Analysis of Highdimensional Data by Semiparametric (Generalized) Regression Models

May 27, 2006

Marlene Müller

Fraunhofer Institut Techno- und Wirtschaftsmathematik

Michael G. Schimek

Medizinsche Universität Graz

Aim of this Study

- discuss different approaches to additive models (AM) and generalized additive models (GAM)
- include categorical variables $\Longrightarrow$ *partial linear* terms (combination of AM/PLM and GAM/GPLM)
- compare different approaches with respect to
  - underlying model is adequate (additive structure)
  - underlying model is non-adequate (non-additive structure)
- analysis of the computational effort of the techniques
- provide software $\Rightarrow$ R package **KernGPLM**

Outline

- semiparametric extensions of the generalized linear model (GLM)
  $\Rightarrow$ GPLM and GAM (generalized partial linear and additive models)
- introduce and compare different estimation approaches
- focus on techniques for **high-dimensional data**

Motivation

$\Rightarrow$ **Financial application: Credit Rating**
estimation of individual credit scores, default probabilities

$\Rightarrow$ **Parametric and Semiparametric Estimation**
logit/probit, nonparametric components, GPLM, GAM

Financial application: Credit Rating

- new interest in this field because of **Basel II**:
  capital requirements of a bank are adapted to the individual credit portfolio $\rightarrow$ **internal ratings-based approach (IRB approach)**

- key problems:
  determine **rating score** and subsequently **default probabilities (PDs)** as a function of some explanatory variables

  $\rightarrow$ classical logit/probit-type models to estimate linear predictors (scores) and probabilities (PDs)

  $\rightarrow$ **classification** problem with 2 groups, but focus on regression models as rating scores need to be **interpretable**

From: The New Basel Capital Accord ("Basel II"):

(www.bis.org)

> The bank must demonstrate that its criteria cover all factors that are relevant to the analysis of borrower risk. These factors should demonstrate an **ability to differentiate risk, have predictive and discriminatory power**, and be both plausible and intuitive in order to ensure that ratings are designed to distinguish risk rather than to minimise regulatory capital requirements.

This yields two objectives:

- study **single factors**
- find the **best model**

Data Example: sample of private loans

References: Fahrmeir/Hamerle (1984); Fahrmeir & Tutz (1995)

- default indicator: $Y \in \{0,1\}$, where $1 =$ default
- explanatory variables:
  personal characteristics, credit history, credit characteristics
- sample size: 1000 (stratified sample with 300 defaults)

Estimated (Logit) Scores

$$\begin{aligned}\text{Score} \;=\; & 0.162 - 0.696^{\star\star\star}\cdot \text{previous} + 0.496^{\star}\cdot (\text{d9-12}) + 0.818^{\star\star\star}\cdot (\text{d12-18}) \\ & +0.919^{\star\star\star}\cdot (\text{d18-24}) + 1.502^{\star\star\star}\cdot (\text{d}>24) - 0.91^{\star\star\star}\cdot \text{savings} \\ & -0.339^{\star}\cdot \text{purpose} + 0.976\cdot \text{foreign} + 0.614^{\star\star\star}\cdot \text{house} - 0.000277^{\star\star}\cdot \text{amount} \\ & -0.0971^{\star\star}\cdot \text{age} + 0.0000000185^{\star\star}\cdot \text{amount}^2 + 0.00086^{\star}\cdot \text{age}^2 \\ & +0.00000272\cdot (\text{amount} \cdot \text{age})\end{aligned}$$

*, **, *** denote significant coefficients at the 10%, 5%, 1% level, respectively

Parametric and Semiparametric Estimation

- parametric score and PD estimation (logit/probit)
- semiparametric score and PD estimation
  - ⋆ find relevant factors
  - ⋆ possibly use transformations for each of the factors

Two objectives:

$\Rightarrow$ search for **effects of single factors**

$\Rightarrow$ search for **best model**

Data Example: binary choice model

estimate the model (credit rating: estimates scores + PDs)

$$P(Y = 1|\boldsymbol{X}) = E(Y|\boldsymbol{X}) = G(\boldsymbol{\beta}^\top \boldsymbol{X})$$

$\Longrightarrow G$ is usually chosen as a cumulative distribution function

Parametric Models

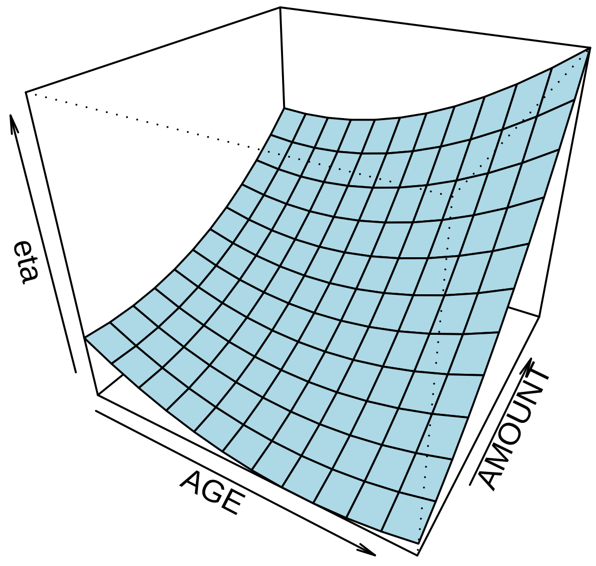
- logit

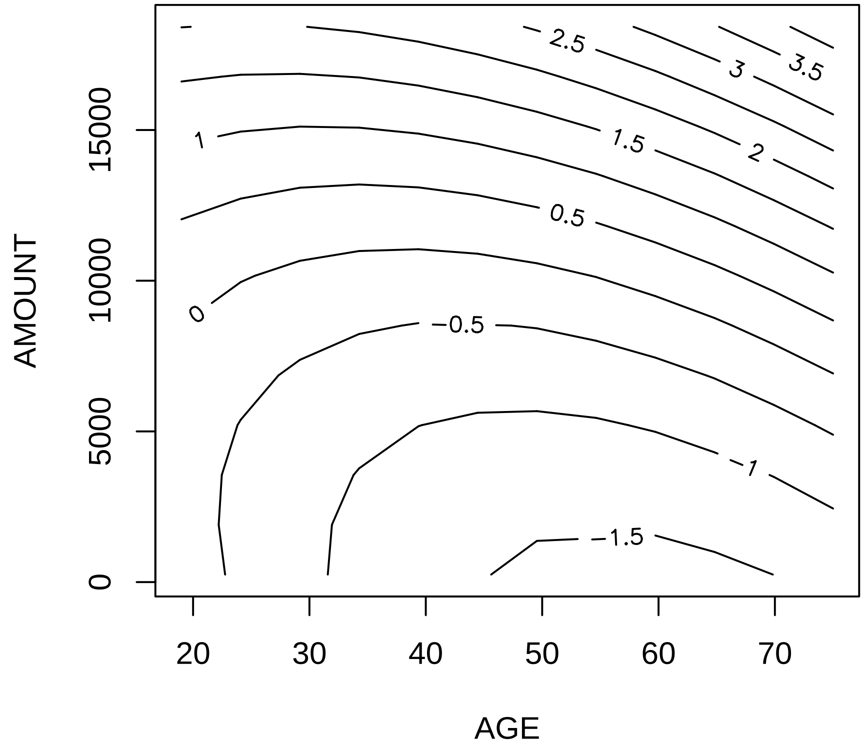
$$P(Y = 1|\boldsymbol{X}) = F(\boldsymbol{X}^\top \boldsymbol{\beta}), \quad F(\bullet) = \frac{1}{1 + e^{-\bullet}}$$

- probit

$$P(Y = 1|\boldsymbol{X}) = \Phi(\boldsymbol{X}^\top \boldsymbol{\beta}), \quad \Phi(\bullet) \text{ standard normal cdf}$$

Data Example: logit (with interaction)





credit default on AGE and AMOUNT using quadratic and interaction terms, left: surface and right: contours of the fitted score function

Semiparametric Models

- local regression

$$E(Y|\boldsymbol{T}) = G\{m(\boldsymbol{T})\}, \quad m \text{ nonparametric}$$

- generalized partial linear model (GPLM)

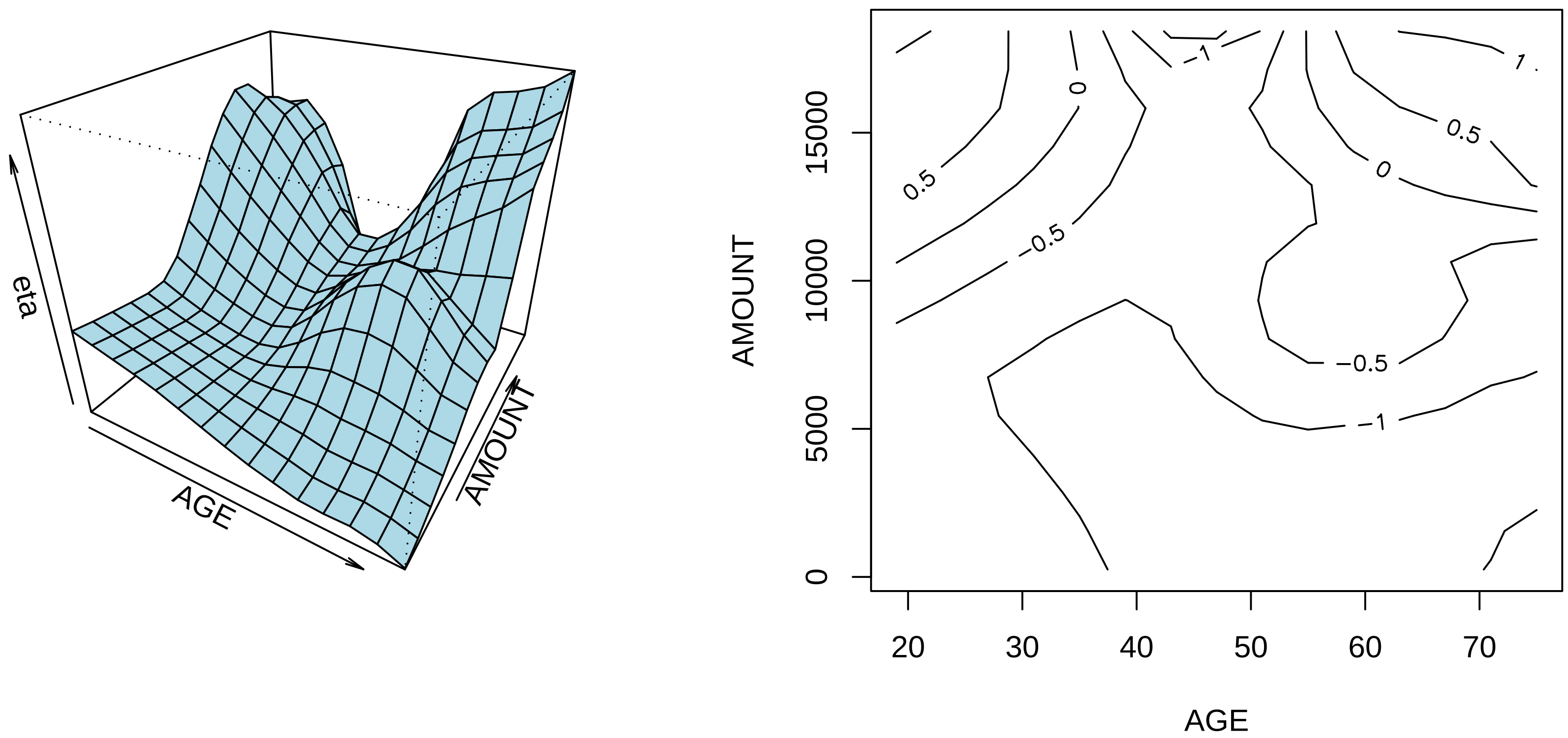
$$E(Y|\boldsymbol{X},\boldsymbol{T}) = G\left\{\boldsymbol{X}^\top\boldsymbol{\beta} + m(\boldsymbol{T})\right\} \quad m \text{ nonparametric}$$

- generalized additive partial linear model (semiparametric GAM)

$$E(Y|\boldsymbol{X},\boldsymbol{T}) = G\left\{\beta_0 + \boldsymbol{X}^\top\boldsymbol{\beta} + \sum_{j=1}^{p} m_j(T_j)\right\} \quad m_j \text{ nonparametric}$$

Some references:
Loader (1999), Hastie and Tibshirani (1990), Härdle et al. (2004), Green and Silverman (1994)

Data Example: generalized partial linear model (GPLM)



credit default on AGE and AMOUNT using a nonparametric function, left: surface and right: contours of the fitted score function on AGE and AMOUNT

Objectives

- obtain shape information: knowlegde about functional dependencies
- select the "optimal" set of predictors: estimate scores and PDs

$\Longrightarrow$ can be obtained by **backfitting and local scoring**

additional aspect (recall the Basel II document)

- estimation of *marginal effects*: identify relevant factors

$\Longrightarrow$ can be obtained by **marginal integration**

Note: the marginal effect represents the conditional expectation $E_{\varepsilon,\boldsymbol{T}_{\underline{\alpha}}}(Y|T_\alpha)$ where the expectation is not only taken on the error distribution but also on all other regressors

Estimation Approaches

- GPLM:
  - ⋆ generalization of Speckman's estimator (type of profile likelihood)
  - ⋆ backfitting for two additive components and local scoring

References:

(PLM) Speckman (1988), Robinson (1988); (PLM/splines) Schimek (2000), Eubank et al. (1998), Schimek (2002); (GPLM) Severini and Staniswalis (1994), Müller (2001)

- semiparametric GAM:
  - ⋆ [modified | smooth] backfitting and local scoring
  - ⋆ marginal [internalized] integration

References:

(marginal integraton) Tjøstheim and Auestad (1994), Chen et al. (1996), Hengartner et al. (1999), Hengartner and Sperlich (2005);
(backfitting) Buja et al. (1989), Mammen et al. (1999), Nielsen and Sperlich (2005)

Estimation of the GPLM

$$E(Y|\boldsymbol{X},\boldsymbol{T}) = G\left(\boldsymbol{X}^\top\boldsymbol{\beta} + m(\boldsymbol{T})\right)$$

- $\widehat{\beta}$ can be estimated if m known (parametric method, weighted LSE),
- $\widehat{m}$ can be estimated if β known (nonparametric method, e.g. Nadaraya–Watson type)

References:
Severini & Staniswalis (1994), Müller (2001)

Speckman estimator (for PLM)

$$
\begin{aligned}
Y &= \boldsymbol{\beta}^T \boldsymbol{X} + m(\boldsymbol{T}) + \varepsilon \\
E(Y|\boldsymbol{T}) &= \boldsymbol{\beta}^T E(\boldsymbol{X}|\boldsymbol{T}) + m(\boldsymbol{T}) + E(\varepsilon|\boldsymbol{T})
\end{aligned}
$$

$$
\underbrace{Y - E(Y|\boldsymbol{T})}_{\widetilde{Y}} = \boldsymbol{\beta}^T \underbrace{\{\boldsymbol{X} - E(\boldsymbol{X}|\boldsymbol{T})\}}_{\widetilde{\boldsymbol{X}}} + \underbrace{\varepsilon - E(\varepsilon|\boldsymbol{T})}_{\widetilde{\varepsilon}}
$$

matrix notation

$$
\widehat{\boldsymbol{\beta}} = \left(\widetilde{\mathcal{X}}^T \widetilde{\mathcal{X}}\right)^{-1} \widetilde{\mathcal{X}}^T \widetilde{\boldsymbol{Y}}
$$

$\widetilde{\mathcal{X}} = (\mathbf{I} - \mathbf{S})\mathcal{X}, \quad \widetilde{\boldsymbol{Y}} = (\mathbf{I} - \mathbf{S})\boldsymbol{Y}$

$\mathcal{X}$ design matrix, $\mathbf{S}$ smoother matrix, $\mathbf{I}$ identity matrix

$$
\widehat{\boldsymbol{m}} = \mathbf{S}\left(\boldsymbol{Y} - \mathcal{X}\widehat{\boldsymbol{\beta}}\right)
$$

Reference: Speckman (1988)

Generalized Speckman Estimator

- partial linear model (identity G)

$$E(Y|\boldsymbol{X},\boldsymbol{T}) = \boldsymbol{X}^T\boldsymbol{\beta} + m(\boldsymbol{T})$$

$$\Longrightarrow \qquad \begin{aligned} \boldsymbol{m}^{new} &= \mathbf{S}(\boldsymbol{Y} - \mathcal{X}\boldsymbol{\beta}) \\ \boldsymbol{\beta}^{new} &= (\widetilde{\mathcal{X}}^T\widetilde{\mathcal{X}})^{-1}\widetilde{\mathcal{X}}^T\widetilde{\boldsymbol{Y}} \end{aligned}$$

- generalized partial linear model

$$E(Y|\boldsymbol{X},\boldsymbol{T}) = G\{\boldsymbol{X}^T\boldsymbol{\beta} + m(\boldsymbol{T})\}$$

$\Longrightarrow$ above for adjusted dependent variable

$$\boldsymbol{Z} = \mathcal{X}\boldsymbol{\beta} + \boldsymbol{m} - \mathcal{W}^{-1}\boldsymbol{v},$$

$\boldsymbol{v} = (\ell_i')$, $\mathcal{W} = \mathrm{diag}(\ell_i'')$

References: Severini and Staniswalis (1994)

Comparison of Algorithms

	parametric step	nonparametric step	est. matrix
Speckman	$\beta^{new} = (\widetilde{\mathcal{X}}^T \mathcal{W} \widetilde{\mathcal{X}})^{-1} \widetilde{\mathcal{X}}^T \mathcal{W} \widetilde{\boldsymbol{Z}}$	$\boldsymbol{m}^{new} = \mathbf{S}(\boldsymbol{Z} - \mathcal{X}\beta)$	$\boldsymbol{\eta} = \mathcal{R}^S \boldsymbol{Z}$
Backfitting	$\beta^{new} = (\mathcal{X}^T \mathcal{W} \widetilde{\mathcal{X}})^{-1} \mathcal{X}^T \mathcal{W} \widetilde{\boldsymbol{Z}}$	$\boldsymbol{m}^{new} = \mathbf{S}(\boldsymbol{Z} - \mathcal{X}\beta)$	$\boldsymbol{\eta} = \mathcal{R}^B \boldsymbol{Z}$
Profile	$\beta^{new} = (\mathcal{X}^T \mathcal{W} \widetilde{\mathcal{X}})^{-1} \mathcal{X}^T \mathcal{W} \widetilde{\boldsymbol{Z}}$	$\boldsymbol{m}^{new} = \ldots$	$\boldsymbol{\eta} = \mathcal{R}^P \boldsymbol{Z}$

Speckman/Backfitting:
$\widetilde{\mathcal{X}} = (\mathbf{I} - \mathbf{S})\mathcal{X}$, $\widetilde{\boldsymbol{Z}} = (\mathbf{I} - \mathbf{S})\boldsymbol{Z}$, $\mathbf{S}$ weighted smoother matrix

Profile Likelihood:

$\widetilde{\mathcal{X}} = (\mathbf{I} - \mathbf{S}^P)\mathcal{X}$, $\widetilde{\boldsymbol{Z}} = (\mathbf{I} - \mathbf{S}^P)\boldsymbol{Z}$, $\mathbf{S}^P$ weighted (different) smoother matrix

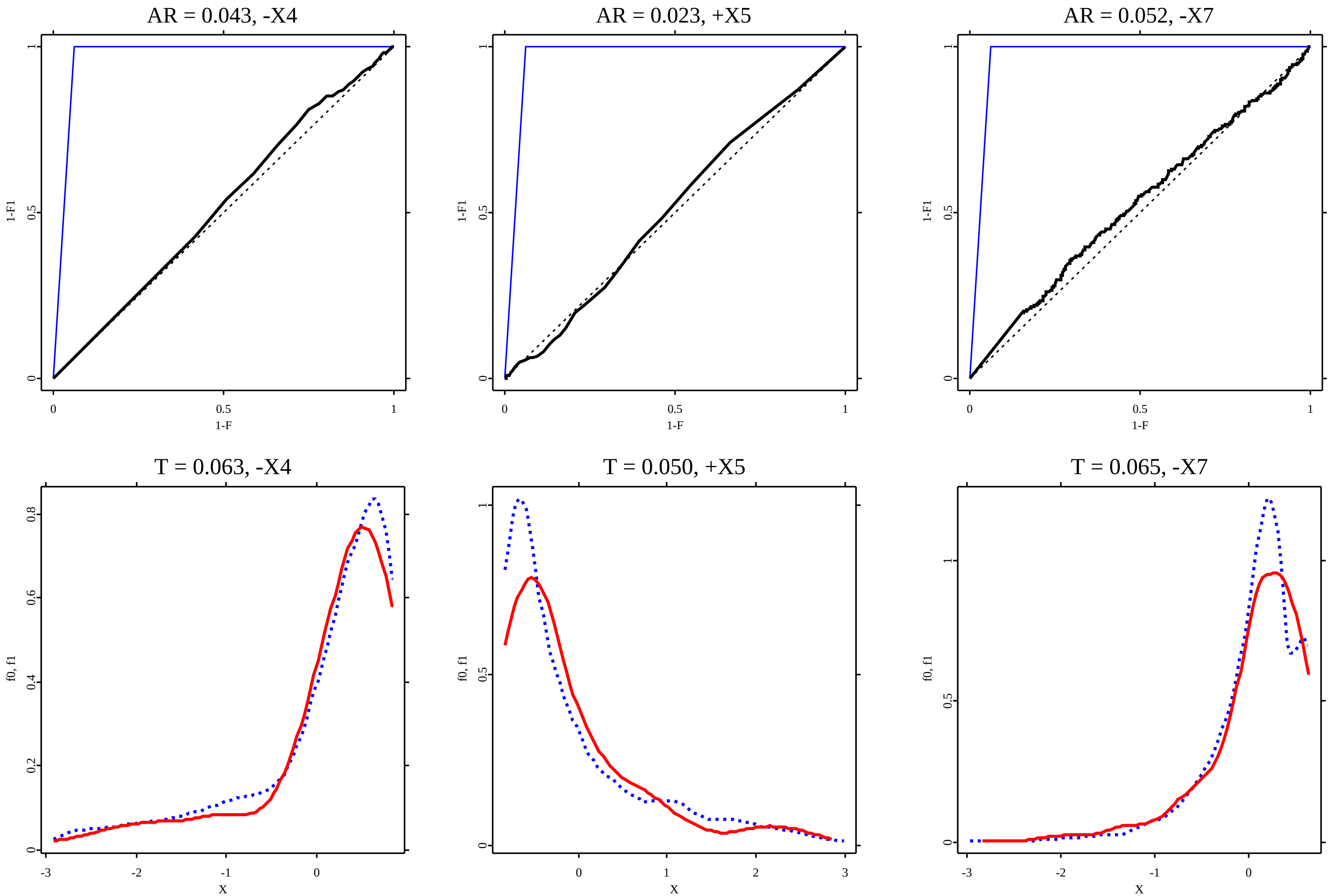
References: Severini and Staniswalis (1994), Müller (2001)

Data Example:

French Credit Data

- response variable Y
  (credit status, 0="Non-Default", 1="Default")
- metric variables X2 to X9
- categorical variables X10 to X24

	Estimation data set		Validation data set	
0 ("Non-Defaults")	5808	(94%)	1891	(94.6%)
1 ("Defaults")	372	(6%)	107	(5.4%)
total	6180		1998	



Lorenz performance curves, density estimates (conditional on Y, red=default) for X4, X5, X7.

GPLM/GAM Application

- to include variable X5 in a nonlinear way:

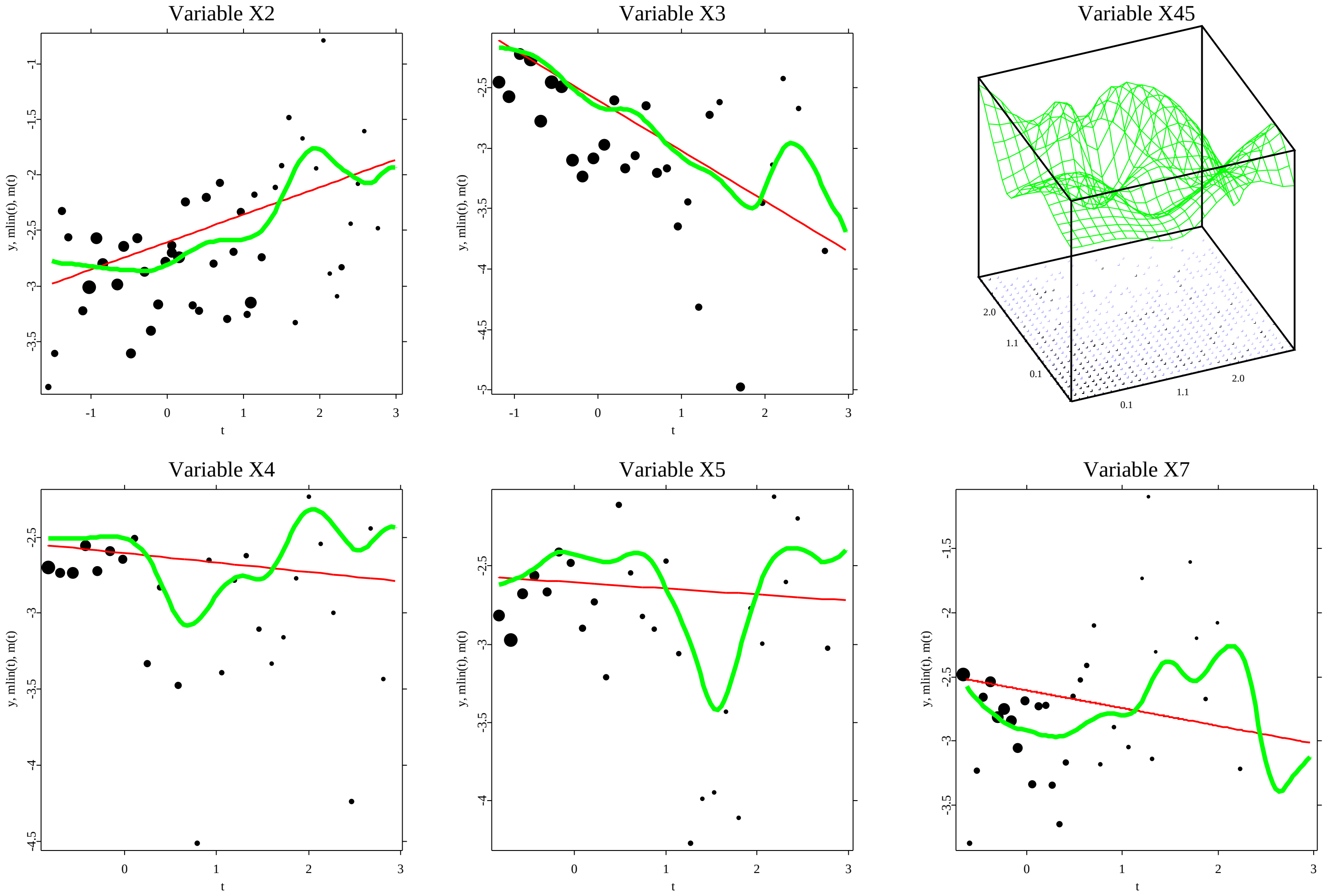
$$P(Y=1|X_{-5},X_5) = F\left(\sum_{j\neq 5}\beta_j^\top X_j + m_5(X_5)\right)$$

- to include variables X4, X5 in a nonlinear way:

$$P(Y=1|X_{-4,-5},(X_4,X_5)) = F\left(\sum_{j\neq 4,5}\beta_j^\top X_j + m_{45}(X_4,X_5)\right)$$

		nonparametric in						
	Logit	X2	X3	X4	X5	X7	X4,X5	X2, X4,X5
constant	**-2.605**	–	–	–	–	–	–	–
X2	**0.247**	–	**0.243**	**0.241**	**0.243**	**0.233**	**0.228**	–
X3	**-0.417**	**-0.414**	–	**-0.414**	**-0.416**	**-0.417**	**-0.408**	**-0.399**
X4	-0.062	-0.052	-0.063	–	-0.065	-0.054	–	–
X5	-0.038	-0.051	-0.045	-0.034	–	-0.042	–	–
X6	**0.188**	**0.223**	**0.193**	**0.190**	**0.177**	**0.187**	0.176	**0.188**
X7	-0.138	-0.138	-0.142	-0.131	-0.146	–	-0.135	-0.128
X8	**-0.790**	**-0.777**	**-0.800**	**-0.786**	**-0.796**	**-0.793**	**-0.792**	**-0.796**
X9	**-1.215**	**-1.228**	**-1.213**	**-1.222**	**-1.216**	**-1.227**	**-1.214**	**-1.215**

Parametric coefficients for X2 to X9. Bold values are significant at 5%.



Marginal dependencies, variables X2, X3, X4+X5, X4, X5 and X5.

Parametric logit fits (red) and GPLM logit fits (green).

Estimation of the GAM

$$E(Y|\boldsymbol{X},\boldsymbol{T}) = G\left\{\beta_0 + \boldsymbol{X}^\top \boldsymbol{\beta} + \sum_{j=1}^{p} m_j(T_j)\right\} \quad m_j \text{ nonparametric}$$

- **classical backfitting:** fit single components by regression on the residuals w.r.t. the other components
- **modified backfitting:** first project on the linear space spanned by all regressors and then nonparametrically fit the partial residuals
- **marginal (internalized) integration:** estimate the marginal effect by integrating a full dimensional nonparametric regression estimate
  $\Longrightarrow$ original proposal is computationally intractable: $O(n^3)$
  $\Longrightarrow$ choice of nonparametric estimate is essential: **marginal internalized integration**

Comparison of Algorithms [State of the Art (?)]

- consistency of marginal integration:
  Hengartner et al. (1999); Hengartner and Sperlich (2005)
  $\Rightarrow$ **shows that marginal (internalized) integration works, no comparison with backfitting**

- optimal rate of convergence for marginal integration:
  Hengartner et al. (1999)
  $\Rightarrow$ **marginal (internalized) integration + one backfitting step yields an oracle efficient estimator for additive components**

- comparison of backfitting and marginal integration:
  Sperlich et al. (1999); Martins-Filho and Yang (2004)
  $\Rightarrow$ **additive components functions are generally more precisely fitted with backfitting, in particular due to boundary effects**

$\Longrightarrow$ all authors use their own implementation, no generally and publicly available code

Marginal (Internalized) Integration

marginal effect of regressor T_j

$$r_j(T_j) = E_{\boldsymbol{T}_{-j}}\{m(\boldsymbol{T})\}$$

if m is truly additive, i.e. $m(\boldsymbol{T}) = c + m_1(T_1) + \ldots + m_p(T_p)$, then

$$r_j(T_j) = c + m_j(T_j)$$

Hengartner et al. (1999); Hengartner and Sperlich (2005) propose to use instruments (m_{-j} denoting $\sum_{\alpha \neq j} m_\alpha$)

$$E(\xi_j | T_j = t_j) = 1 \quad \text{and} \quad E\{m_{-j}(\boldsymbol{T}_{-j}) \cdot \xi_j | T_j = t_j\} = 0$$

$\Longrightarrow$

$$E(Y\xi_j | T_j = t_j) = r_j(t_j) = c + m_j(t_j)$$

Marginal (Internalized) Integration:

$$E(Y\xi_j | T_j = t_j) = r_j(t_j) = c + m_j(t_j)$$

- estimate additive component function by regression of $Y\widehat{\xi_j}$ on T_j
- no iteration required! (but additional estimate of the instrument ξ_j)
- numerous smoothing parameters have to be chosen: for the instruments, for the regression
- estimator can be seen as an internalized version of the original marginal integration estimator by Tjøstheim and Auestad (1994), Chen et al. (1996)

Estimation of Instruments

$$\xi_j = \frac{f_j(T_j) f_{-j}(\boldsymbol{T}_{-j})}{f(\boldsymbol{T})} = \frac{f_j(T_j)}{f(T_j | \boldsymbol{T}_{-j})}$$

f_j, f_{-j} and f denoting the pdfs of T_j, $\boldsymbol{T}_{-j}$ and $\boldsymbol{T}$, respectively

$$\widehat{r}_j(t_j) = \frac{1}{n} \sum_{i=1}^{n} K_{h_j}(t_j - T_{ij}) \underbrace{\frac{1}{\widehat{f}(T_{ij} | \boldsymbol{T}_{i,-j})}}_{\text{to estimate!}} Y_i = \mathbf{S}_j \left(\widehat{\xi}_{ij} Y_i \right) ,$$

$\mathbf{S}_j$ denoting a univariate smoother here

Estimation of Instruments (cont'd)

we need an estimate for

$$\widehat{f}(T_j | \boldsymbol{T}_{-j})$$

$\Longrightarrow$ **How to solve this?**

- by nonparametric KDE ("classical" marginal integration)
  $\rightarrow$ not feasible in large dimensions!
- under normality assumption
  $\rightarrow$ restrictive!
- normality $\approx$ "linear regression" of T_j on $\boldsymbol{T}_{-j}$

  estimate 1: estimate pdf of T_j given linear regression of T_j on $\boldsymbol{T}_{-j}$

  estimate 2: estimate pdf of T_j given additive regression of T_j on $\boldsymbol{T}_{-j}$

  $\rightarrow$ means to use **dimension reduction techniques** here

Extension to Partial Linear and Generalized Cases

important: avoid any multi-dimensional smoothing!

- two-step approach for PLM (with additive components)
  - estimate additive components by a Speckman-type approach

$$
\begin{aligned}
\widetilde{\boldsymbol{\beta}}_j &= \left\{\mathcal{X}^\top (\mathbf{I}-\mathbf{S}_j)^\top (\mathbf{I}-\mathbf{S}_j)\mathcal{X}\right\}^{-1} \mathcal{X}^\top (\mathbf{I}-\mathbf{S}_j)^\top (\mathbf{I}-\mathbf{S}_j)\,\widehat{\boldsymbol{\xi}}_j \boldsymbol{Y} \\
&= \left(\widetilde{\mathcal{X}}_j^\top \widetilde{\mathcal{X}}_j\right)^{-1} \widetilde{\mathcal{X}}_j^\top (\mathbf{I}-\mathbf{S}_j)\,\widehat{\boldsymbol{\xi}}_j \boldsymbol{Y} \qquad \text{where} \quad \widetilde{\mathcal{X}}_j = (\mathbf{I}-\mathbf{S}_j)\mathcal{X}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{r}}_j &= \mathbf{S}_j(\widehat{\boldsymbol{\xi}}_j \boldsymbol{Y} - \mathcal{X}\widetilde{\boldsymbol{\beta}}_j) \\
\widehat{\boldsymbol{m}}_j &= \widehat{\boldsymbol{r}}_j - \overline{\widehat{\boldsymbol{r}}_j}
\end{aligned}
$$

  - fit the linear part by a regression on the residuals w.r.t. the additive part
- similar two-step approach for GPLM (with additive components) by applying the above on the adjusted dependent variable

Simulation Example: true underlying regression is additive

nonparametric part:

$$m(\boldsymbol{T}) = \underbrace{\cos(T_1) - E\cos(T_1)}_{m_1(T_1)} + \underbrace{\cos(T_2) - E\cos(T_2)}_{m_2(T_2)}$$

where

$$T_1 \sim N(0,1), \quad T_2 = 0.8\,(T_1^2 - 1) + 0.2\,U,\ U \sim N(0,1)$$

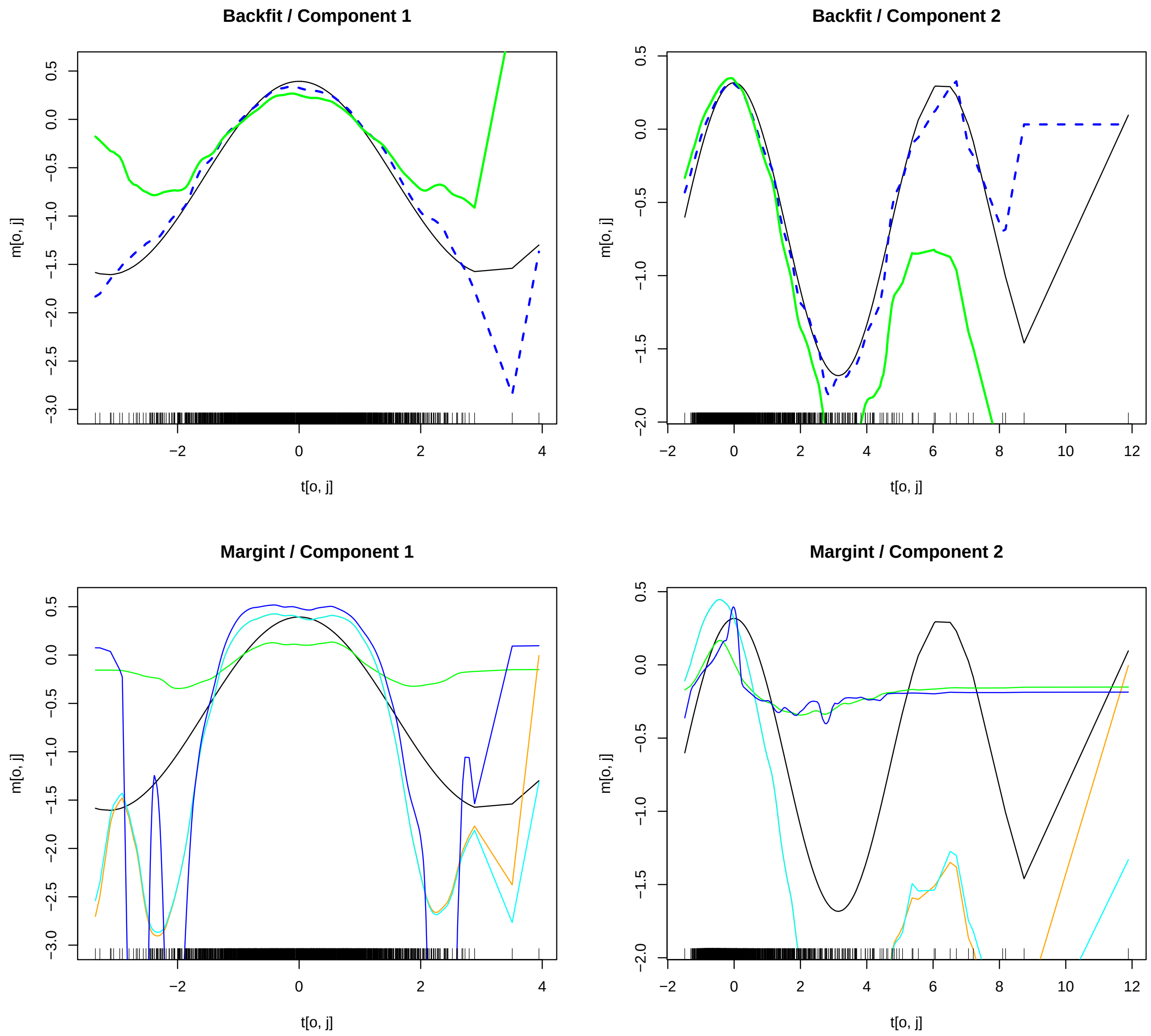
(quadratic dependence of the regressors!)

sample size: $n = 2000$

→ marginal integration?

→ marginal integration with <u>one</u> additional backfitting step?

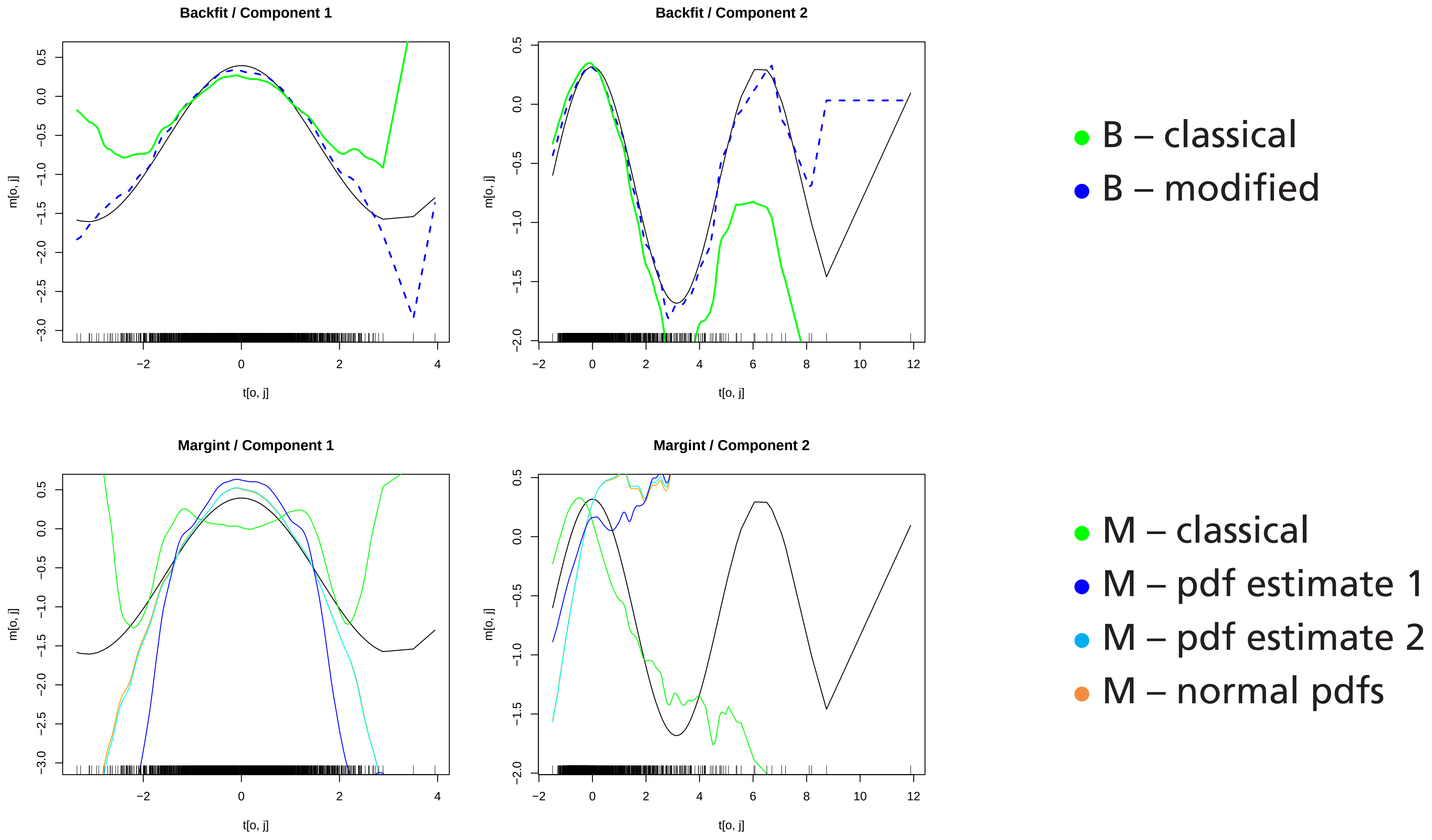
→ marginal integration to initialize backfitting (replacing the usual zero-functions)



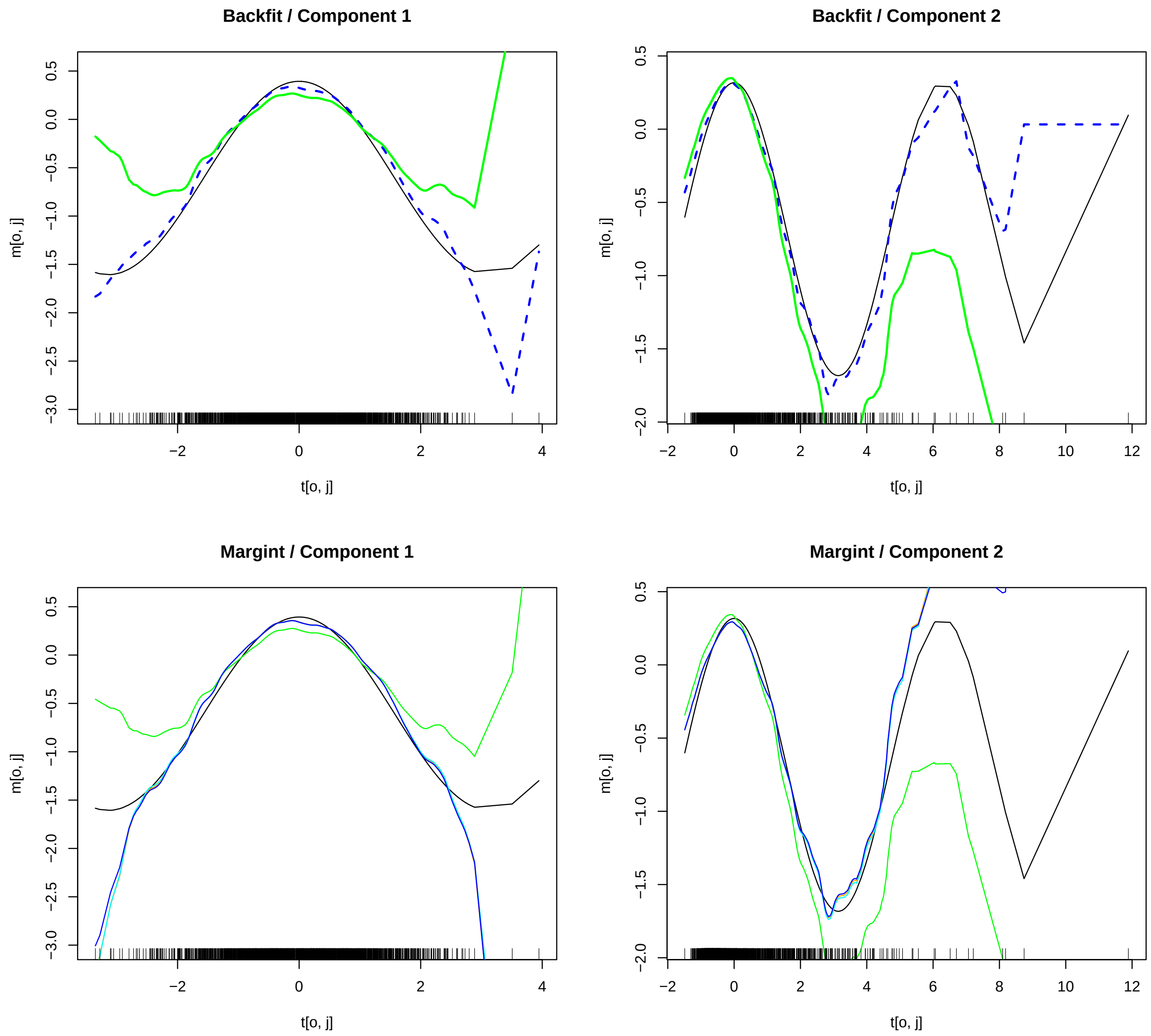
- B – classical
- B – modified

- M – classical
- M – pdf estimate 1
- M – pdf estimate 2
- M – normal pdfs

Marginal integration – no subsequent backfitting



Marginal integration – one subsequent backfitting



● B – classical
● B – modified

● M – classical
● M – pdf estimate 1
● M – pdf estimate 2
● M – normal pdfs

Marginal integration – as initialization for backfitting

Simulation Example: true underlying regression is non-additive

nonparametric part:

$$m(\boldsymbol{T}) = \cos\{\alpha T_1 + (1-\alpha)T_2\}, \quad \alpha = 0.4$$

where

$$T_1, T_2 \sim N(0,1), \mathrm{cov}(T_1, T_2) = 0.8$$

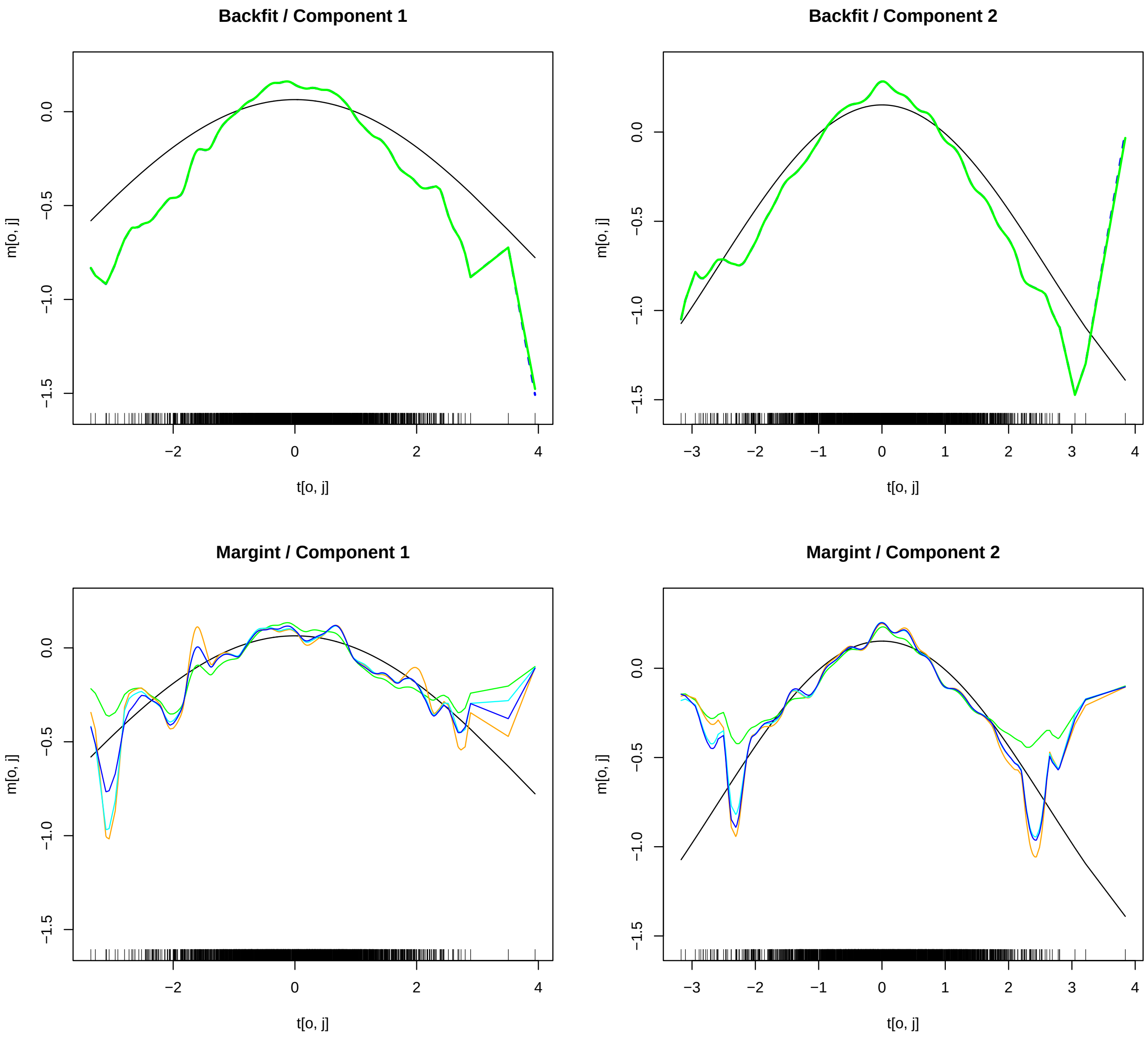
(linear correlation of the regressors)

$\Longrightarrow$ marginal effects:

$$\widetilde{m}_1(T_1) = e^{-(1-\alpha)^2/2}\{\cos(\alpha T_1) - e^{-\alpha^2/2}\}, \quad \widetilde{m}_2(T_2) = e^{-\alpha^2/2}[\cos\{(1-\alpha)T_1\} - e^{-(1-\alpha)^2/2}]$$

sample size: $n = 2000$

$\longrightarrow$ marginal integration estimates marginal effects?



- B – classical
- B – modified

- M – classical
- M – pdf estimate 1
- M – pdf estimate 2
- M – normal pdfs

Marginal integration – estimate of marginal effects

Comparison of Algorithms [Recent Findings]

- consistency of marginal integration:
  $\Rightarrow$ *if underlying function is truly additive, backfitting outperforms marginal integration*
  $\Rightarrow$ *consider marginal integration to initialize backfitting (replacing the usual zero-functions)*

- comparison of backfitting and marginal integration:
  $\Rightarrow$ *marginal integration indeed estimates marginal effects, but large number of observations is needed*
  $\Rightarrow$ *estimation method of the instruments is essential, dimension reduction techniques are required*

$\Longrightarrow$ R package **KernGPLM** for application

Summary

- GPLM and semiparametric GAM are natural extensions of the GLM
- different techniques to estimate these models are particularly useful because of:
  - ⋆ shape informations to obtain transformations
  - ⋆ marginal effects to identify relevant factors
- *large amount of data* is needed for estimating marginal effects

Work in Progress:

⇒ R package **KernGPLM** with routines for

- ⋆ (kernel based) generalized partial linear and additive models
- ⋆ additive components [modified | smooth] backfitting and local scoring
- ⋆ additive components through marginal [internalized] integration

References

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics*, 17:453–555.

Chen, R., Härdle, W., Linton, O., and Severance-Lossin, E. (1996). Estimation and variable selection in additive nonparametric regression models. In Härdle, W. and Schimek, M., editors, *Proceedings of the COMPSTAT Satellite Meeting Semmering 1994*, Heidelberg. Physica Verlag.

Eubank, R. L., Kambour, E. L., Kim, J. T., Klipple, K., Reese, C. S., and Schimek, M. G. (1998). Estimation in partially linear models. *Computational Statistics & Data Analysis*, 29:27–34.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*, volume 58 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric Modeling: An Introduction*. Springer, New York.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

Hengartner, N., Kim, W., and Linton, O. (1999). A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. *Journal of Computational and Graphical Statistics*, 8:1–20.

Hengartner, N. and Sperlich, S. (2005). Rate-optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis*, 95:246–272.

Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York.

Mammen, E., Linton, O., and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics*, 27:1443–1490.

Martins-Filho, C. and Yang, K. (2004). Finite sample performance of backfitting, marginal integration and two stage estimators under common bandwidth selection criterion. http://oregonstate.edu/~martinsc/martins-filho-yang(04)_v3.pdf.

Müller, M. (2001). Estimation and testing in generalized partial linear models — a comparative study. *Statistics and Computing*, 11:299–309.

Nielsen, J. and Sperlich, S. (2005). Smooth backfitting in practice. *Journal of the Royal Statistical Society, Series B*, 67:43–61.

Robinson, P. M. (1988). Root n–consistent semiparametric regression. *Econometrica*, 56:931–954.

Schimek, M. G. (2000). Estimation and inference in partially linear models with smoothing splines. *Journal of Statistical Planning and Inference*, 91:525–540.

Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, 89:501–511.

Speckman, P. E. (1988). Regression analysis for partially linear models. *Journal of the Royal Statistical Society, Series B*, 50:413–436.

Sperlich, S., Linton, O., and Härdle, W. (1999). Integration and backfitting methods in additive models: Finite sample properties and comparison. *Test*, 8:419–458.

Tjøstheim, D. and Auestad, B. (1994). Nonparametric identification of nonlinear time series: Projections. *Journal of the American Statistical Association*, 89:1398–1409.