoothing parameter is of utmost importance.

The main objective of this paper is the development of an iterative plug-in (IPI) algorithm for P-splines for cross sectional data. Ruppert et al. (1995) proposed an effective bandwidth selector for local least squared regression, namely the direct plug in (DPI) method, which is a special case of the plug-in (PI) method. Wand (1999) adapted the approach of Ruppert et al. (1995) to P-spline regression and provided a closed-form asymptotic approximation to the optimal smoothing parameter. Based on this approximation Wand (1999) derived a fast and simple DPI rule to determine the smoothing parameter directly from the data. This paper follows the idea of Gasser et al. (1991) and extends the DPI-rule developed by Wand (1999). Following Wand, 1999 three different approximations are derived. For each approximation we propose an IPI-rule, namely IPI_A, IPI_B and IPI_C. For IPI_A an adaptation of equation (4) in Wand (1999) is implemented. IPI_C is based on a simplified version of equation (4) in Wand (1999) and IPI_B is a combination of both algorithms. In order to assess the goodness and performance of our proposals, we conduct a comprehensive simulation study where we apply the estimators in 36 different cases. For the estimation of the variance of the error term, we use a difference based variance estimator proposed by Gasser et al. (1986). According to our results IPI_B is to be recommended as it clearly outperforms IPI_A and IPI_C. As the performance of IPI_C is generally very poor and unstable in some cases, results obtained by means of IPI_C are omitted. Moreover, our proposal is applied to various real data examples and it is found that the IPI delivers satisfying results in this context as well.

In Section 1 the model is introduced as well as asymptotic results of the P-Spline estimator are presented. The IPI-algorithms are introduced in Section 2. In Section 3 a simulation study is conducted and the results are analysed. The application to real data examples is carried out in section 4. Concluding remarks are given in Section 5.

2 The model and asymptotics

In this section the underlying model is defined. Moreover, asymptotic properties of the P-spline estimator are illustrated and a closed-form asymptotic approximation to the optimal smoothing parameter based on the proposal of Wand (1999) is presented. We consider the following fixed design non-parametric regression model

$$y_i = m(x_i) + \epsilon_i, \tag{1}$$

where $m(\cdot)$ is an unknown smooth function. ϵ_i are assumed to be i.i.d. (independent and identically distributed) random variables with $E(\epsilon_i) = 0$ and a constant variance $var(\epsilon_i) = \sigma_\epsilon^2$. The observed response in this model is given by $y_i, i = 1, ..., n$, with the standardized fixed design points $x_1 < ... < x_n$, such that $x_i = (i - 0.5)/n$ and $m : [0, 1] \to \mathbb{R}$.

2.1 Penalised spline estimation of the trend

In this paper the unknown smooth function is estimated by P-spline regression with truncated polynomial basis functions $(x_i - \kappa_k)_+^p$, with a set of K equidistant knots $\kappa_1, ..., \kappa_K$. Alternatively, one could use a B-spline basis (Eilers and Marx, 1996) or the Demmler-Reinsch basis (Demmler and Reinsch, 1975). We chose this kind of basis functions in order to avoid penalising the polynomial coefficients. Let p be an odd integer and $r = p + 1$ and let $m(\cdot)$ be a r-times continously differentiable function. We can divide the unknown smooth function into

$$m(x) = \sum_{j=0}^{p} \beta_j x^j + \sum_{j=p+1}^{p+\kappa} \beta_j (x - \kappa_j)_+^p b_a(x),$$

where $b_a(x)$ is defined as the approximation bias. $m(\cdot)$ can now be estimated by minimising the penalised least squares

$$\sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x^j - \sum_{j=p+1}^{p+K} \beta_j (x - K_j)_+^p \right)^2 + n\lambda^{2r} \sum_{j=1}^{K} \beta_{p+j}^2, \tag{2}$$

where λ denotes the penalty- or smoothing parameter. This minimisation problem can be expressed in matrix notation. Let $Y = (y_1, ..., y_n)^T$,

$$Z = \begin{pmatrix} 1 & x_1 & \cdots & x_1^p & (x_1 - \kappa_1)_+^p & \cdots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \\ 1 & x_n & \cdots & x_n^p & (x_n - \kappa_1)_+^p & \cdots & (x_n - \kappa_K)_+^p \end{pmatrix} \quad \text{and}$$

$$D = \text{diag}\{\mathbf{0}_{(p+1)\times 1}, \mathbf{1}_{K\times 1}\}.$$

Then the P-spline estimator of m can be formulated as

$$\hat{m}_\lambda(x) = Z(Z^T Z + n\lambda^{2r} D)^{-1} Z^T Y. \tag{3}$$

Please note that there are many other possibilities to define the penalty matrix D. Our definition of D has the advantage that the low rank components are not penalised. The definition of the smoothing parameter in (2) and (3) is chosen to be $\lambda^* = n\lambda^{2r}$. Other definitions of the smoothing parameter can be found in Wand (1999), Hall and Opsomer (2005), Li and Ruppert (2008) and Claeskens et al. (2009). However, please note that all formulations of λ are equivalent to each other.

2.2 Asymptotic properties

The IPI-algorithm proposed in this paper is based on minimising an asymptotic approximation of the MASE (mean averaged squared error) of $\hat{m}$ obtained by Wand (1999). A well known decomposition of the MASE is the division into its bias and variance. Let W denote the Hat-matrix with

$W_\lambda = Z(Z^T Z + n\lambda^{2r} D)^{-1} Z^T$. The finite sample MASE of $\hat{m}_\lambda$ is then given by

$$\text{MASE}(\hat{m}_\lambda) = \frac{\sigma_\epsilon^2}{n} tr(W_\lambda W_\lambda^T) + \frac{1}{n} ||(W_\lambda - I)m||^2, \tag{4}$$

where the first term on the right side represents the average variance and the second term the average squared bias. The theoretical optimal smoothing parameter given by

$$\lambda_{opt} = \text{argmin MASE}(\hat{m}_\lambda)$$

can be obtained by numerical minimisation of (4). A useful approximation for the MASE is the AMASE (asymptotic mean averaged squared error). Following Wand (1999) the AMASE is given by

$$\begin{aligned} \text{AMASE}(\hat{m}_\lambda) = \sigma_\epsilon^2 \big(&(p+K+1) - n 2\lambda^{2r} tr[(W^TW)^{-1}D] \\ &+ n^2\lambda^{4r} tr\{[(W^TW)^{-1}D]^2\}\big) \\ &+ n^2\lambda^{4r} ||W(W^TW)^{-1}D(W^TW)^{-1}W^Tm||^2. \end{aligned} \tag{5}$$

Differentiating with respect to λ leads to

$$\lambda_A = \Big(\frac{1}{n}\frac{\sigma_\epsilon^2 tr\{(W^TW)^{-1}D\}}{||W(W^TW)^{-1}D(W^TW)^{-1}W^Tm||^2 + \sigma_\epsilon^2 tr[\{(W^TW)^{-1}D\}^2]}\Big)^{\frac{1}{(2r)}}. \tag{6}$$

Please note that (5) and (6) differ from the corresponding equations in Wand (1999) as the penalty term in this paper is written as $n\lambda^{2r}$. This definition has the advantage that the smoothing parameter does not tend to infinity under increasing sample size, i.e. $\lambda_A = O(1)$. Throughout this paper the following expression for estimating the optimal smoothing parameter is mainly considered:

$$\lambda_B = \frac{(\lambda_A + \lambda_C)}{2}, \tag{7}$$

where

$$\lambda_C = \left(\frac{1}{n}\frac{\sigma_\epsilon^2 tr\{(W^TW)^{-1}D\}}{||W(W^TW)^{-1}D(W^TW)^{-1}W^Tm||^2}\right)^{\frac{1}{(2r)}}. \tag{8}$$

Please note that equation (8) is a result of dropping the second term in the denominator of equation (6), as this term is asymptotically negligible. Based on (6), (7) and (8) two data-driven IPI-procedures are proposed.

3 The proposed iterative plug-in algorithms

The unknown parameters in (6) are σ_ϵ and m. With appropriate estimates of these two quantities we can obtain a suitable smoothing parameter by plugging these estimates into (6) or (7). For a given smoothing parameter, $\hat{m}$ can be directly obtained with (3). Note that $\hat{\sigma}_\epsilon$ might possibly depend on λ_0 within the first two to three iterations. Therefore, the variance estimator proposed by Gasser et al. (1986) is used, which is given by

$$\sigma_\epsilon^2 = \frac{2}{3(n-2)} \sum_{i=1}^{n-2} [y_{i+1} - \frac{1}{2}(y_i + y_{i+2})]^2 =: \sigma_G^2, \tag{9}$$

in order to obtain a suitable estimate for σ_ϵ already within the first iteration. Note that this estimator is based on a local linearity assumption for a possible trend. Following the original idea of Gasser et al. (1991) and extending this approach to the P-Spline framework, we propose two IPI-algorithms for independent data. Namely, IPI_A and IPI_B based on equations (6) and (7), respectively. Let λ_0 be the starting smoothing parameter. The IPI-algorithms process as follows:

i) Choose an initial value λ_0. Estimate σ_ϵ^2 with (9) and set $\hat{\sigma}_\epsilon^2 = \sigma_G^2$.

ii) In the j_{th} iteration insert $\hat{\lambda}_{j-1}$ into (3) and obtain $\hat{m}_j$.

iii) Obtain $\hat{\sigma}_{\epsilon,j}$ from the trend-adjusted residuals.

iv) Plug $\hat{m}_j$ and $\hat{\sigma}_{\epsilon,j}^2$ into (6) or (7) and obtain $\hat{\lambda}_j$.

v) Repeat steps ii) to iv), until the J_{th} iteration or convergence is reached and set $\hat{\lambda}_A = \hat{\lambda}_J$ or $\hat{\lambda}_B = \hat{\lambda}_J$.

The number of total iterations J depends on the convergence rate of the IPI. According to our definition, convergence is reached for $j > 1$, if $|\hat{\lambda}_j - \hat{\lambda}_{j-1}| = o(n^{-1})$ or 20 iterations are achieved. IPI_A is usually not affected by the initially chosen smoothing parameter, which could be chosen from a suitable and logical interval, for example $0 \leq \lambda_0 \leq 2$. However IPI_B is not independent from the initially chosen smoothing parameter due to (8) being more sensitive to λ_0. Therefore, we suggest to set the initial smoothing parameter to $\lambda_0 = 0.2$ for $p = 3$. Both algorithms usually converge within a few iterations. Moreover, in comparison with the properties of the IPI-algorithm for local polynomial regression developed by Feng and Beran (2013), our proposal has one major advantage. Namely, that only a pilot estimation of m is needed rather than estimations of the first and second order derivatives m' and m''.

4 A simulation study

In this section a comprehensive simulation study is conducted. Different cases are constructed to investigate the practical performance of IPI_A and IPI_B. Both algorithms are compared with each other in terms of quality of the selected smoothing parameter and the goodness of fit. Moreover, it is investigated how the number of knots impact the finite sample MASE of $\hat{m}_\lambda$.

4.1 Design of the simulation study

To ensure comparability and to demonstrate the applicability of the IPI-algorithms, six trend functions are chosen which have already been used in other simulation studies. The trend functions $f_1 = \tanh(4(x-0.5))$, $f_2 = 2.9(\sin(2(x-0.5))^2)$, $f_3 = \sin(2(x-0.5)\pi)$ and $f_5 = x + 1.5\exp(-100(x-0.5)^2)$ were already subject to a simulation study conducted by Beran et al. (2009), who developed and compared the asymptotic performance of a modified double smoothing bandwidth selector with various similar approaches. The trend functions $f_4 = (\sin(2x\pi))^2 \exp(x)$ and $f_6 = \sin(6x\pi)$ were employed within the scope of an empirical study by Schwarz and Krivobokova (2016). For each trend function data is generated with different sample sizes and variances. The IPI-algorithms for independent data are then applied in 1000 Monte-Carlo simulations for each trend function. The simulation is carried out with sample sizes $n_1 = 250$, $n_2 = 500$ and $n_3 = 1000$. Variances of the error terms are set to $\sigma^2_{\epsilon,1} = 0.01$ and $\sigma^2_{\epsilon,2} = 0.25$. The number of knots is fixed in the main part of the simulation study with $K = 40$ for each trend function. In Section 4.3 it is exemplified that the number of knots only has a negligible effect on the goodness of fit if the number of knots is sufficiently large. In total, 36 cases are tested. Please note that throughout this paper different cases of n and σ^2_ϵ are classified with a case number. For instance, the case with sample size n_1 and $\sigma^2_{\epsilon,1}$ is referred to as Case 11 and the case with sample size n_3 and $\sigma^2_{\epsilon,2}$ is referred to as Case 32. Simulated data for Case 31 and Case 32 are exemplified in Figures 1 and 2.

4.2 Simulation results

In this section the performance of IPI_A and IPI_B is examined and numerical results are presented. The asymptotic (optimal) values are denoted by λ_A and λ_B and are calculated by means of (6) and (7). Moreover, the theoretically optimal smoothing parameter, λ_{opt}, is obtained by numerical minimisation of (4). In Figure 3 the results for Case 31 are illustrated. The theoretical values for λ_A, indicated by the blue line, λ_B, indicated by the red line and λ_{opt}, indicated by the green line, are approximately equal for $f4$, $f5$ and $f6$. For $f1$, $f2$ and $f3$ the deviation from λ_{opt} of λ_A is significantly larger than the deviation of λ_B. The results for Case 32 are shown in Figure 4. Here we observe larger deviations from λ_{opt} of λ_A for $f1$, $f2$ and $f3$ due to the larger variance $\sigma_{\epsilon,2}$. However, $\sigma_{\epsilon,2}$ has almost no impact on λ_B. Across all cases and for all trend functions λ_B nearly coincides with λ_{opt}, except for $f2$ (Figures 3 b and 4 b). For $f2$ the MASE, indicated by the black line, is approximately constant around λ_{opt}, consequently, the relatively small deviation of λ_A is negligible in this case and would not have a strong effect on the estimation quality of m. These findings give us a first indication that approximation (7) could perform better than (6). Graphical analyses for all other cases (Cases 11, 12, 21 and 22) are enclosed in the Appendix

In Figure 5 and Figure 6 boxplots are shown for the estimated smoothing parameters obtained by IPI_A and IPI_B for all trend functions with $\sigma^2_{\epsilon,2}$ and for each sample size. Boxplots for the case with $\sigma^2_{\epsilon,1}$ are to be found in the Appendix. The poor behaviour of the IPI_A estimator for the first three trend functions $f1$, $f2$ and $f3$ (see Figure 5, a, b and c) becomes very obvious. The median slightly increases with increasing sample size and the variance does not decrease with increasing sample size. We observe a lot of outliers for each sample size below the bottom whisker and the distribution of $\hat{\lambda}_A$ is left-skewed for all three trend functions. On the contrary, the performance of IPI_B-estimator for the first three trend functions (see Figure 6, a, b and c) is better. We observe a decreasing variance with increasing sample size, less outliers are observed and the distribution of $\hat{\lambda}_B$ is not skewed. For trend functions $f4$, $f5$ and $f6$ (see Figures 5 and 6, d, e and f) one can clearly recognize that the values of both IPI-estimators in the 25% and 75% quartiles are distributed closer around the median with increasing sample size, i.e. the variance of $\hat{\lambda}_A$ and $\hat{\lambda}_B$ decreases. The boxplots for $f4$ and $f6$ (see Figures 5 and 6, d and f) are very similar for IPI_A and IPI_B. For $f5$ (see Figure 6, f) the deviation from the median is quite large for $n = 250$, which is due to (8) being too unstable for this relatively small sample size. The deviation diminishes with increasing sample size.

Numerical results for $\sigma^2_{\epsilon,1}$ and $\sigma^2_{\epsilon,2}$ are presented in Table 1 and 2. The arithmetic means of the MASE values of $\hat{m}_{\lambda_A}$ and $\hat{m}_{\lambda_B}$ multiplied with 10000 are denoted by MASE_A and MASE_B, respectively. MASE_{opt} stands for the theoretical optimal MASE. The average of the MSE values of $\hat{\lambda}_A$ and $\hat{\lambda}_B$ multiplied with 1000 are denoted by MSE_A and MSE_B, the means of the estimated smoothing parameters are denoted by $\bar{\hat{\lambda}}_A$ and $\bar{\hat{\lambda}}_B$, as well as the mean of the variance of the error term multiplied with 100, denoted by $\bar{\hat{\sigma}}_{\epsilon A}$ and $\bar{\hat{\sigma}}_{\epsilon B}$. The results confirm our expectation that IPI_B performs better than IPI_A. For the first three trend functions IPI_A seems to perform much worse than IPI_B, in particular for $\sigma^2_{\epsilon,2} = 0.25$ (see Table 2). In this case the values of MASE_A and especially MSE_A are much higher than the corresponding values obtained with IPI_B. Overall, IPI_B performs better or at least equally as good in all cases. The MASE and MSE values of both estimators strongly decrease with increasing sample size, which implies that both IPI-estimators are consistent. The means of the estimated variances $\bar{\hat{\sigma}}_{\epsilon A}$ as well as $\bar{\hat{\sigma}}_{\epsilon B}$ are already very close to the true variances for the smallest sample size and converge to $\sigma^2_{\epsilon,1} = 0.01$ and $\sigma^2_{\epsilon,2} = 0.25$ with increasing sample size.

4.3 Knot selection

Selecting an appropriate number of knots can be done manually by visually analysing the complexity of the data. The number of knots are neither to be too small, as there would not be enough observations between two knots, nor too big, in order to save computing time. Ruppert (2002) developed a myopic and a fullsearch algorithm for automatically determining the optimal amount of knots. Based on his findings the author suggests to use a default number of knots for large data sets from 20 to 40 knots (see Ruppert et al. (2003)). In this part of the simulation study we illustrate that the amount of knots does not have a strong effect on the efficiency of the estimation results, at least for the trend functions which are examined in this paper. More specifically, we investigate how the MASE is impacted by various selected numbers of knots. This part of the simulation is carried out with a steadily increasing sequence of numbers of knots $K_i \in \{10, 12, ..., 20, 30, ..., 100, 150, 200, 250, n\}$. For each K_i a sequence of smoothing parameters in the interval $\lambda_j \in \{0.0030, 0.0031, 0.0032, ..., 0.4998, 0.4999, 0.5\}$ is defined. Then for each K_i the theoretically optimal smoothing parameter with the lowest MASE, denoted by $\lambda_{\text{opt}}^{K_i}$ and $\text{MASE}_{\text{opt}}^{K_i}$, respectively, is determined by plugging λ_j into (4). Let the ratio of $\lambda_{\text{opt}}^{K_i}$ to $\bar{\lambda}_{\text{opt}} = \frac{1}{n}\sum_{i=1}^{n} \lambda_{\text{opt}}^{K_i}$ as well as $\text{MASE}_{\text{opt}}^{K_i}$ to $\overline{\text{MASE}}_{\text{opt}} = \frac{1}{n}\sum_{i=1}^{n} \text{MASE}_{\text{opt}}^{K_i}$ be defined by

$$\varphi_{\lambda,i} = \frac{\lambda_{\text{opt}}^{K_i}}{\bar{\lambda}_{opt}} \quad \text{and} \quad \varphi_{\text{MASE},i} = \frac{\text{MASE}_{\text{opt}}^{K_i}}{\overline{\text{MASE}}_{\text{opt}}}.$$

In Figure 17 the ratios $\varphi_{\lambda,i}$ and $\varphi_{\text{MASE},i}$, indicated by the solid red line and solid blue line, for Case 12 are shown. Graphical analyses for all other cases are enclosed in the Appendix. Overall, our results coincide with the findings of Ruppert (2002), who proposed to use a minimum number of knots of $K = \min\{N/4, 35\}$. Our findings confirm that if the number of knots is sufficiently large, then the effect on the MASE is negligible and it remains constant. Within the scope of our simulation study it is found that this is already the case for $K > 10$.

5 Application

In this section the IPI-algorithm is applied to real data examples for uncorrelated data, in order to present that the IPI-algorithm can be applied to a wide range of data sets from different fields of science.

The first data set used in this section is the LIDAR* data set with 221 observations. The independent variable is *range* which stands for the distance a laser light travels when illuminating a target. The dependent variable is *logratio*, which stands for the logarithm of the ratio of light received from two laser sources. The above mentioned regression methods are applied to this data set and compared with each other in Figure 8 (a), where the fitted P-spline obtained by applying IPI_A is indicated by the black solid line and the fitted P-spline obtained by IPI_B is indicated by the red dashed line. The performance of both estimators is quite well. The finally selected smoothing parameters for IPI_A and IPI_B are approximately equal with $\hat{\lambda}_A \approx \hat{\lambda}_B \approx 0.098$. A further descriptive analysis is beyond the scope of this paper.

*The data set is implemented in the R-package *SemiPar*.

As the second data set the California[†] test score data set is chosen, which contains data on the test performance, characteristics of the school and the demographic backgrounds of students from 420 districts in California from 1998 to 1999. For the purpose of this paper *test score* is chosen to be the dependent variable and *income* the independent variable. *Test score* is defined as the average of reading and math scores in standardized tests designed for 5_{th} grade students. *Income* is defined as the average income per capita in a district. Figure 8 (b) illustrates the fitted P-spline, indicated by the black solid line, the fitted local polynomial, indicated by the purple dashed line, cubic regression by the light blue dashed line and simple linear regression by the red dashed line. Again, the P-spline smoother delivers very satisfying results. By analysing both fitted splines, first a strong upward trend is observed, which indicates a strong positive correlation between *income* and *test score*. This upward trend steadily decreases and turns into a slightly negative trend. These findings confirm our expectations that the relation between *income* and *test score* is nonlinear with a strong positive correlation at the beginning. The finally selected smoothing parameter for IPI_A and IPI_B are $\hat{\lambda}_A = 0.105$ and $\hat{\lambda}_B = 0.132$, respectively.

The third data set is the German Socio-Economic Panel (SOEP), seeWagner et al., 2007. We use the wave of 2006 which contains measures of physical fitness, mental fitness and body mass index, each on a scale from 1 to 100. The population polled consists of more than $21,000$ citizens characterised by their age, their place of residence and their gender. *Physical fitness* is defined as the dependent variable and *age* as the independent variable. Furthermore, the observed individuals are separated into categories of gender and place of residence (West- or East-Germany). The individuals' age is limited, the minimum age is 18 and the maximum age is 65, and the maximum of working hours per week are 50 hours. Figure 8 shows the fitted P-splines for both algorithms, indicated by the black and red solid line. It can be seen that the relation between *age* and *physical fitness* is apparently non-linear. From the age of 18 to 25 years only a slightly negative or no correlation can be observed. Between the age of 25 to 35 years the negative correlation starts to increase steadily. This trend continues until the age of 65 in all groups in West and East Germany. For West Germany the P-spline fit shows a slightly different curvature. Unlike in East Germany the physical fitness of the male and female group members (c) and (d) seem to stabilize around the age of 55. Neither a significant negative nor positive correlation can be observed between the ages of 60 and 65. Interestingly enough, there is a difference

[†]The data set is implemented in the R-package *AER*.

between the male and female group in West Germany concerning the downward tendency. Apparently, physical fitness seems to decrease more rapidly in the female group than in the male group. Overall algorithm IPI_B is to be preferred for this data example as IPI_A produces a too responsive fit.

6 Final remarks

This paper proposes new IPI-rules for selecting the smoothing parameter in the P-spline framework. A comprehensive simulation study is conducted. Based on our results we recommend to use the IPI_B algorithm in most cases. The application to real data examples illustrates the wide applicability of our proposal and that it works very well in practice. A further improvement of our proposal could be the innovation of the estimation of σ_ϵ^2. In our paper we estimated σ_ϵ^2 in the first iteration with a difference based variance estimator. A possible improvement could be the repeated estimation of σ_ϵ^2 in each iteration based on the residuals. A combination of both methods might also perform well. Moreover, the development of a P-spline IPI-algorithm for time series data could offer an interesting possibility for future research. Possible extensions of our proposal could be the short-memory and long-memory case.

References

Aerts, M., G. Claeskens, and M. P. Wand (2002). “Some theory for penalized spline generalized additive models”. In: *Journal of statistical planning and inference* 103.1, pp. 455–470.

Akaike, H. (1974). “A new look at the statistical model identification”. In: *IEEE transactions on automatic control* 19.6, pp. 716–723.

Beran, J., Y. Feng, and S. Heiler (2009). “Modifying the double smoothing bandwidth selector in nonparametric regression”. In: *Statistical Methodology* 6.5, pp. 447–465.

Claeskens, G., T. Krivobokova, and J. D. Opsomer (2009). “Asymptotic properties of penalized spline estimators”. In: *Biometrika* 96.3, pp. 529–544.

Cleveland, W. S. (1979). “Robust locally weighted regression and smoothing scatterplots”. In: *Journal of the American statistical association* 74.368, pp. 829–836.

Craven, P. and G. Wahba (1978). “Smoothing noisy data with spline functions”. In: *Numerische Mathematik* 31.4, pp. 377–403.

Demmler, A and C Reinsch (1975). “Oscillation matrices with spline smoothing”. In: *Numerische Mathematik* 24.5, pp. 375–382.

Eilers, P. H. and B. D. Marx (1996). “Flexible smoothing with B-splines and penalties”. In: *Statistical science*, pp. 89–102.

Eilers, P. H., B. D. Marx, and M. Durbán (2015). “Twenty years of P-splines”. In: *SORT-Statistics and Operations Research Transactions* 39.2, pp. 149–186.

Feng, Y. and J. Beran (2013). “Optimal convergence rates in non-parametric regression with fractional time series errors”. In: *Journal of Time Series Analysis* 34.1, pp. 30–39.

Gasser, T., A. Kneip, and W. Köhler (1991). “A flexible and fast method for automatic smoothing”. In: *Journal of the american statistical association* 86.415, pp. 643–652.

Gasser, T., L. Sroka, and C. Jennen-Steinmetz (1986). “Residual variance and residual pattern in nonlinear regression”. In: *Biometrika*, pp. 625–633.

Hall, P. and J. D. Opsomer (2005). “Theory for penalised spline regression”. In: *Biometrika* 92.1, pp. 105–118.

Kauermann, G. (2005). “A note on smoothing parameter selection for penalized spline smoothing”. In: *Journal of statistical planning and inference* 127.1, pp. 53–69.

Krivobokova, T. (2013). “Smoothing parameter selection in two frameworks for penalized splines”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.4, pp. 725–741.

Li, Y. and D. Ruppert (2008). “On the asymptotics of penalized splines”. In: *Biometrika* 95.2, pp. 415–436.

Mallows, C. L. (1973). “Some comments on C p”. In: *Technometrics* 15.4, pp. 661–675.

Mosier, C. I. (1951). “The need and means of cross validation. I. Problems and designs of cross-validation.” In: *Educational and Psychological Measurement*.

Nadaraya, E. A. (1964). “On estimating regression”. In: *Theory of Probability & Its Applications* 9.1, pp. 141–142.

O’sullivan, F., B. S. Yandell, and W. J. Raynor Jr (1986). “Automatic smoothing of regression functions in generalized linear models”. In: *Journal of the American Statistical Association* 81.393, pp. 96–103.

Parker, R. and J. Rice (1985). “Discussion of ”Some aspects of the spline smoothing approach to nonparametric curve fitting” by BW Silverman”. In: *Journal of the Royal Statistical Society, Series B* 47, pp. 40–42.

Ruppert, D. (2002). “Selecting the number of knots for penalized splines”. In: *Journal of computational and graphical statistics*.

Ruppert, D., S. J. Sheather, and M. P. Wand (1995). “An effective bandwidth selector for local least squares regression”. In: *Journal of the American Statistical Association* 90.432, pp. 1257–1270.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. 12. Cambridge university press.

Schwarz, K. and T. Krivobokova (2016). “A unified framework for spline estimators”. In: *Biometrika* 103.1, pp. 121–131.

Wager, C., F. Vaida, and G. Kauermann (2007). “Model selection for penalized spline smoothing using Akaike information criteria”. In: *Australian & New Zealand Journal of Statistics* 49.2, pp. 173–190.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). “The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements”. In.

Wahba, G. (1977). “Optimal smoothing of density estimates”. In: *Classification and Clustering* 1, pp. 423–458.

Wand, M. P. (1999). “On the optimal amount of smoothing in penalised spline regression”. In: *Biometrika* 86.4, pp. 936–940.

Wang, X., J. Shen, D. Ruppert, et al. (2011). “On the asymptotics of penalized spline smoothing”. In: *Electronic Journal of Statistics* 5, pp. 1–17.

Watson, G. S. (1964). “Smooth regression analysis”. In: *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 359–372.

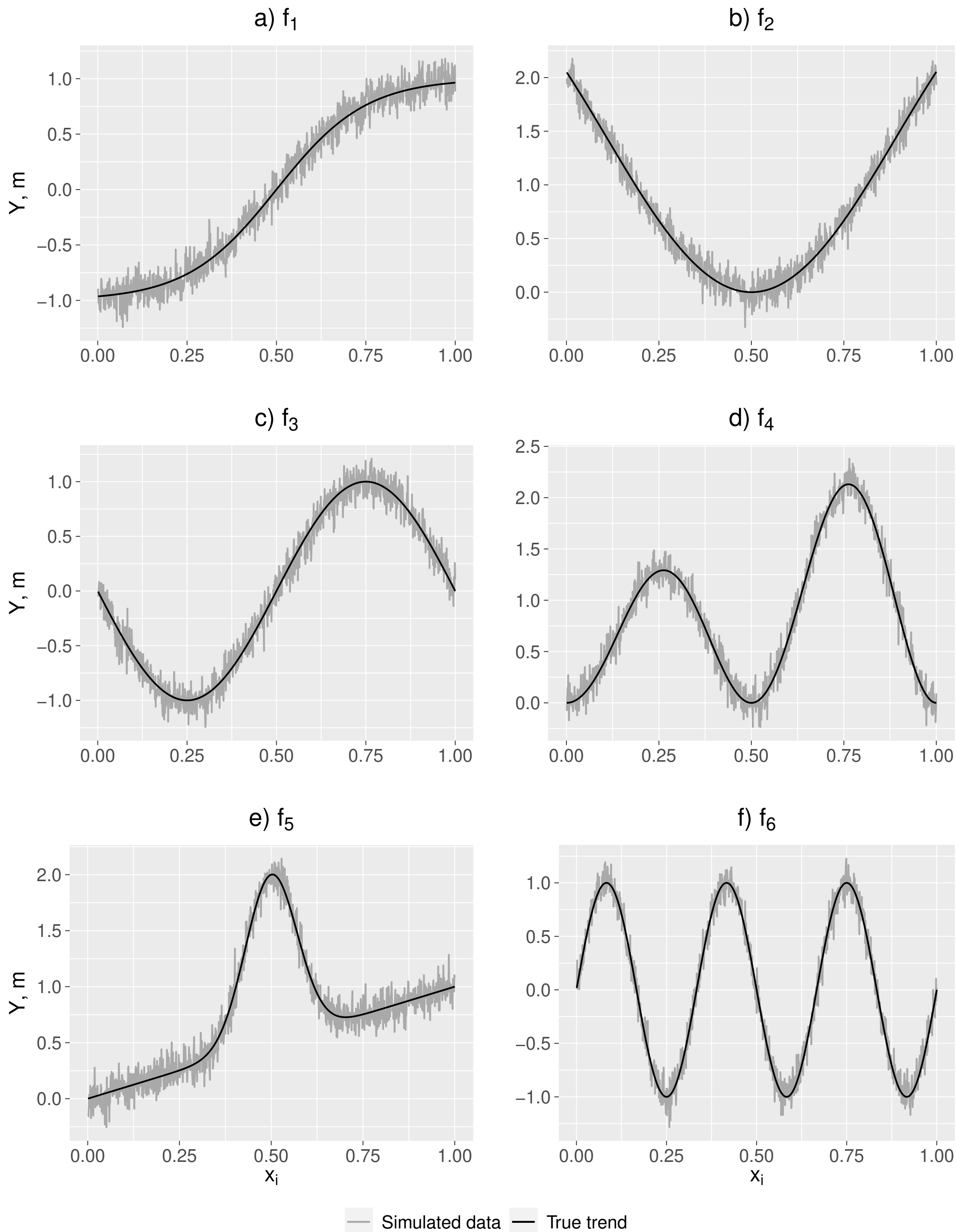

Figure 1: Case 31 - Simulated data and true trend functions.

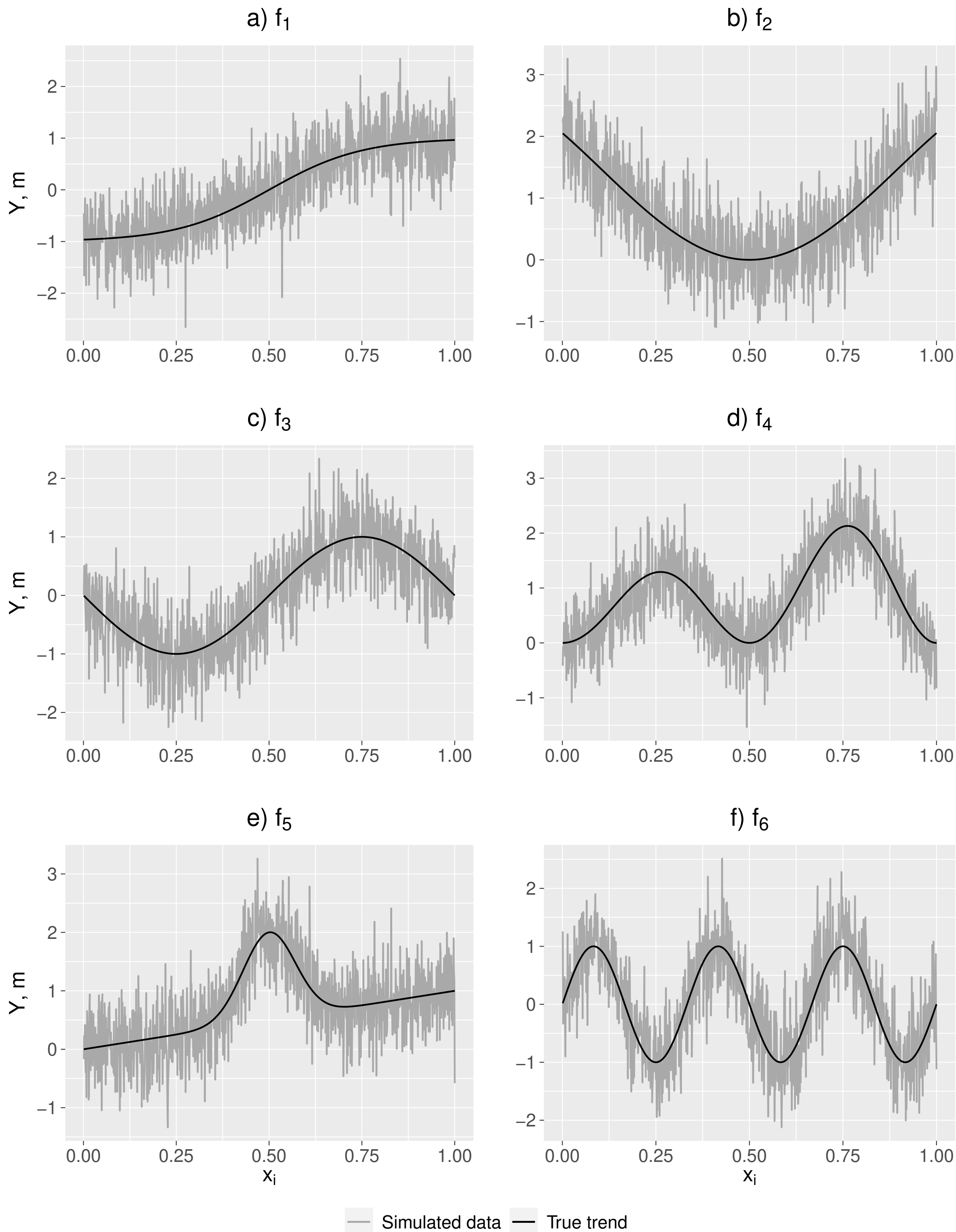

Figure 2: Case 32 - Simulated data and true trend functions.

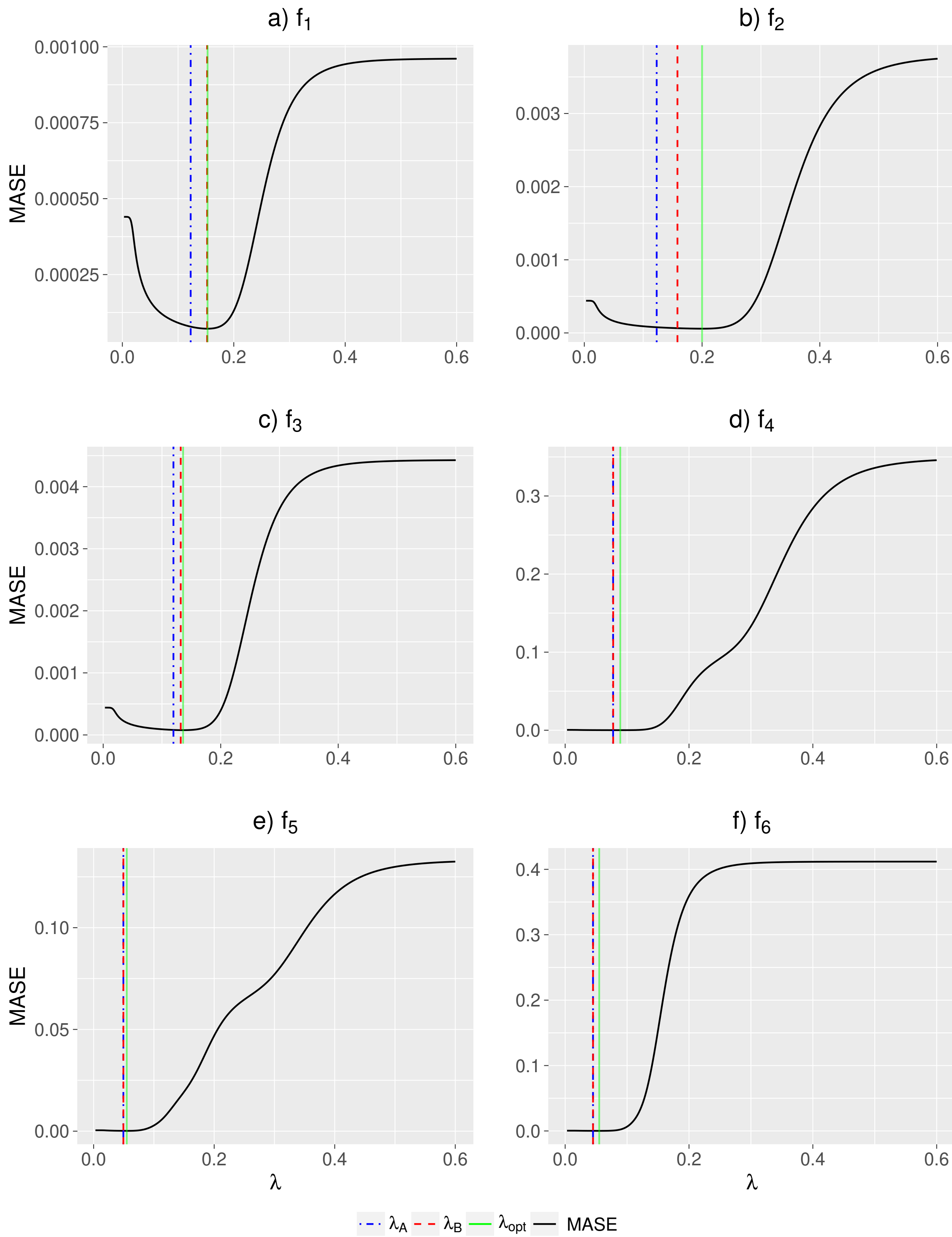

Figure 3: Case 31 - MASE. λ_A, λ_B and λ_{opt} are indicated by the blue-dashed, red and green lines, respectively.

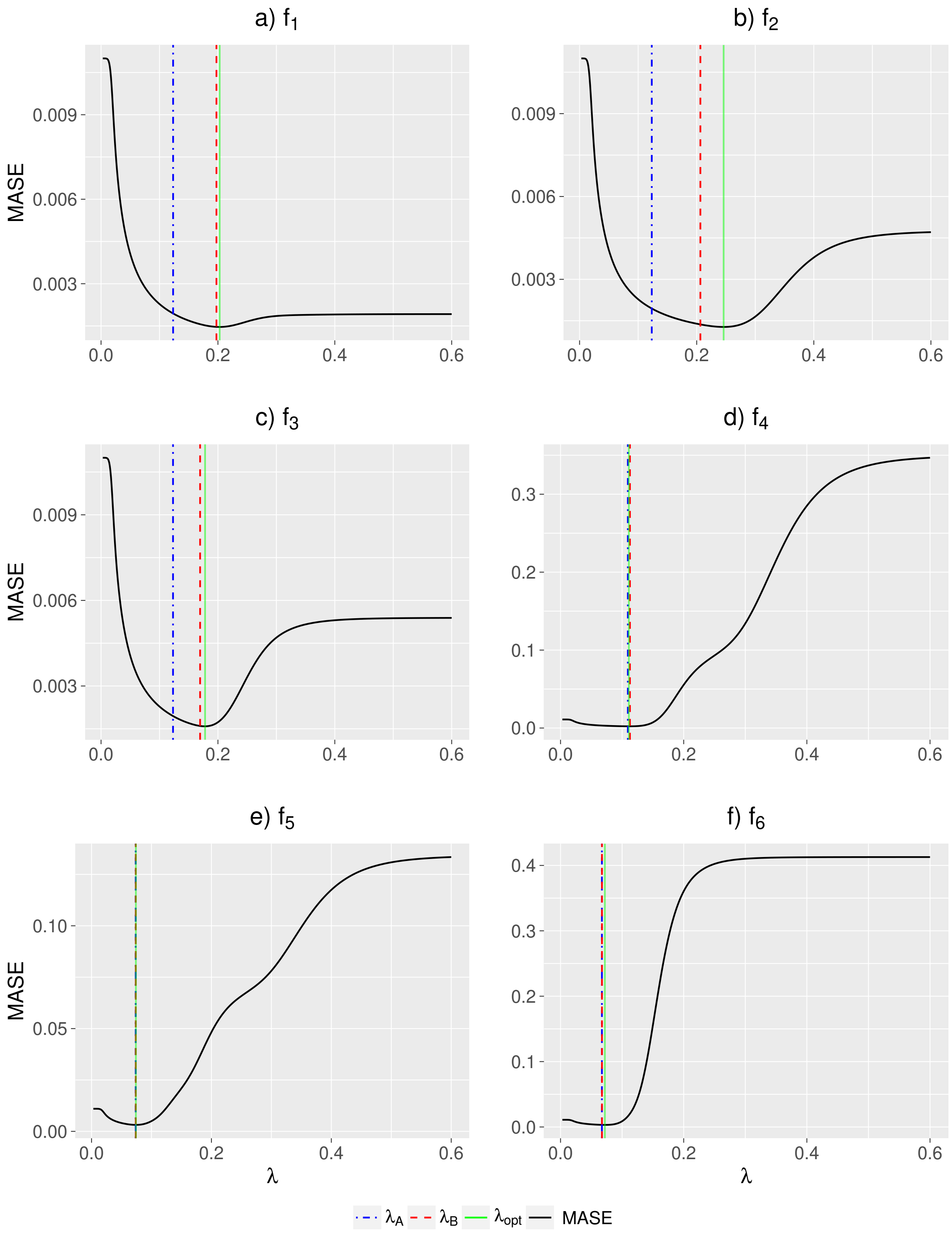

Figure 4: Case 32 - MASE. λ_A, λ_B and λ_{opt} are indicated by the blue-dashed, red and green lines, respectively.

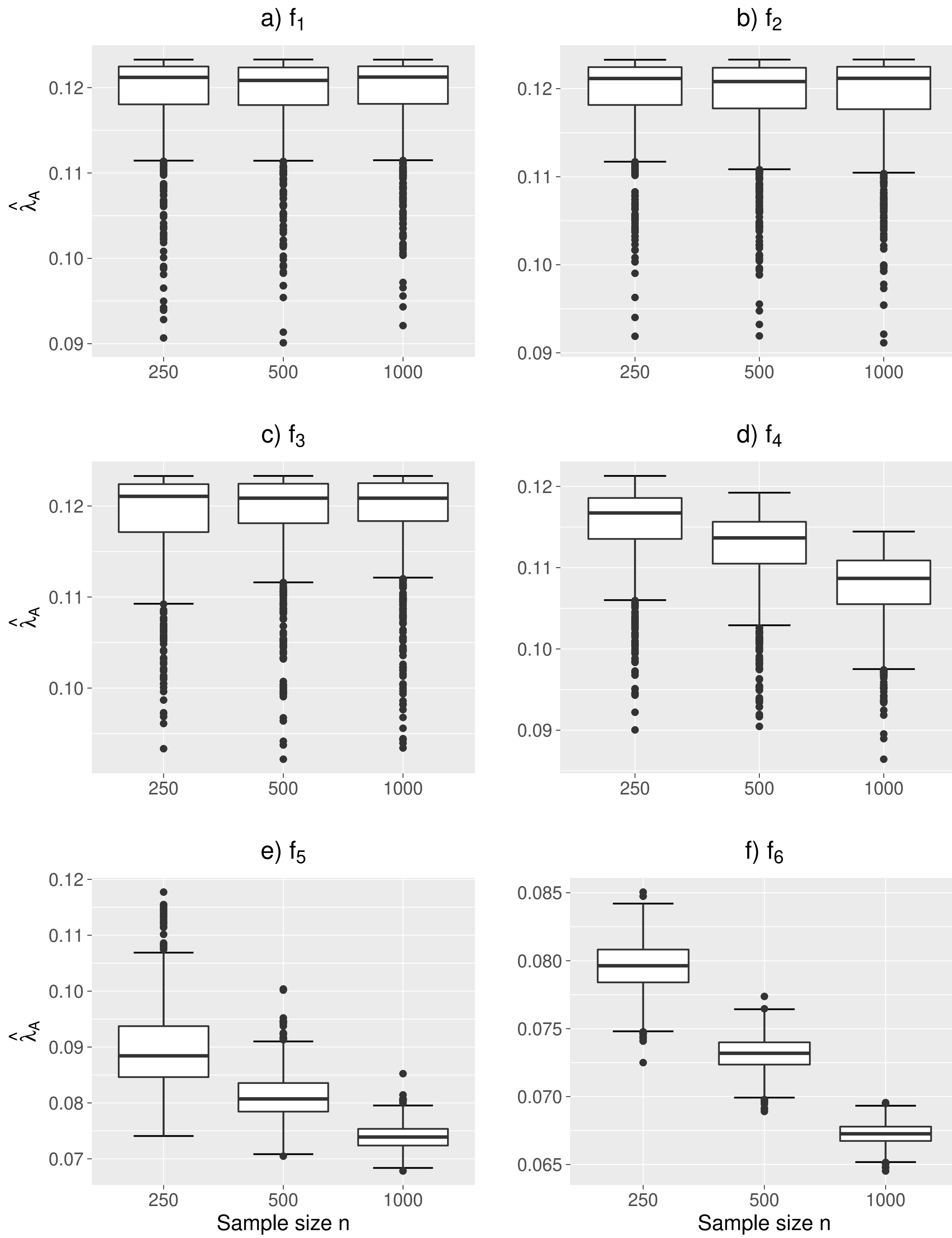

Figure 5: Boxplots for IPI_A with $\sigma^2_{\epsilon,2}$.

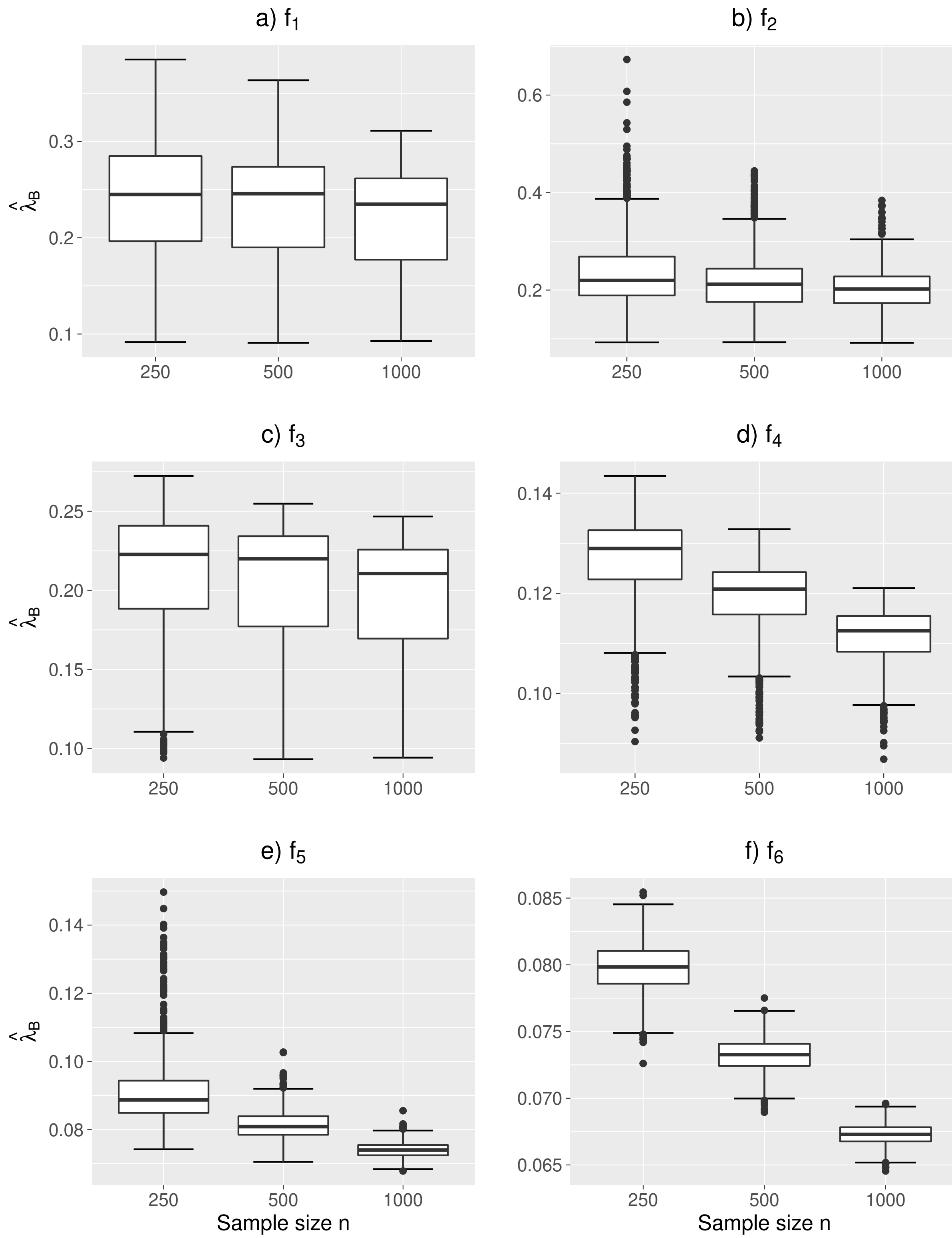

Figure 6: Boxplots for IPI_B with $\sigma^2_{\epsilon,2}$.

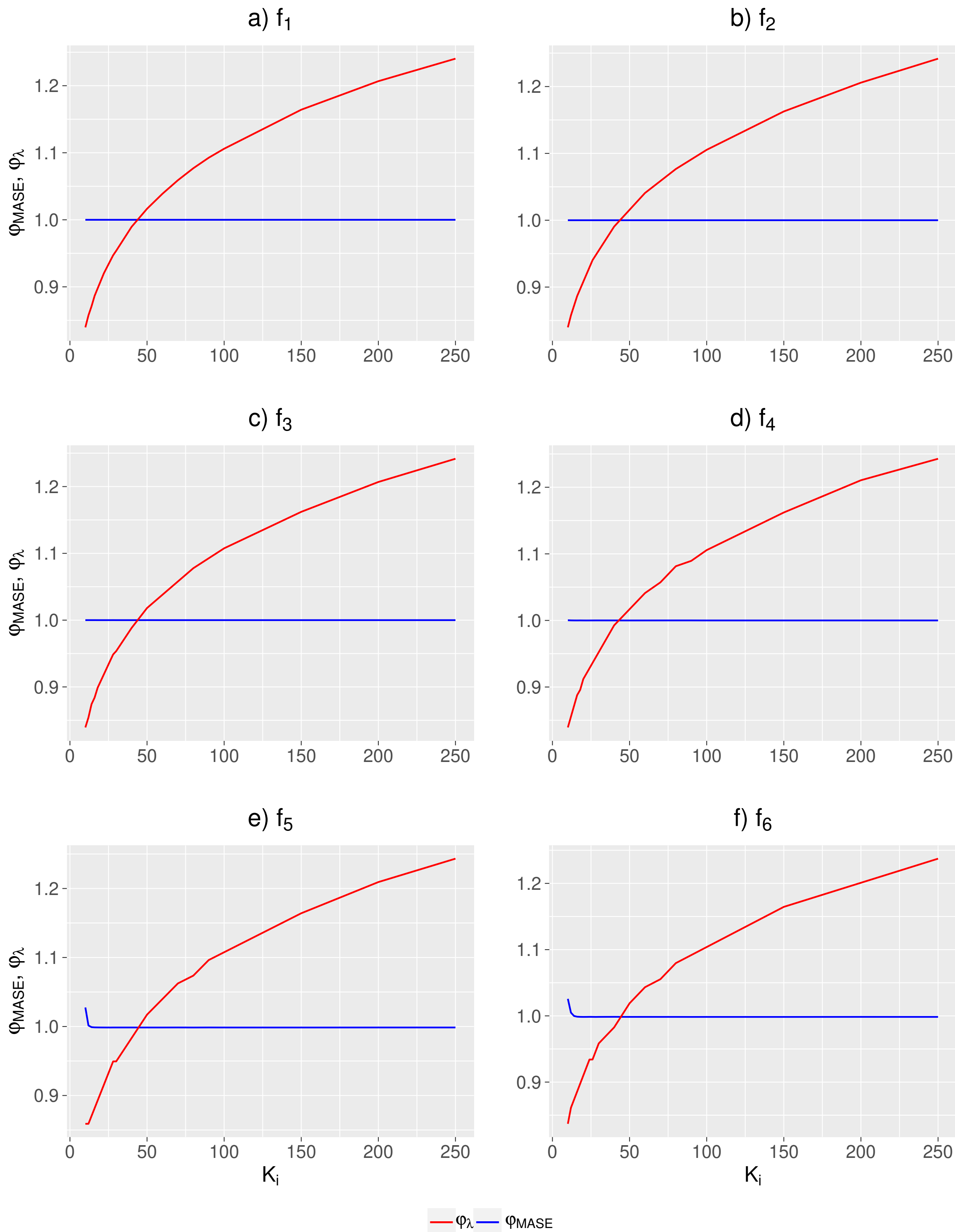

Figure 7: Case 12 - φ_λ and φ_{MASE} are indicated by the red and blue line, respectively.

Table 1: Numerical results for all trend functions and sample sizes with $\sigma^2_{\epsilon,1}$.

n	f	MASE$_{opt}$	MASE$_A$	MASE$_B$	MSE$_A$	MSE$_B$	$\bar{\hat{\lambda}}_A$	$\bar{\hat{\lambda}}_B$	$\bar{\hat{\sigma}}_{\epsilon,A}$	$\bar{\hat{\sigma}}_{\epsilon,B}$
250	f_1	2.609	3.280	3.676	2.632	1.714	0.119	0.188	0.966	0.984
	f_2	2.193	3.176	2.521	9.233	2.562	0.119	0.172	0.966	0.974
	f_3	2.827	3.258	3.497	1.135	0.630	0.119	0.158	0.960	0.972
	f_4	3.933	4.081	4.073	0.049	0.044	0.091	0.091	0.959	0.958
	f_5	5.785	5.937	5.935	0.009	0.009	0.058	0.058	0.934	0.936
	f_6	5.863	6.220	6.220	0.063	0.063	0.053	0.053	0.931	0.931
500	f_1	1.366	1.670	1.862	1.790	1.047	0.119	0.174	0.987	0.994
	f_2	1.129	1.623	1.344	7.627	2.573	0.119	0.161	0.982	0.985
	f_3	1.479	1.664	1.666	0.689	0.378	0.117	0.143	0.980	0.983
	f_4	2.059	2.132	2.128	0.071	0.067	0.084	0.084	0.977	0.977
	f_5	3.033	3.150	3.150	0.020	0.020	0.054	0.054	0.962	0.963
	f_6	3.087	3.363	3.363	0.066	0.066	0.049	0.049	0.960	0.960
1000	f_1	0.718	0.825	0.856	0.995	0.589	0.119	0.160	0.991	0.994
	f_2	0.582	0.808	0.693	6.161	2.365	0.119	0.153	0.991	0.993
	f_3	0.774	0.840	0.836	0.341	0.212	0.116	0.134	0.992	0.994
	f_4	1.082	1.146	1.145	0.099	0.096	0.077	0.077	0.990	0.990
	f_5	1.586	1.678	1.678	0.023	0.023	0.049	0.049	0.979	0.979
	f_6	1.626	1.824	1.824	0.082	0.082	0.045	0.045	0.978	0.978

Table 2: Numerical results for all trend functions and sample sizes with $\sigma^2_{\epsilon,2}$.

n	f	MASE$_{opt}$	MASE$_A$	MASE$_B$	MSE$_A$	MSE$_B$	$\bar{\hat{\lambda}}_A$	$\bar{\hat{\lambda}}_B$	$\bar{\hat{\sigma}}_{\epsilon,A}$	$\bar{\hat{\sigma}}_{\epsilon,B}$
250	f_1	51.237	81.147	61.753	32.623	7.679	0.119	0.238	24.154	24.477
	f_2	48.248	80.287	60.901	22.979	8.294	0.119	0.237	24.151	24.448
	f_3	58.998	80.549	70.301	5.958	1.947	0.119	0.209	24.058	24.409
	f_4	82.299	84.800	87.793	0.058	0.110	0.115	0.127	24.152	24.296
	f_5	111.930	116.784	118.922	0.068	0.121	0.090	0.091	24.060	24.067
	f_6	115.731	119.089	119.076	0.004	0.004	0.080	0.080	23.812	23.843
500	f_1	28.060	39.689	31.948	9.612	3.618	0.119	0.231	24.544	24.719
	f_2	24.834	41.691	32.828	18.537	5.704	0.119	0.215	24.582	24.719
	f_3	30.604	40.455	38.974	4.224	2.003	0.119	0.203	24.483	24.681
	f_4	42.695	44.607	45.777	0.031	0.068	0.112	0.119	24.511	24.556
	f_5	59.569	59.932	60.064	0.022	0.025	0.081	0.081	24.469	24.470
	f_6	60.872	61.051	61.032	0.005	0.005	0.073	0.073	24.367	24.371
1000	f_1	14.648	20.194	18.053	6.395	3.400	0.119	0.218	24.744	24.834
	f_2	12.753	20.136	15.562	15.100	3.907	0.119	0.199	24.779	24.843
	f_3	15.871	20.192	21.994	3.142	1.854	0.119	0.195	24.810	24.919
	f_4	22.160	22.899	23.056	0.025	0.032	0.108	0.111	24.814	24.826
	f_5	31.561	32.381	32.388	0.006	0.006	0.074	0.074	24.678	24.680
	f_6	32.011	32.896	32.889	0.015	0.014	0.067	0.067	24.593	24.594

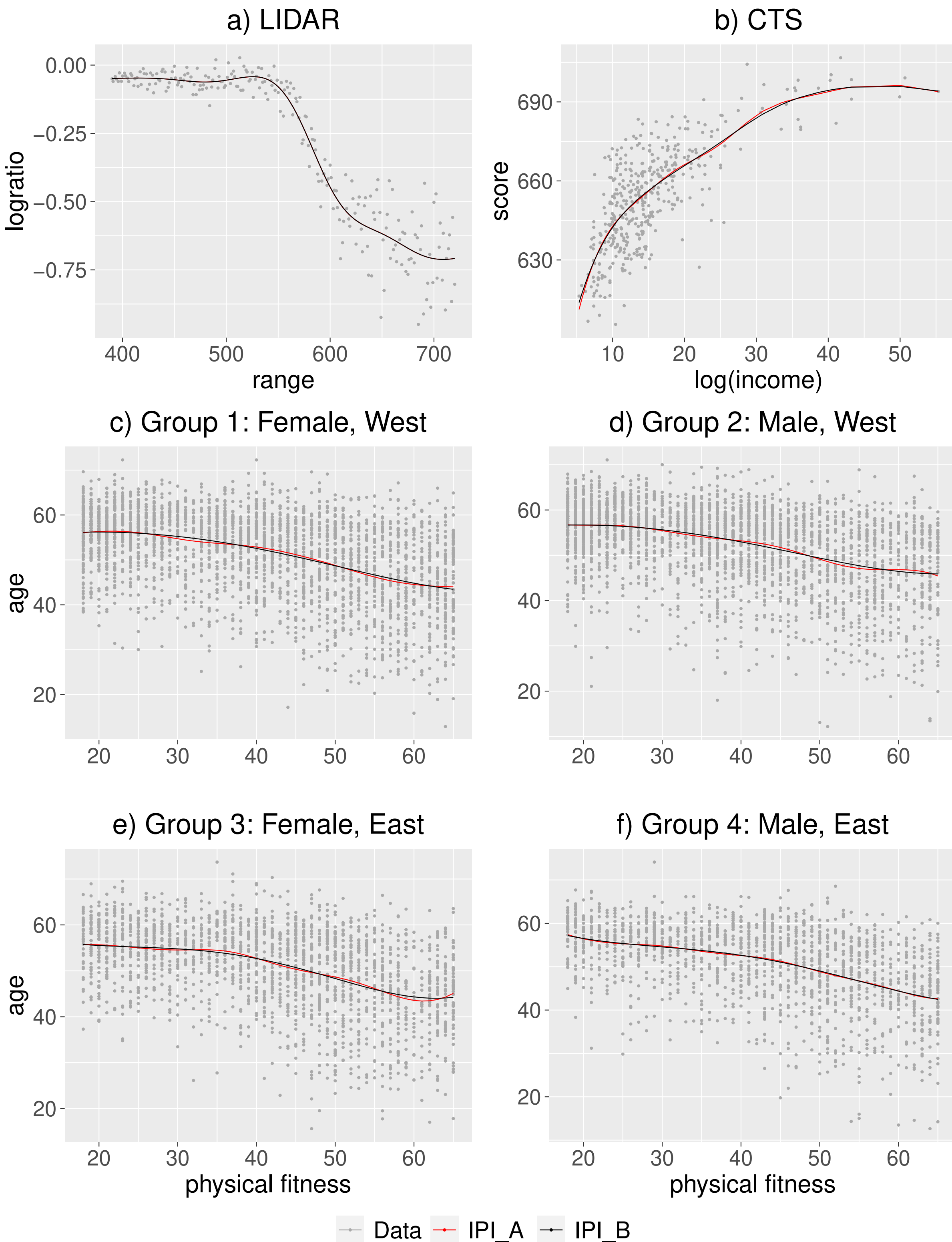

Figure 8: LIDAR data set and California test score data set (a and b). SOEP data of 2006 (c to f). The estimated trends obtained by means of IPI_A and IPI_B are indicated by the red and black line, respectively.

Appendix

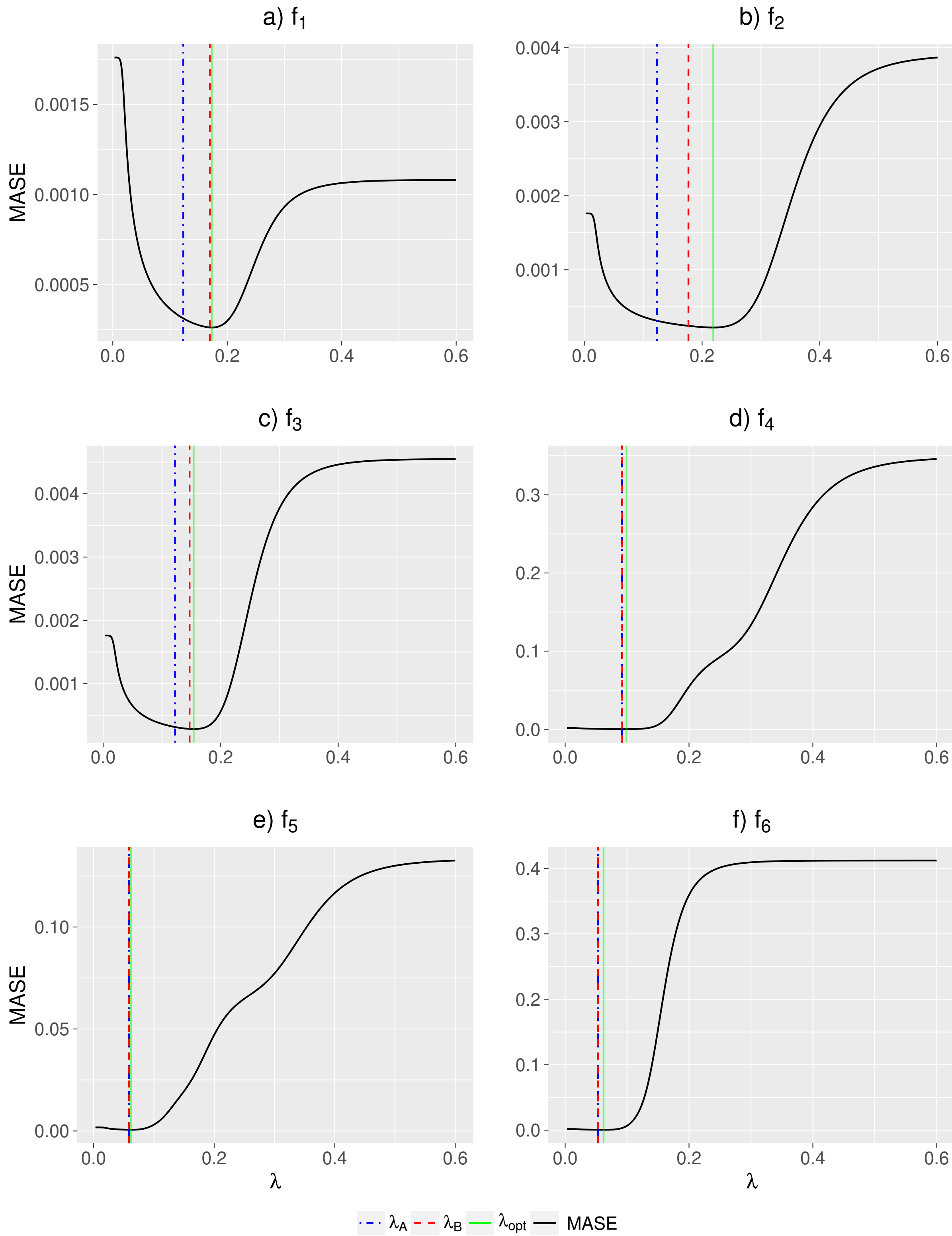

Figure 9: Case 11 - MASE. λ_A, λ_B and λ_{opt} are indicated by the blue-dashed, red and green lines, respectively.

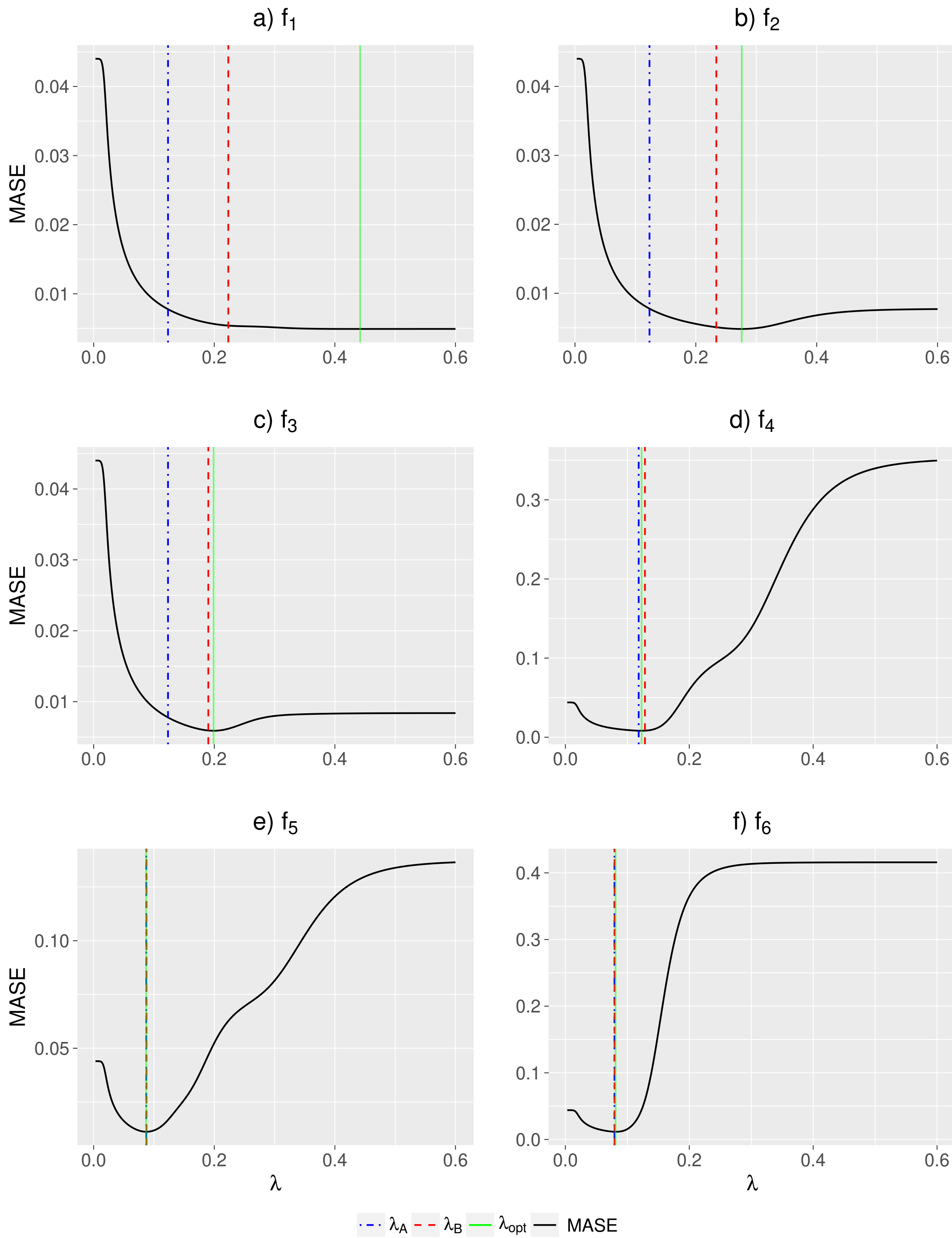

Figure 10: Case 12 - MASE. λ_A, λ_B and λ_{opt} are indicated by the blue-dashed, red and green lines, respectively.

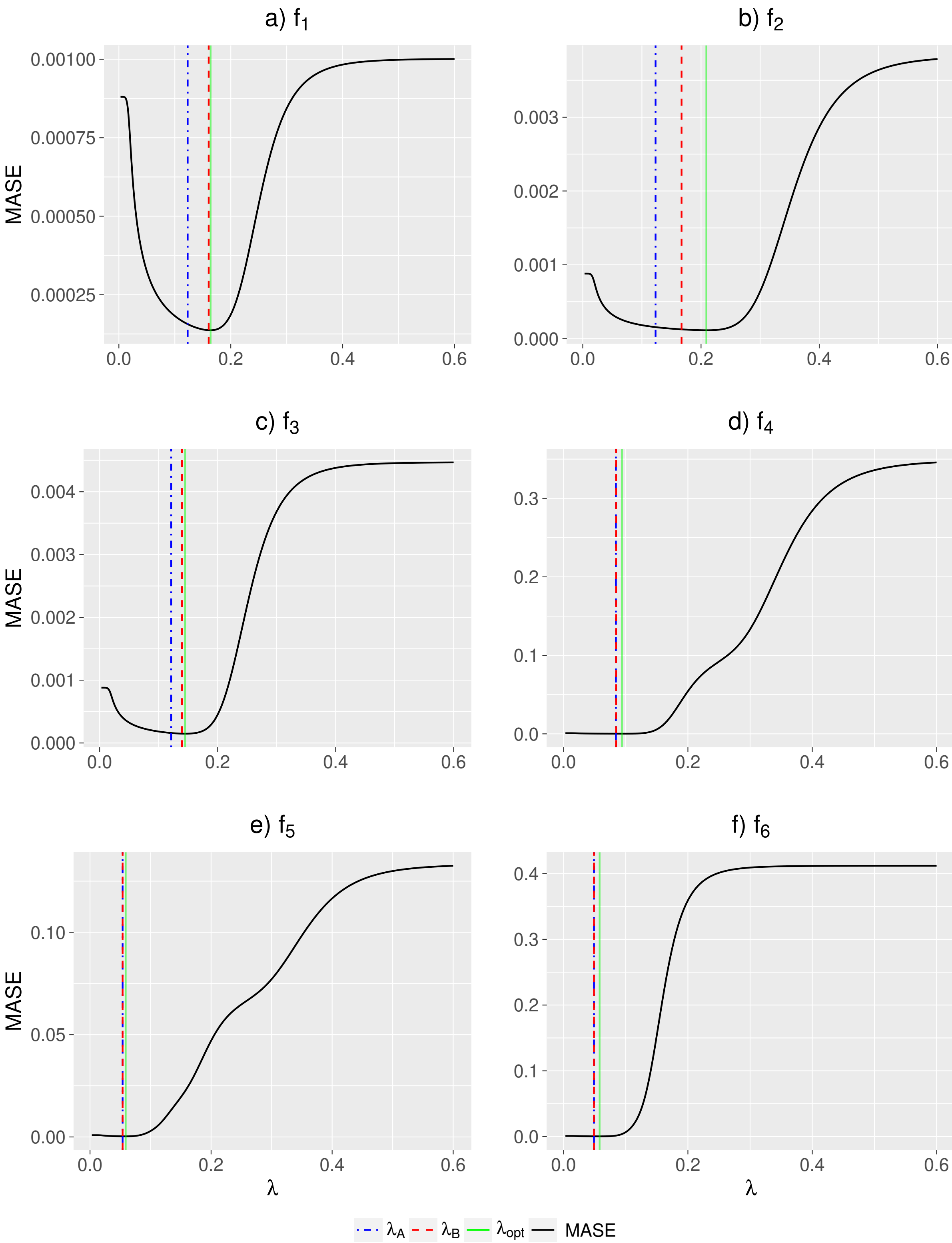

Figure 11: Case 21 - MASE. λ_A, λ_B and λ_{opt} are indicated by the blue-dashed, red and green lines, respectively.

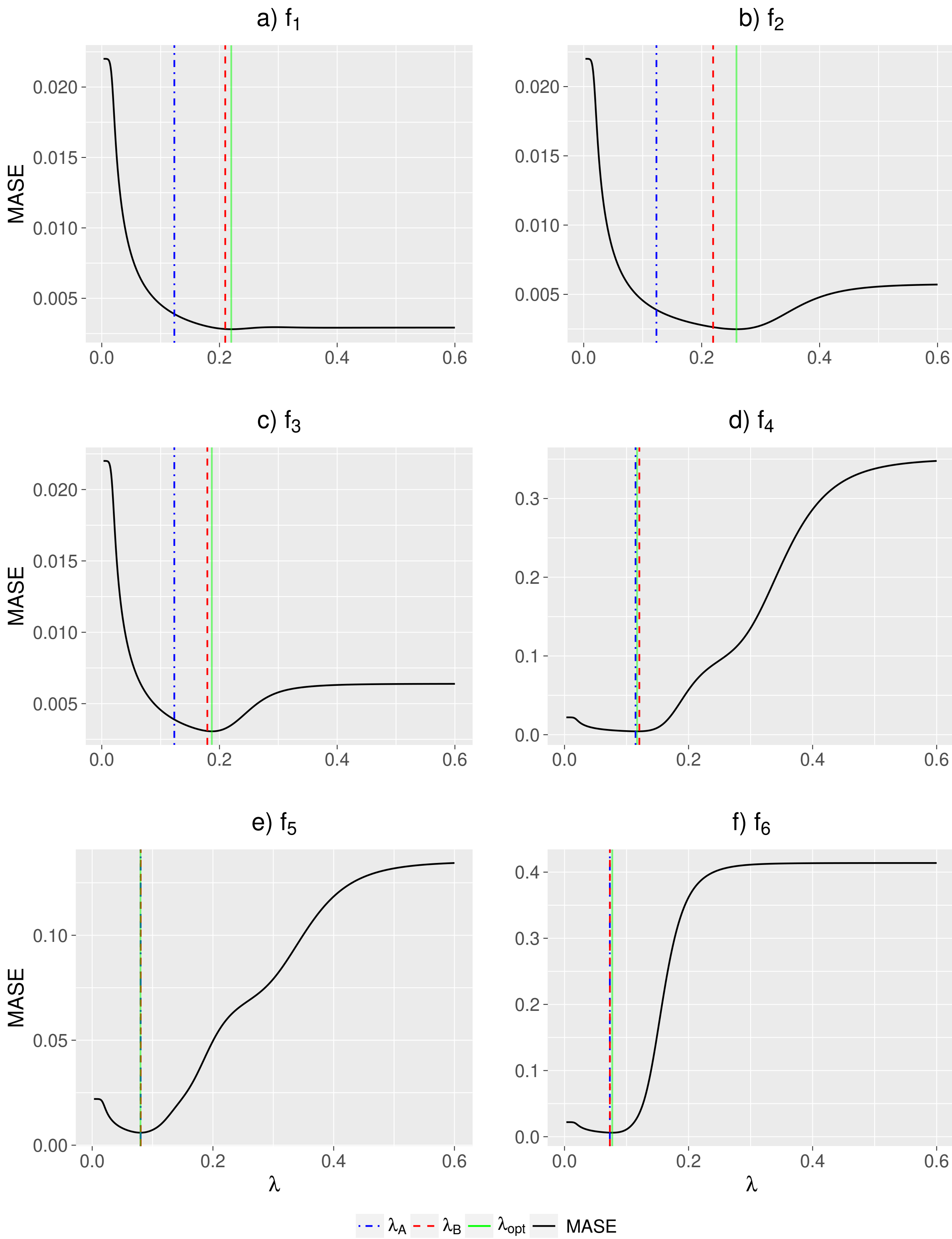

Figure 12: Case 22 - MASE. λ_A, λ_B and λ_{opt} are indicated by the blue-dashed, red and green lines, respectively.

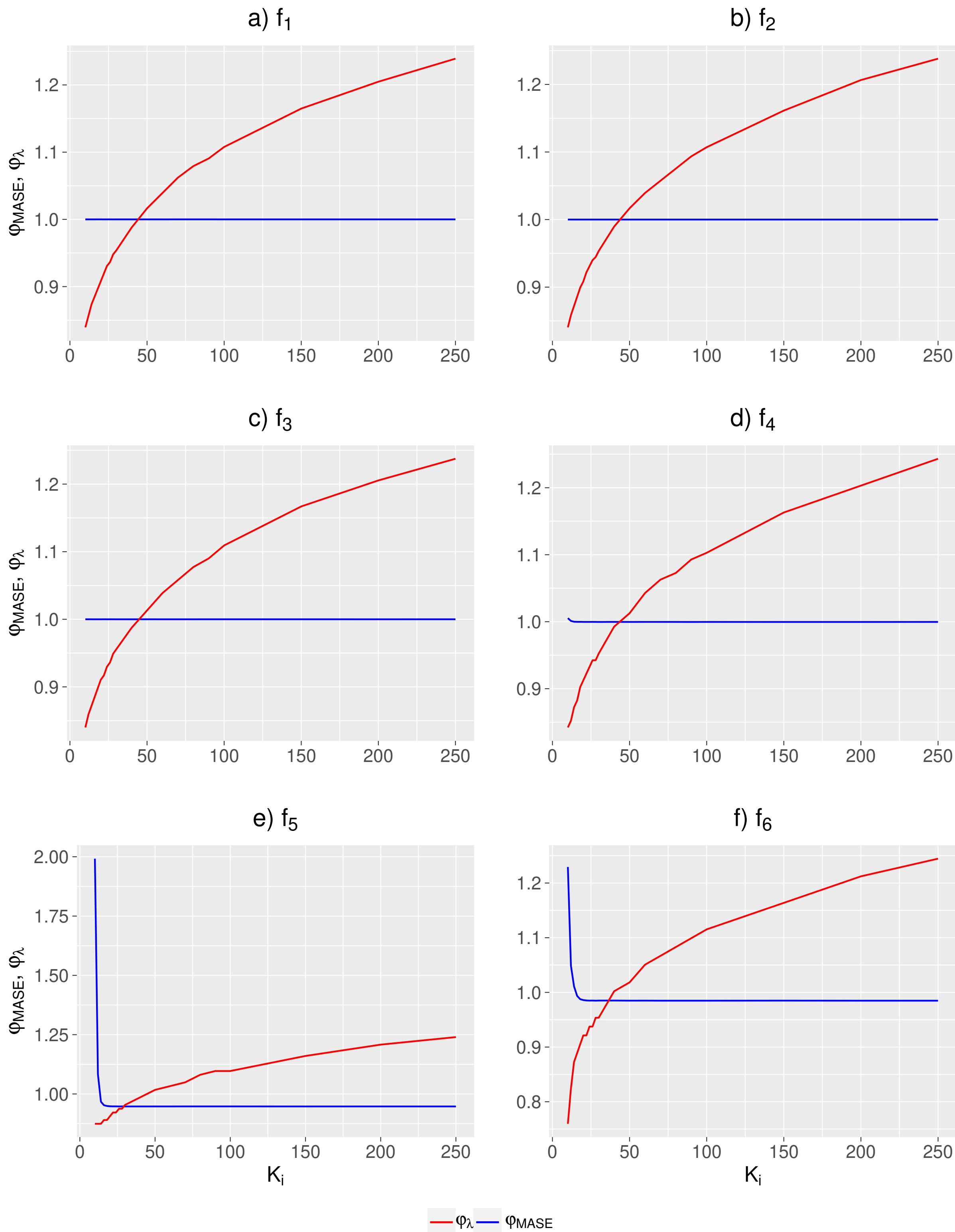

Figure 13: Case 11 - φ_λ and φ_{MASE} are indicated by the red and blue line, respectively.

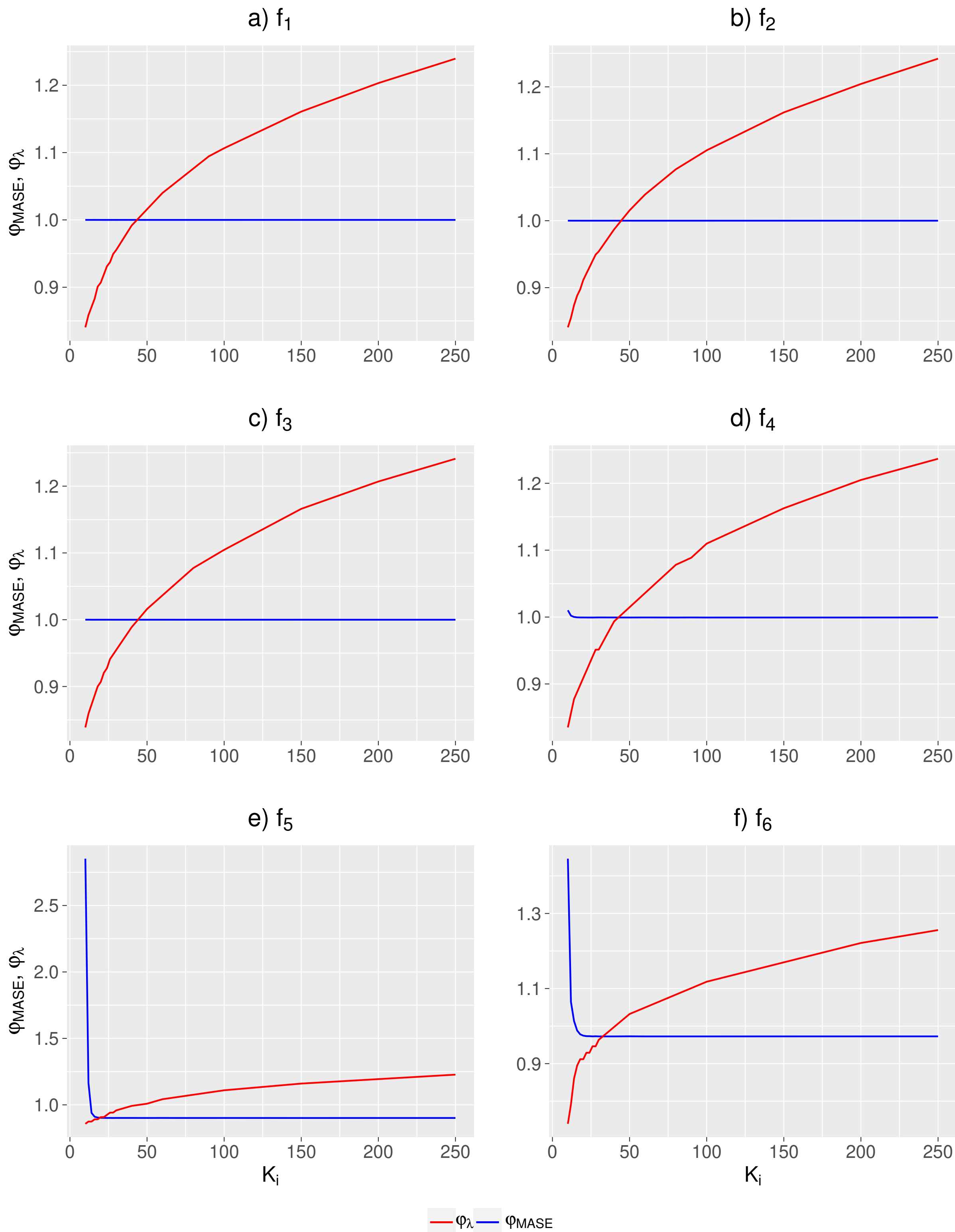

Figure 14: Case 21 - φ_λ and φ_{MASE} are indicated by the red and blue line, respectively.

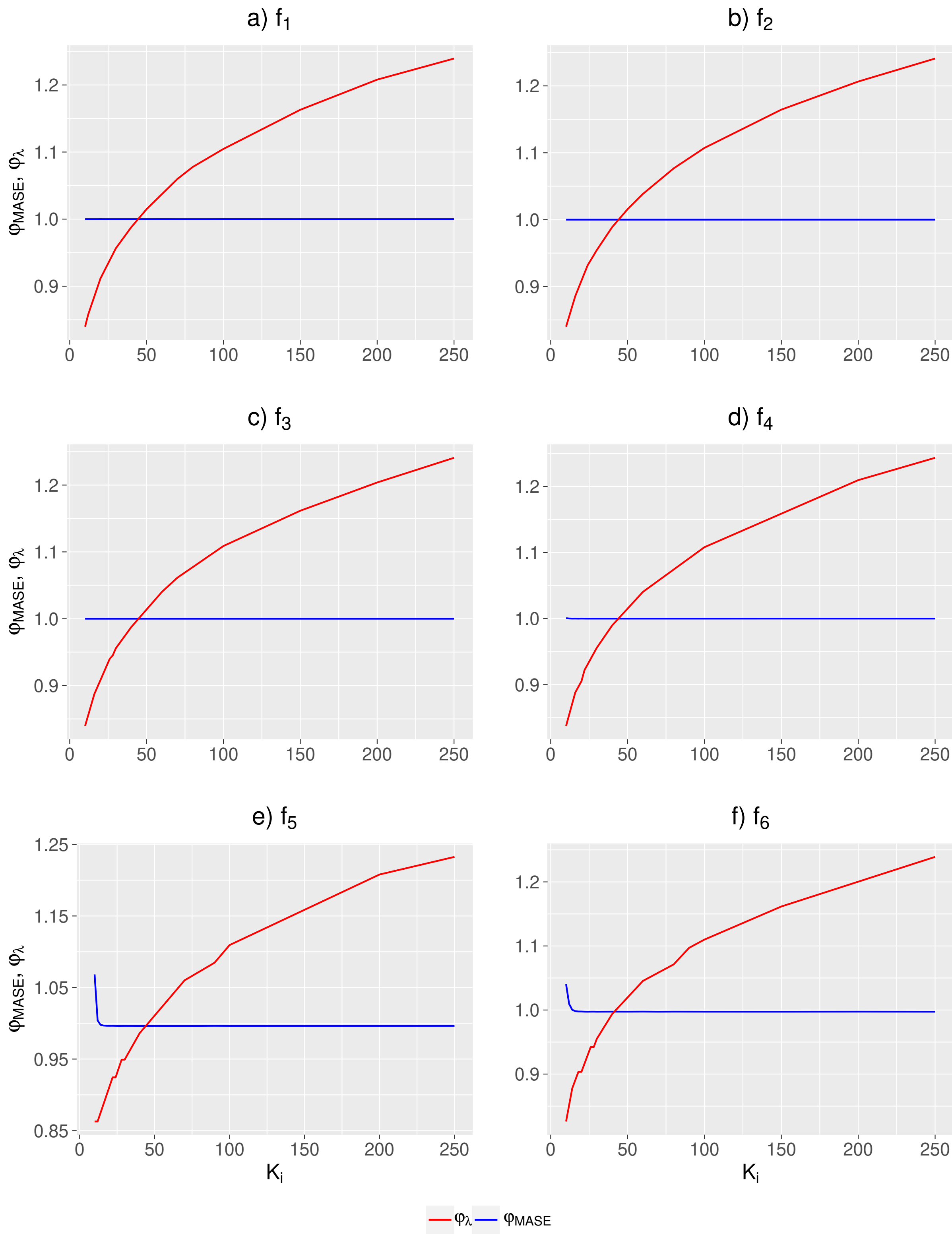

Figure 15: Case 22 - φ_λ and φ_{MASE} are indicated by the red and blue line, respectively.

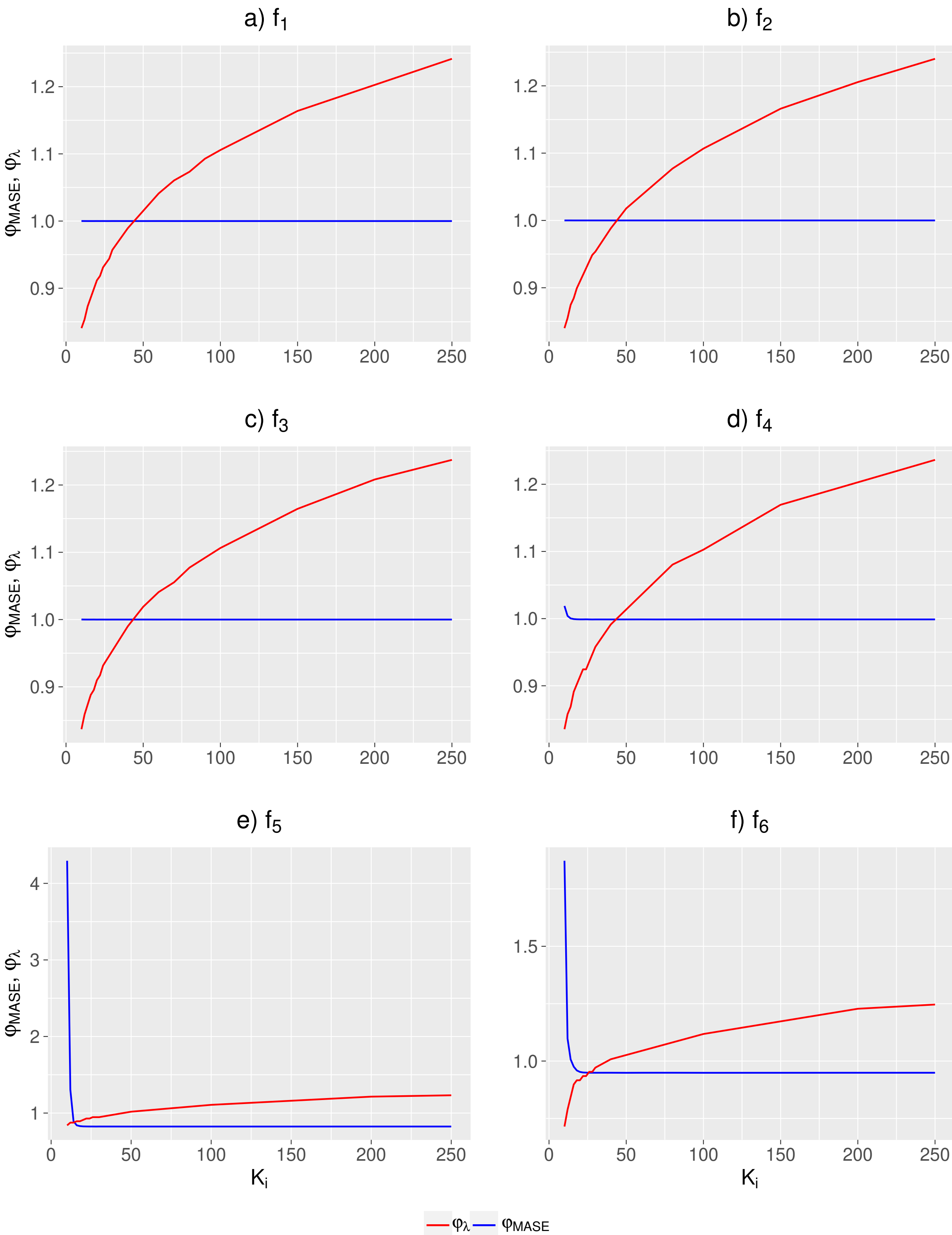

Figure 16: Case 31 - φ_λ and φ_{MASE} are indicated by the red and blue line, respectively.

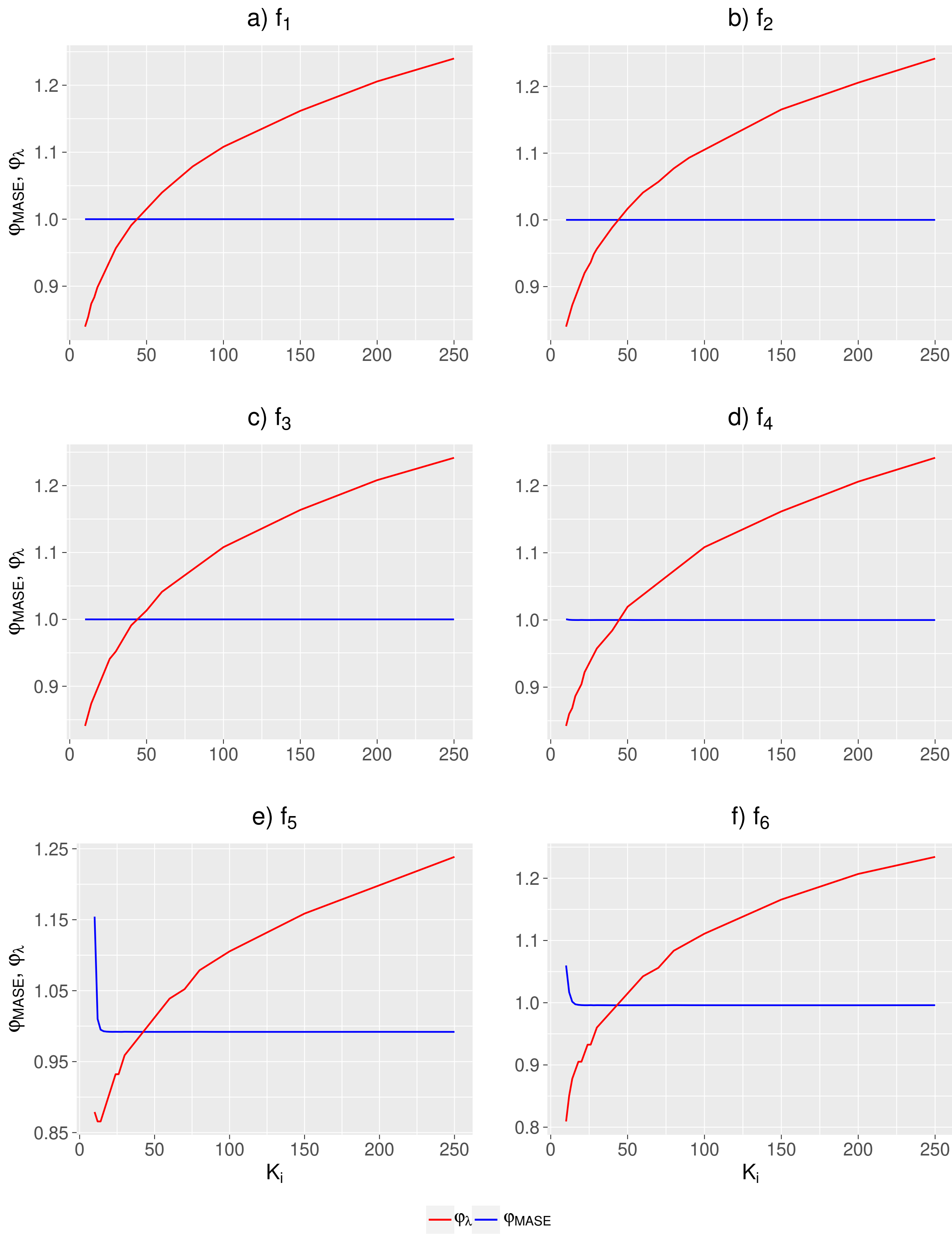

Figure 17: Case 32 - φ_λ and φ_{MASE} are indicated by the red and blue line, respectively.

Recent discussion papers

2022-04	Sebastian Letmathe Yuanhua Feng	An iterative plug-in algorithm for P-Spline regression
2022-03	Claus-Jochen Haake Martin Schneider	Playing games with QCA: Measuring the explanatory power of single conditions with the Banzhaf index
2022-02	Carina Burs Thomas Gries	Decision-making under Imperfect Information with Bayesian Learning or Heuristic Rules
2022-01	Claus-Jochen Haake Thomas Streck	Distortion through modeling asymmetric bargaining power
2021-09	Bernard M. Gilroy Marie Wegener Christian Peitz	COVID-19 and Triage - A Public Health Economic Analysis of a Scarcity Problem
2021-08	Bastian Schäfer	Bandwidth selection for the Local Polynomial Double Conditional Smoothing under Spatial ARMA Errors
2021-07	Sebastian Letmathe Jan Beran Yuanhua Feng	An extended exponential SEMIFAR model with application in R
2021-06	Yuanhua Feng Bastian Schäfer	Boundary modification in local polynomial regression*
2021-05	Bastian Schäfer Yuanhua Feng	Fast Computation and Bandwidth Selection Algorithms for Smoothing Functional Time Series*
2021-04	Yuanhua Feng	Uni- and multivariate extensions of the sinh-arcsinh normal distribution applied to distributional regression
2021-03	Sebastian Letmathe Yuanhua Feng André Uhde	Semiparametric GARCH models with long memory applied to Value at Risk and Expected Shortfall
2021-02	Claus-Jochen Haake Walter Trockel	Socio-legal Systems and Implementation of the Nash Solution in Debreu-Hurwicz Equilibrium
2021-01	Thomas Gries Paul Welfens	Testing as an Approach to Control the Corona Epidemic Dynamics and Avoid Lockdowns
2020-10	Fatma Aslan Papatya Duman Walter Trockel	Non-cohesive TU-games: Efficiency and Duality
2020-09	Yuanhua Feng Jan Beran Sebastian Letmathe Sucharita Ghosh	Fractionally integrated Log-GARCH with application to value at risk and expected shortfall
2020-08	Fatma Aslan Papatya Duman Walter Trockel	Removed from the CIE series
2020-07	Thomas Gries Veronika Müller	Conflict Economics and Psychological Human Needs
2020-06	Thomas Mayrhofer Hendrik Schmitz	Prudence and prevention – Empirical evidence*
2020-05	Daniel A. Kamhöfer Hendrik Schmitz Matthias Westphal	Marginal College Wage Premiums under Selection into Employment*