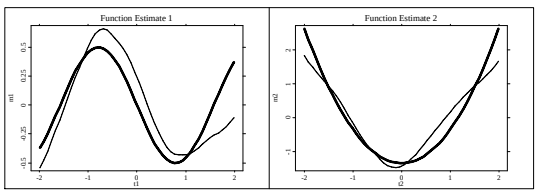
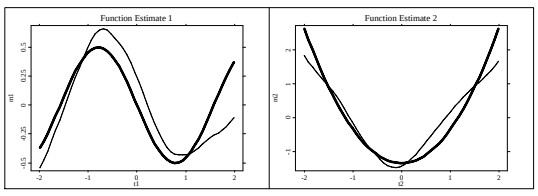
Semiparametric additive indices for binary response and generalized additive models

Wolfgang Härdle
Sylvie Huet
Enno Mammen
Stefan Sperlich

Sonderforschungsbereich 373
Wirtschaftswissenschaftliche Fakultät
Humboldt–Universität zu Berlin
sfb.wiwi.hu-berlin.de



1. Partial linear additive models
2. Additive binary response
3. Generalized additive models
4. Bootstrap bias correction
5. Bootstrap tests for linearity
6. Uniform bootstrap confidence bands
7. Simulations
8. East-West Migration
9. Apprenticeship



1 Partial linear additive models

econometric model

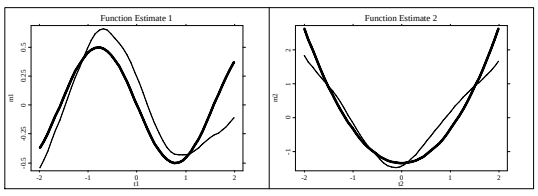
$$m(x,t) = E[Y|X = x, T = t]$$

traditional approach

$$m(x,t) = G(x^T\beta + \alpha + t^T\gamma)$$

where

$X \in I\!R^q$	covariable
$T \in I\!R^p$	cts covariable
$Y \in I\!R$	response
G	link function
β, γ	parameters



problem of misspecification

Burda (1993) East-West Migration

Fahrmeir and Tutz (1994) credit scoring

Bertschek (1996) firm innovations

Horowitz and Härdle (1996) binary response

Partial linear model

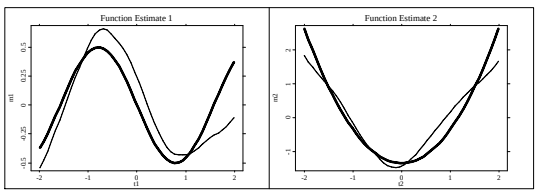
$$m(x,t) = G\{\beta^T x + m(t)\}$$

Severini and Staniswalis (1994)

Ai and McFadden (1997)

Carroll, Fan, Gijbels and Wand (1995)

Härdle, Mammen and Müller (1998)



problem of dimensionality

$$m : I\!R^p \longrightarrow I\!R$$

$$\inf_{\widehat{m}} \sup_{m} \|\widehat{m} - m\|_2 = O(n^{\frac{-l}{2l+p}})$$

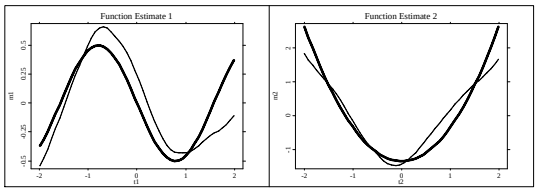
additive structure

$$m(t) = \sum_{j=1}^{p} m_j(t_j)$$

$$\inf_{\widehat{m}} \sup_{m(add)} \|\widehat{m} - m\|_2 = O(n^{\frac{-l}{2l+1}})$$

partial linear additive model

$$m(x,t) = G\{x^T\beta + \alpha + \sum_{j=1}^{p} m_j(t_j)\}$$



partial linear additive model

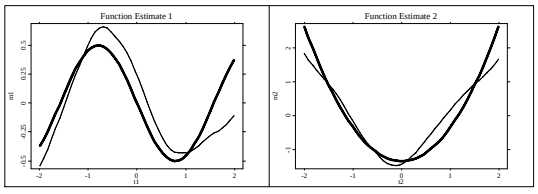
$$m(x,t) = G\{x^T\beta + \alpha + \sum_{j=1}^{p} m_j(t_j)\}$$

Backfitting

Buja, Hastie and Tibshirani (1989)
Opsomer and Ruppert (1997)
Linton, Mammen and Nielsen (1998)

Integration

Tjøstheim and Auestad (1994)
Linton and Nielsen (1995)
Fan, Härdle and Mammen (1998)
Severance-Lossin and Sperlich (1997)
Yang, Härdle and Nielsen (1998)



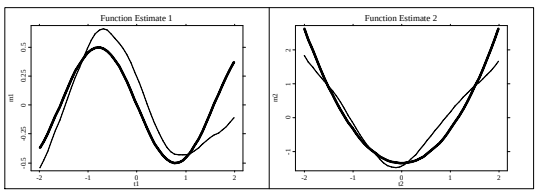
2 Additive binary response

Example1 Y = migration decision
Example2 Y = unemployed apprentice

Observe $(Y_i, X_i, T_i),\ Y_i \in \{0, 1\}$

$$P(Y = 1|X = x, T = t)$$
$$= G\{x^T\beta + \alpha + m_1(t_1) + \cdots + m_d(t_d)\}$$

G	known link function
$\beta \in I\!R^p$	unknown parameter
$m_j : I\!R \longrightarrow I\!R$	unknown functions
$\alpha \in I\!R$	
$E[w_j(T_j)m_j(T_j)] = 0$	for weight functions w_j



Conditional likelihood of $\{Y_i\}_{i=1}^n$

$$\begin{aligned} Q(\mu_i, Y_i) &= Y_i \log \mu_i + (1 - Y_i) \log(1 - \mu_i) \\ \mu_i &= G\{X_i^T \beta + \alpha + m_1(T_{i,1}) + \cdots + m_d(T_{i,d})\} \\ \mathcal{L}(m^+, \beta) &= \sum_{i=1}^{n} Q(\mu_i, Y_i) \end{aligned}$$

where

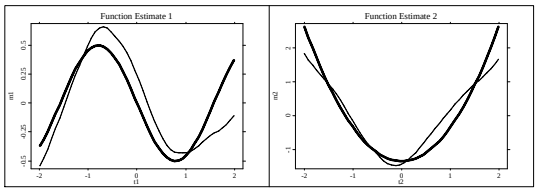
$$m^+(t) = \alpha + m_1(t_1) + \cdots + m_d(t_d)$$

Estimation of m_1

$$\begin{aligned} &\mathcal{L}^S(m^+, \beta) \\ &= \int \sum_{i=1}^{n} K_h(t_1 - T_{i,1}) L_g(t_{-1} - T_{i,-1}) \\ &\quad Q\left[G\{X_i^T \beta + m^+(t)\}; Y_i\right] dt \end{aligned}$$

where

$$\begin{aligned} u_{-1} &= (u_2, \ldots, u_d)^T \\ L_g(v) &= g^{-(d-1)} L(g^{-1} v) \\ K_h(v) &= h^{-1} K(h^{-1} v) \end{aligned}$$



Severini and Wong (1992)

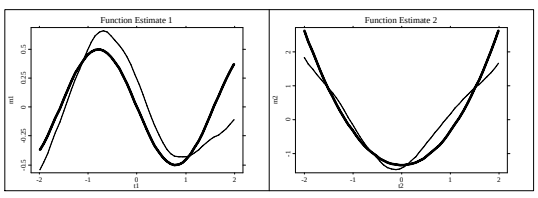
Severini and Staniswalis (1994)

Härdle, Mammen and Müller (1998)

Define for β

$$\begin{aligned}
\widehat{m}_\beta(t) &= \arg\max_\eta \sum_{i=1}^{n} K_h(t_1 - T_{i,1}) L_g(t_{-1} - T_{i,-1}) \\
&\quad Q\left[G\{X_i^T\beta + \eta\}; Y_i\right] \\
\widehat{\beta} &= \arg\max_{\beta \in B} \mathcal{L}(\widehat{m}_\beta, \beta) \\
\widehat{m} &= \widehat{m}_{\widehat{\beta}}
\end{aligned}$$

Result: $\widehat{m}$ a multivariate kernel estimator of m^+ with no additive structure.
Integration method applied to $\widehat{m}$ will give an additive estimate.



Integration method

$$
\begin{aligned}
m_1(t_1) &= \{\int w_{-1}(v)dv\}^{-1} \int w_{-1}(v) m^+(t_1, v)dv \\
&\quad + const. \\
&= \{\frac{1}{n}\sum_{i=1}^{n} w_{-1}(T_{i,-1})\}^{-1} \frac{1}{n}\sum_{i=1}^{n} w_{-1}(T_{i,-1}) \\
&\quad m^+(t_1, T_{i,-1}) + const.
\end{aligned}
$$

Define

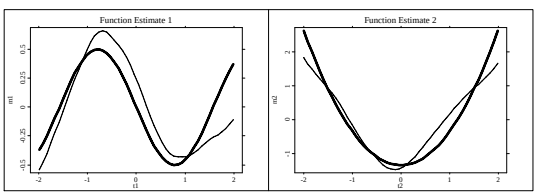
$$
\begin{aligned}
\overline{m}_1(t_1) &= \frac{\frac{1}{n}\sum_{i=1}^{n} w_{-1}(T_{i,-1})\widehat{m}(t_1, T_{i,-1})}{\frac{1}{n}\sum_{i=1}^{n} w_{-1}(T_{i,-1})} \\
\widehat{m}_1(t_1) &= \overline{m}_1(t_1) - \frac{1}{n}\sum_{i=1}^{n} \frac{w_1(T_{i,1})\overline{m}_1(T_{i,1})}{\frac{1}{n}\sum_{i=1}^{n} w_1(T_{i,1})}
\end{aligned}
$$

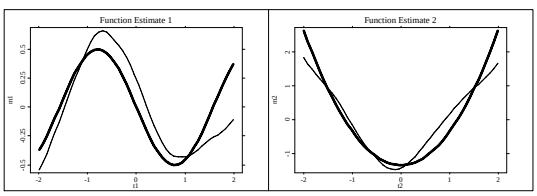
Additive constant

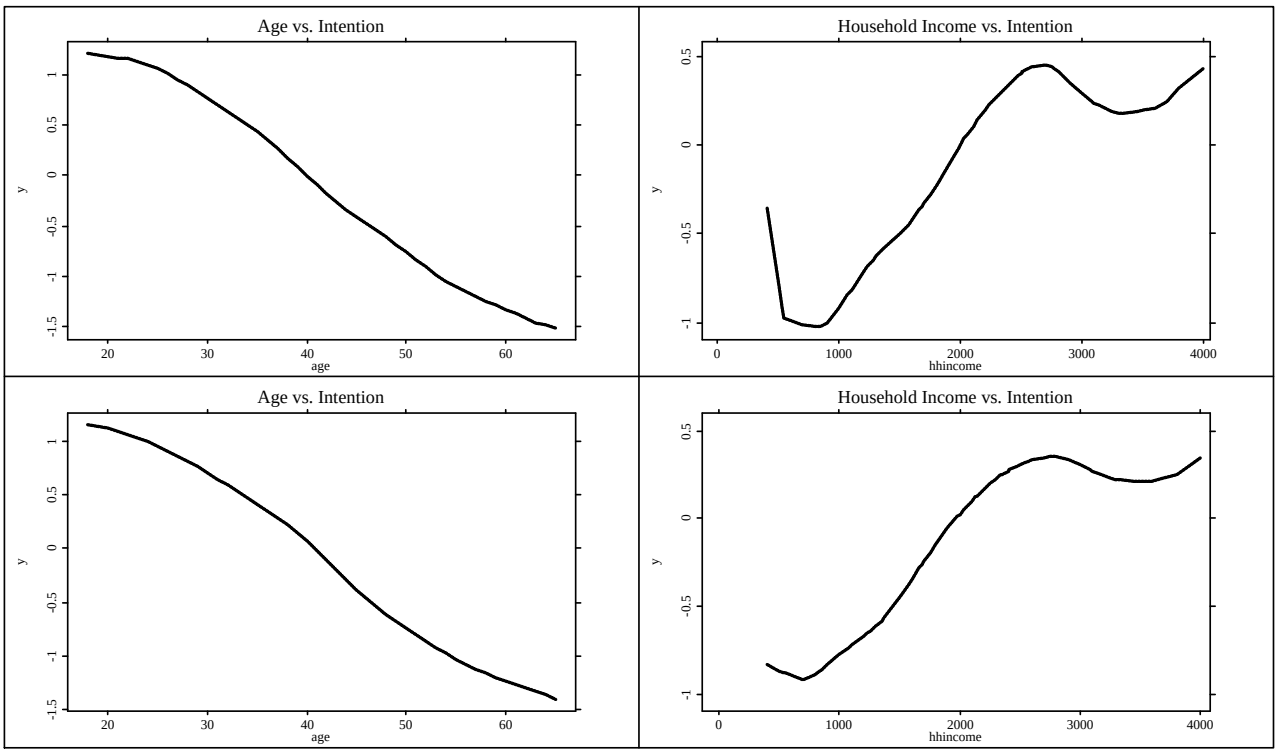
$$
\widehat{\alpha} = \frac{\frac{1}{n}\sum_{i=1}^{n} w_0(T_i)\left[\widehat{m}(T_i) - \widehat{m}_1(T_{i,1}) - \ldots - \widehat{m}_d(T_{i,d})\right]}{\frac{1}{n}\sum_{i=1}^{n} w_0(T_i)}
$$

Final additive estimate

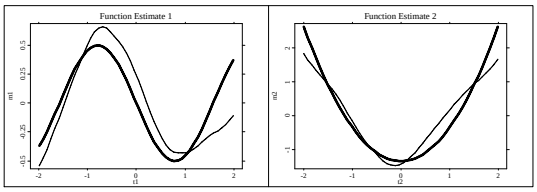
$$
\widehat{m}^+ = \widehat{\alpha} + \widehat{m}_1(t_1) + \cdots + \widehat{m}_d(t_d)
$$

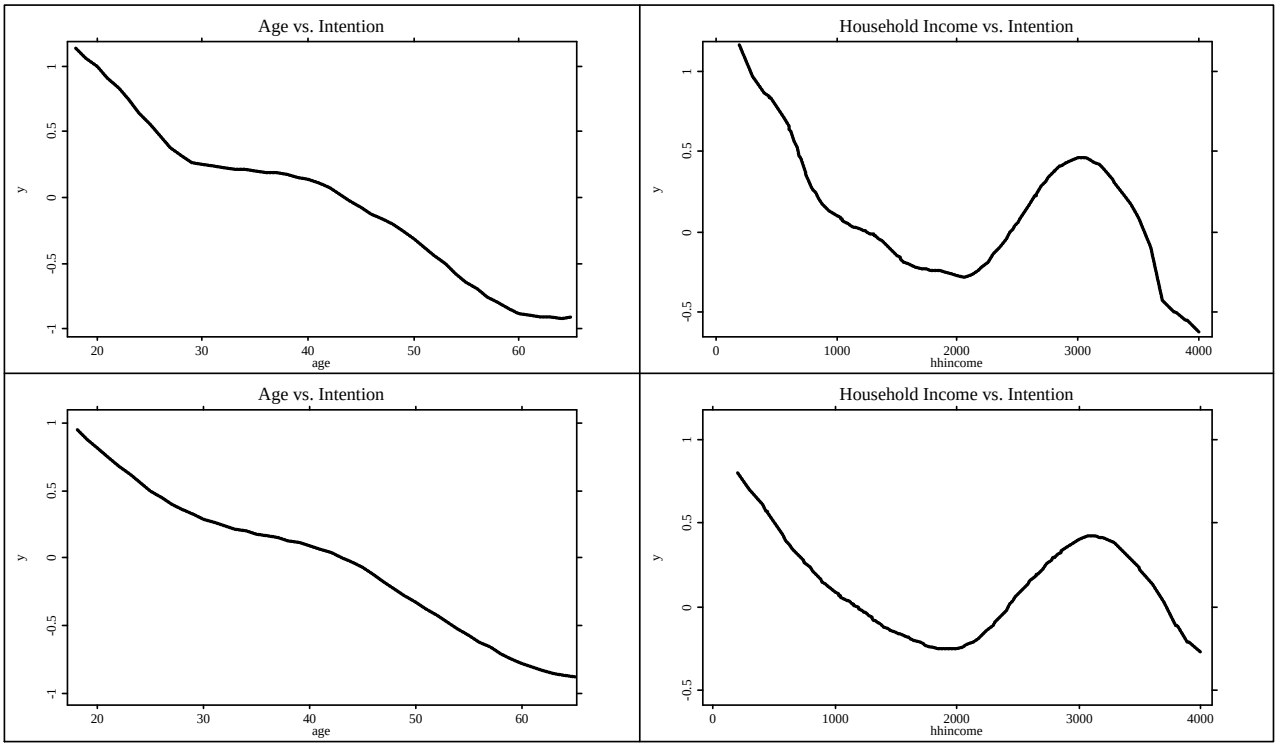



Function Estimate 1
m1
t1
Function Estimate 2
m2
t2

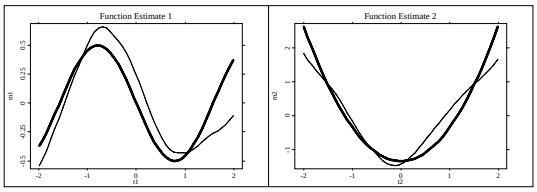


The semiparametric estimates for the influence of Age *(left) and* HH-Income *(right) in Mecklenburg-Vorpommern. The upper plots were estimated with* $h = 1.0$*, the lower plots with* $h = 1.25$*.*





The semiparametric estimates for the influence of Age *(left) and* HH-Income *(right) in Sachsen. The upper plots were estimated with* $h = 0.75$*, the lower plots with* $h = 1.0$*.*



3 Generalized additive models

$$m(x,t) = G\{\beta^T x + m(t)\}$$

Quasi-Likelihood function

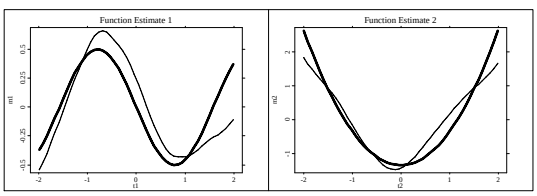
$$Q(\mu; y) = \int_{\mu}^{y} \frac{(s-y)}{V(s)}\, ds$$

where $\mu = G\{X^T\beta + m(T)\}$

Model A $(Y_1, X_1, T_1), \ldots, (Y_n, X_n, T_n)$ is an i.i.d. sample with $E[Y_i|X_i, T_i] = G\{X_i^T\beta + m(T_i)\}$.

Model B Model A and $Var[Y_i|X_i, T_i] = \sigma^2 V(\mu_i)$ where $\mu_i = G\{X_i^T\beta + m(T_i)\}$, σ^2 unknown scale parameter

Model C Model A and the cdtl distribution of Y_i from exp family $(\mu_i, V(\mu_i))$



Theorem 1

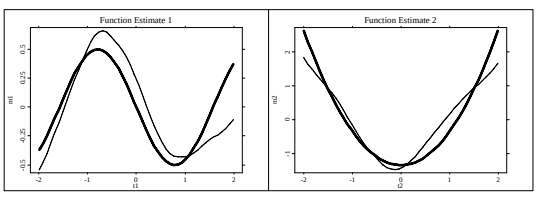
$$\sqrt{nh}\{\widehat{m}_1(t_1) - m_1(t_1) - \delta_n^1(t_1)\} \xrightarrow{D} N(0, \sigma_1^2(t_1))$$

$$\sigma_1^2(t_1) = \int K^2(u)du \frac{f_1(t_1)}{\{Ew_{-1}(T_{-1})\}^2} E\left[\frac{Z_1}{Z_2}|T_1 = t_1\right],$$

$f_{T_{-1}}, f_T$ densities of T_{-1} or $T = (T_1, T_{-1})$

$$\begin{aligned} Z_1 &= w_1^2(T_{-1}) \frac{Z^2}{V[G\{X^T\beta + m^+(T)\}]} \\ &\quad f_{T_{-1}}^2(T_{-1}) Var(Y|X,T), \\ Z_2 &= E\left[Z^2 \middle| T_1 = t_1, T_{-1}\right]^2 f_T^2(t_1, T_{-1}), \\ Z^2 &= \frac{G'(X^T\beta + m^+(T))^2}{V[G\{X^T\beta + m^+(T)\}]} \end{aligned}$$

$\delta_n^1(t_1)$ depends on other components: $O(g^2)$

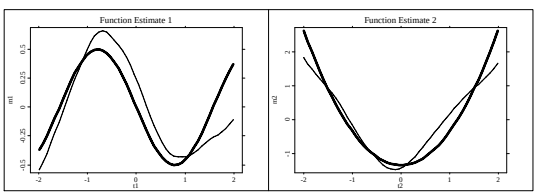


Multivariate situation

$$\sqrt{nh}\begin{pmatrix} \widehat{m}_1(t_1) - m_1(t_1) - \delta_n^1(t_1) \\ \vdots \\ \widehat{m}_d(t_d) - m_d(t_d) - \delta_n^d(t_d) \end{pmatrix}$$

converges to a centered Gaussian variable with covariance matrix

$$\begin{bmatrix} \sigma_1(t_1) & 0 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_d(t_d) \end{bmatrix}.$$

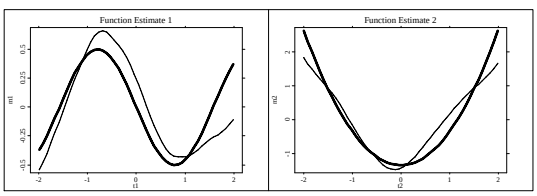


The variance of the estimate $\widehat{m}_1(t_1)$ can be estimated by

$$\hat{\sigma}_1^2(t_1) = nh \sum_{i=1}^{n} \hat{\tau}_i^2,$$

where

$$\begin{aligned}
\hat{\tau}_i \quad = \quad & \frac{1}{\frac{1}{n}\sum_{j=1}^{n} w_1(T_{j,-1})} \frac{1}{n} \sum_{j=1}^{n} w_1(T_{j,-1}) \kappa_j(t_1, T_{i,-1}) \\
& \left[\frac{1}{n} \sum_{l=1}^{n} \frac{G'(X_l^T \hat{\beta} + \widehat{m}^+(T_l))}{V[G\{X_l^T \hat{\beta} + \widehat{m}^+(T_l)\}]} \kappa_l(t_1, T_{j,-1}) \right] \\
& \frac{G'(X_i^T \hat{\beta} + \widehat{m}^+(t_1, T_{j,-1}))}{V[G\{X_i^T \hat{\beta} + \widehat{m}^+(t_1, T_{j,-1})\}]} \hat{s}_i,
\end{aligned}$$



$$\kappa_j(t) = \frac{K_h(t_1 - T_{i,1})L_g(t_{-1} - T_{i,-1})}{\frac{1}{n}\sum_{j=1}^{n} K_h(t_1 - T_{j,1})L_g(t_{-1} - T_{j,-1})}$$

$$\hat{s}_i^2 = \begin{cases} [Y_i - \hat{\mu}_i]^2 & \text{in case of Model A,} \\ \hat{s}^2 V(\hat{\mu}_i) & \text{in case of Model B,} \\ V(\hat{\mu}_i) & \text{in case of Model C} \end{cases}$$

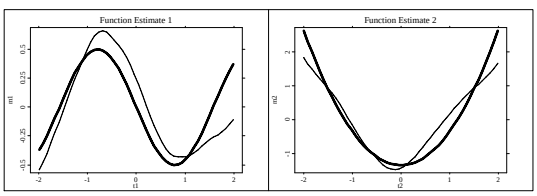
with

$$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{[Y_i - \hat{\mu}_i]^2}{V(\hat{\mu}_i)}$$

and

$$\hat{\mu}_i = G\{X_i^T\hat{\beta} + \hat{\alpha} + \widehat{m}_1(T_{i,1}) + \cdots + \widehat{m}_d(T_{i,d})\}.$$

Problem: bias of $\widehat{m}_1(t_1)$ depends on shape of the other components



4 Bootstrap bias correction

Wild Bootstrap model (A)

Step 1. Calculate residuals $\hat{\varepsilon}_i = Y_i - \hat{\mu}_i$.

Step 2. Generate n i.i.d. random variables $\varepsilon_1^*, \ldots, \varepsilon_n^*$ with mean 0, variance 1 and which fulfill for a constant C that $|\varepsilon_i^*| \leq C$ (a.s.) for $i = 1, \ldots, n$.

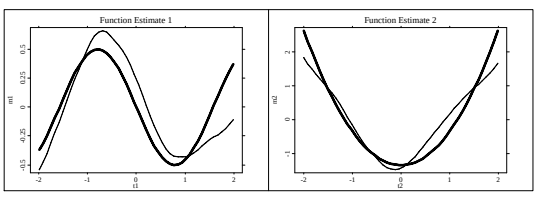
Step 3. Put $Y_i^* = \hat{\mu}_i + \hat{\varepsilon}_i \varepsilon_i^*$ for $i = 1, \ldots, n$.

Bootstrap model (B)

modify Step 3 above by putting
$Y_i^* = \hat{\mu}_i + \hat{\sigma} V\{\hat{\mu}_i\}^{1/2} \varepsilon_i^*$ for $i = 1, \ldots, n$.

Bootstrap model (C)

in exp model generate n independent $Y_1^*, \ldots, Y_n^*$ with distributions defined by $\hat{\mu}_i$



Bootstrap uses the data $(X_1, T_1, Y_1^*), \ldots,$ (X_n, T_n, Y_n^*) to calculate the estimate $\widehat{m}_1^*$.

Bandwidth h for the component t_1 and the same g for the other $d-1$ components.

Bootstrap estimate of the mean of $\widehat{m}_1(t_1)$ is $E^*\widehat{m}_1^*(t_1)$

Bias corrected estimate of $m_1(t_1)$ is defined by

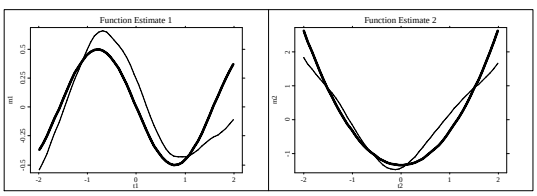
$$\widehat{m}_1^B(t_1) = \widehat{m}_1(t_1) - \widehat{\delta}_n^1(t_1),$$

where $\widehat{\delta}_n^1(t_1) = E^*\widehat{m}_1^*(t_1) - \widehat{m}_1(t_1)$.

Theorem 2

bias terms of order g^2 are removed by this construction.

$$\widehat{m}_1^B(t_1) - m_1(t_1) = O_p\{h^4 + g^4 + (nh)^{-1/2}\}.$$



Theorem 3

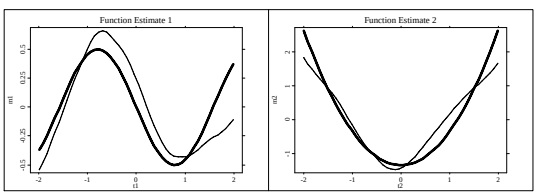
If $hg^{d-1}n^{1/2}(\log n)^{-1}$ tends to infinity and h and $g = o(n^{-1/8})$,

$$n^{1/2}\{\widehat{\beta} - \beta\} \xrightarrow{D} N(0, I^{-1})$$

where

$$\begin{aligned} I &= EZ^2\widetilde{X}\widetilde{X}^T \quad \text{with} \\ \widetilde{X} &= X - \{E(Z^2|T)\}^{-1}E(Z^2X|T) \end{aligned}$$

(Z^2 as defined before)



5 Bootstrap tests for linearity

$$H_0 : \quad m_1(t_1) = \gamma_1 t_1$$

Hastie and Tibshirani (1990)

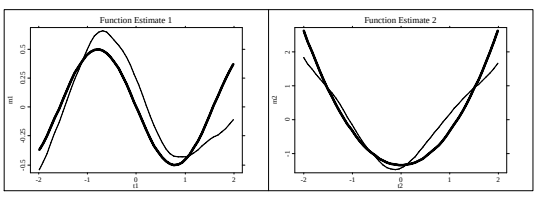
Likelihood ratio test, χ^2 approximations

On the hypothesis the estimate $\widehat{m}_1(t_1)$ may have a non-negligible bias. Compare $\widehat{m}_1(t_1)$ with a bootstrap estimate of its expectation under the hypothesis.

Calculate semiparametric estimates ($\tilde{\beta}$, etc.)in the hypothesis model

$$m(X,T) = G\{X^T\beta + \alpha + \gamma_1 T_1 + m_2(T_2) + \cdots + m_d(T_d)\}.$$

[The α occurring her is different from the α defined before, because X_i is now replaced by $(X_i, T_{i,1})$.]



Put

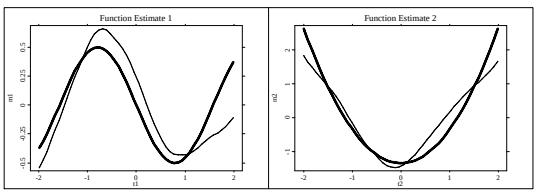
$$\tilde{\mu}_i = G\{X_i^T \tilde{\beta} + \tilde{\alpha} + \tilde{\gamma}_1 T_{i,1} + \widetilde{m}_2(T_{2,i}) + \cdots + \widetilde{m}_d(T_{i,d})\}.$$

Generate independent samples $(Y_1^*, \ldots, Y_n^*)$ with μ_i replaced by $\tilde{\mu}_i$. Calculate the estimate $\widehat{m}_1^*$.

Bootstrap estimate of the mean of $\widehat{m}_1(t_1)$ is given by $E^*\widehat{m}_1^*(t_1)$

Define test statistic:

$$\begin{aligned} R &= \sum_{i=1}^{n} w(T_i) \frac{[G'\{X_i^T \widehat{\beta} + \widehat{m}^+(T_i)\}]^2}{V(G\{X_i^T \widehat{\beta} + \widehat{m}^+(T_i)\})} \\ &\quad \{\widehat{m}_1(T_{i,1}) - E^*\widehat{m}_1^*(T_{i,1})\}^2 \end{aligned}$$



test statistic:

$$R = \sum_{i=1}^{n} w(T_i) \frac{[G'\{X_i^T \widehat{\beta} + \widehat{m}^+(T_i)\}]^2}{V(G\{X_i^T \widehat{\beta} + \widehat{m}^+(T_i)\})} \{\widehat{m}_1(T_{i,1}) - E^* \widehat{m}_1^*(T_{i,1})\}^2$$

Theorem 4

$$v_n^{-1}(R - e_n) \xrightarrow{D} N(0,1)$$

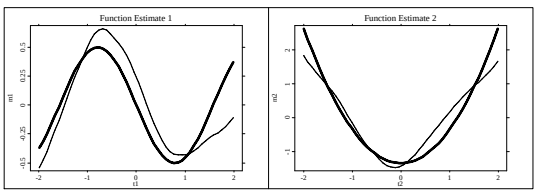
with

$$e_n = h^{-1} \int K(u)^2 \; du E[A f_{T_1}(T_1)],$$

$$v_n^2 = h^{-1} \int K^{(2)}(u)^2 \; du E\left\{E[A|T_1]^2 f_{T_1}(T_1)^3\right\},$$

$$A = \frac{1}{E[w_{-1}(T_{-1})]} \frac{w_{-1}(T_{-1}) w(T) Z^4 f_{T_{-1}}^2(T_{-1})}{E[Z^2|T]^2 f_T^2(T)} \frac{Var[Y|X,T]}{V\{X^T\beta + m^+(T)\}},$$

where $K^{(2)}(u) = \int K(u-v)K(v) \; dv$ is the convolution of K with itself.



Bootstrap test statistic

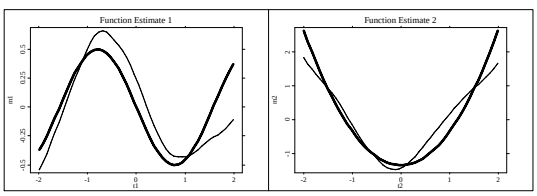
$$R^* = \sum_{i=1}^{n} w(T_i) \frac{[G'\{X_i^T \widehat{\beta} + \widehat{m}^+(T_i)\}]^2}{V\{X_i^T \widehat{\beta} + \widehat{m}^+(T_i)\}} \{\widehat{m}_1^*(T_{i,1}) - E^* \widehat{m}_1^*(T_{i,1})\}^2 .$$

Theorem 5 It holds that

$$d_K\{\mathcal{L}^*(R^*), \mathcal{L}(R)\} \xrightarrow{P} 0$$

where d_K denotes the Kolmogorov distance, which is defined for two probability measures μ and ν (on the real line) as

$$d_K(\mu, \nu) = \sup_{t \in I\!R} \Big| \mu(X \leq t) - \nu(X \leq t) \Big|.$$



6 Uniform bootstrap confidence bands

Define

$$S = \sup_{t_1} w_1(t_1)|\widehat{m}_1(t_1) - m_1(t_1) - \delta_n^1(t_1)|\hat{\sigma}_1^{-1}(t_1),$$

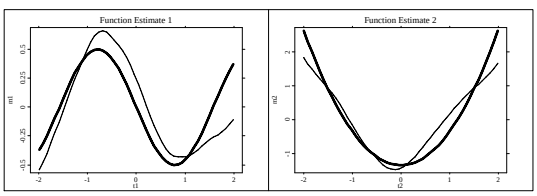
$$S^* = \sup_{t} w_1(t_1)|\widehat{m}_1^*(t_1) - E^*\ \widehat{m}_1^*(t_1)|\hat{\sigma}_1^{-1}(t_1)$$

$$S^{**} = \sup_{t} w_1(t_1)|\widehat{m}_1^*(t_1) - E^*\ \widehat{m}_1^*(t_1)|[\hat{\sigma}_1^*]^{-1}(t_1)$$

Theorem 6 It holds that

$$d_K\{\mathcal{L}^*(S^*), \mathcal{L}(S)\} \quad \xrightarrow{P} \quad 0,$$

$$d_K\{\mathcal{L}^*(S^{**}), \mathcal{L}(S)\} \quad \xrightarrow{P} \quad 0.$$



7 Simulation

Binary choice model

G = Logit

$X_1, X_2 \ N(0,1)$

$T_1, T_2 \ U(-2,2)$

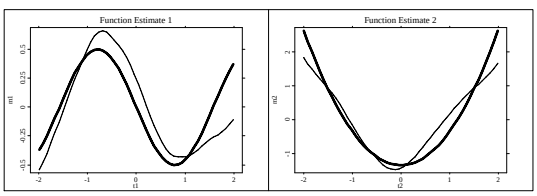
$n = 250, B = 249, \beta = (0.3, -0.7)^T, \alpha = 0$

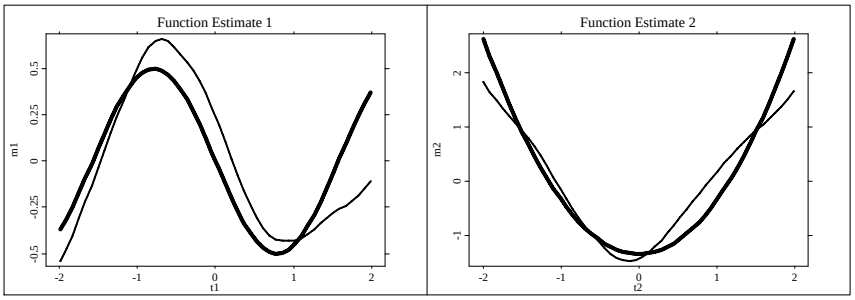
$m_1(t_1) = 2 * sin(-2 * t_1), m_2(t_2) = t_2^2 - E[T_2^2]$

All computations done in

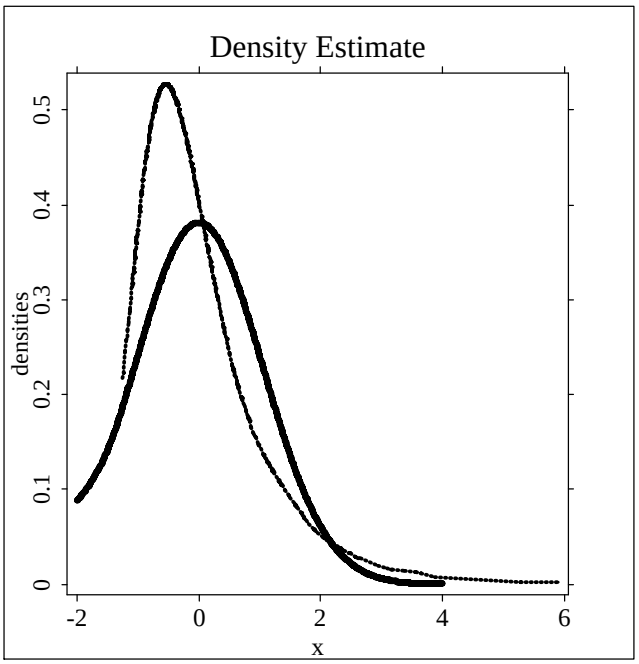
XploRe - the statistical computing environment

www.xplore-stat.de

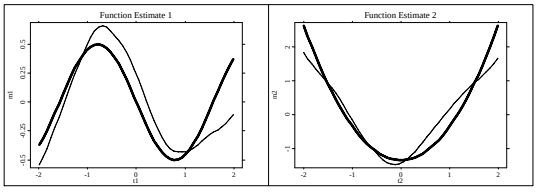




Plots of the nonparametric components.



Standardized density estimate of the test statistic (thin line) and standard normal density (thick line).

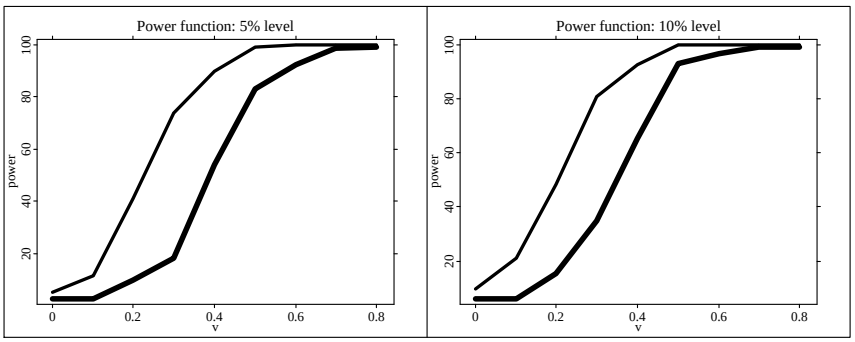


Power calculations

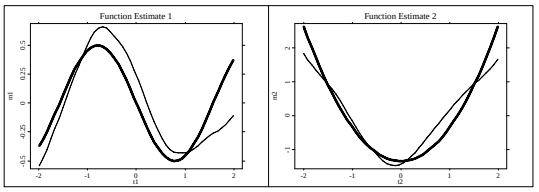
$$m_1(t_1) = (1-v)t_1 + v\{2\sin(-2t_1)\}, 0 < v < 1$$

Compare with LRT for

$$G[\beta x + \gamma_1 t_1 + \gamma_2\{2\sin(-2t_1)\} + \gamma_3 m_2(t_2) + \gamma_4].$$



Power functions for theoretical levels 0.05 and 0.1 , for the nonparametric bootstrap test (thick line) and the likelihood ratio test (thin line).

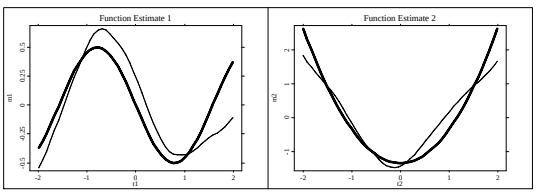


8 East West Migration

Mecklenburg-Vorpommern

n	402			
	min.	max.	mean	stdev.
response y	0	1	0.390	0.488
fam West x_1	0	1	0.888	0.315
unempl x_2	0	1	0.211	0.408
city size x_3	0	1	0.358	0.480
female x_4	0	1	0.502	0.500
age t_1	18	65	39.93	12.89
hhincome t_2	400	4000	2262.2	769.8

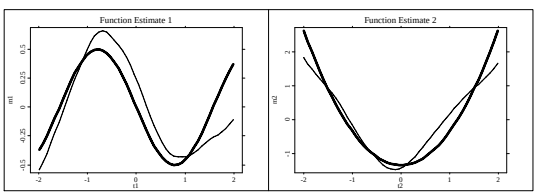
Table 1: *Descriptive statistic for our data of Meck-Pomm.*



Sachsen

n	955			
	min.	max.	mean	stdev.
response y	0	1	0.395	0.489
fam West x_1	0	1	0.824	0.380
unempl x_2	0	1	0.183	0.387
city size x_3	0	1	0.259	0.438
female x_4	0	1	0.516	0.499
age t_1	18	65	40.36	12.69
hhinc t_2	200	4000	2136.31	738.71

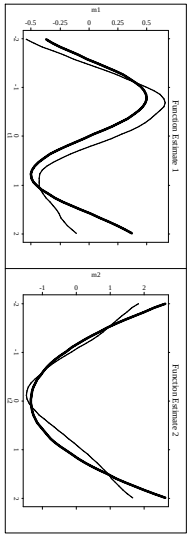
Table 2: *Descriptive statistic for our data of Sachsen.*



Parametric estimation results

	Meck-Pomm			Sachsen		
	Coeff.	stdev.	$P > \|z\|$	Coeff.	stdev.	$P > \|z\|$
fam West	0.5893	0.3820	0.124	0.7604	0.1972	<0.001
unemployed	0.7799	0.2779	0.005	0.1354	0.1783	0.447
city size	0.8216	0.2421	0.001	0.2596	0.1556	0.085
female	-0.3884	0.2315	0.093	-0.1868	0.1382	0.178
age	-0.9227	0.1330	<0.001	-0.5051	0.0728	<0.001
hhinc	0.2318	0.1221	0.057	0.0936	0.0707	0.187
constant	-1.3673	0.2969	0.001	-1.0924	0.2003	<0.001

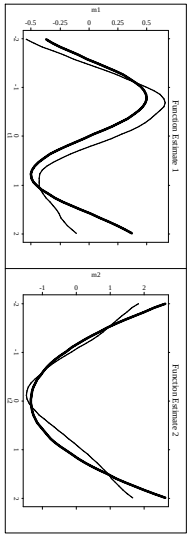
Table 3 : *Results of a generalized linear regression.*

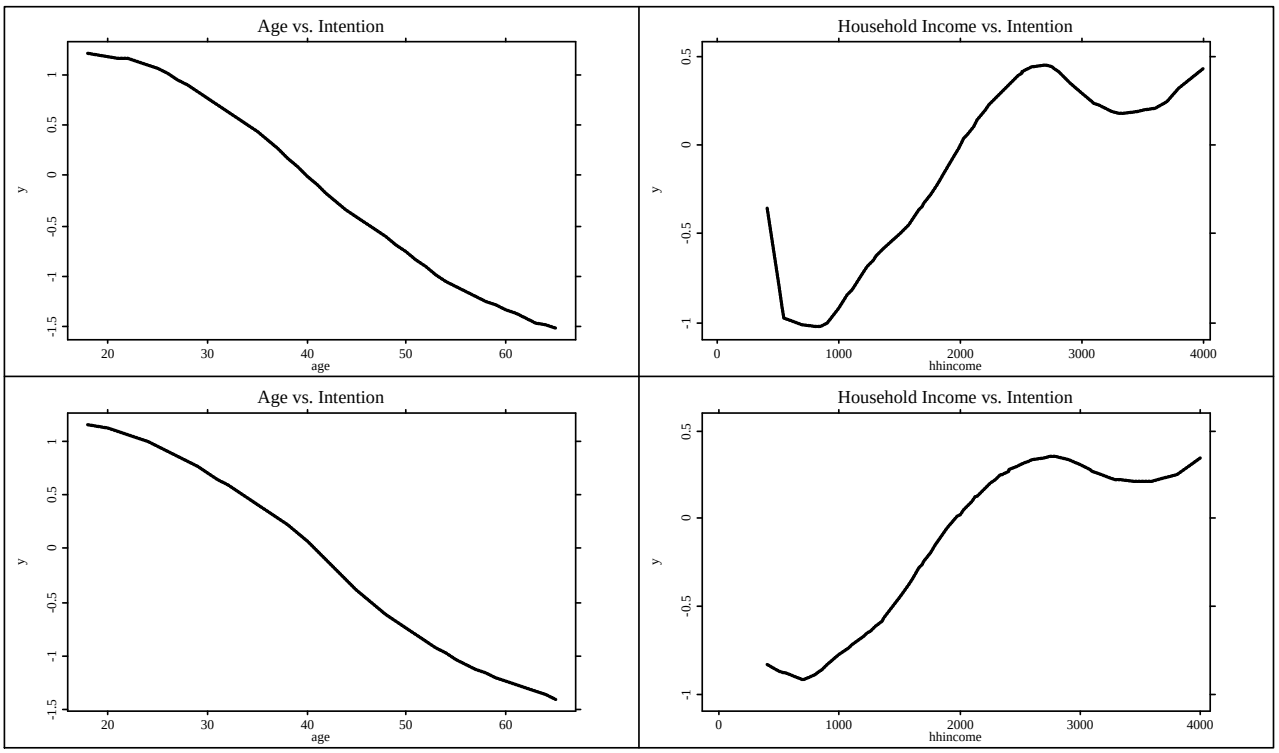


Coefficients of the linear part

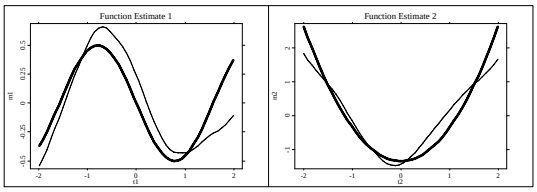
	Meck-Pomm		Sachsen	
	semi. a	semi. b	semi. a	semi. b
family/friends West	0.5920	0.5809	0.7137	0.7289
unemployed/...	0.7771	0.7992	0.1469	0.1308
middle sized city	0.7156	0.7127	0.3134	0.2774
female	-0.3309	-0.3485	-0.1898	-0.1871
constant	-1.4616	-1.4111	-1.1045	-1.1007

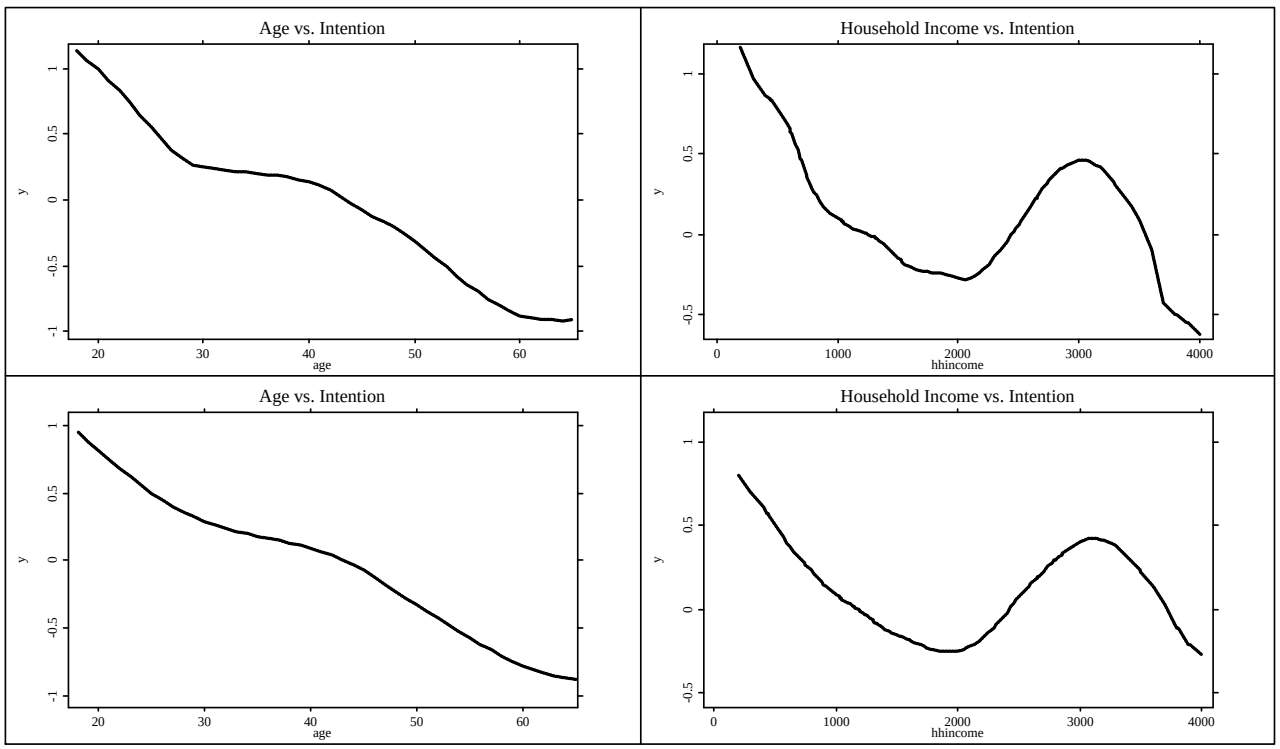
Table 4: *Results for the pure parametric estimation (par.) and for the parametric part of the generalized additive partially linear regression model: semi. a (with bandwidth $h = 1.0$), semi. b ($h = 1.25$) for Mecklenburg-Vorpommern; semi. a ($h = 0.75$) and semi. b ($h = 1.0$) for Sachsen.*





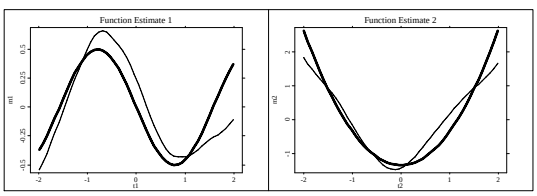
The semiparametric estimates for the influence of Age *(left) and* HH-Income *(right) in Mecklenburg-Vorpommern. The upper plots were estimated with* $h = 1.0$*, the lower plots with* $h = 1.25$*.*





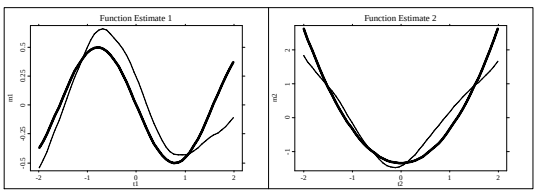
The semiparametric estimates for the influence of Age *(left) and* HH-Income *(right) in Sachsen. The upper plots were estimated with* $h = 0.75$*, the lower plots with* $h = 1.0$*.*

Result of bootstrap test: linearity of age and hhinc rejected.



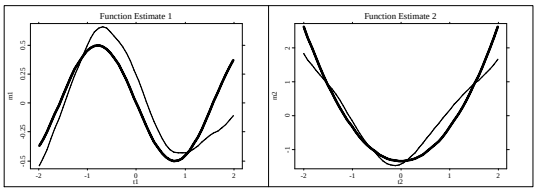
9 Apprenticeship

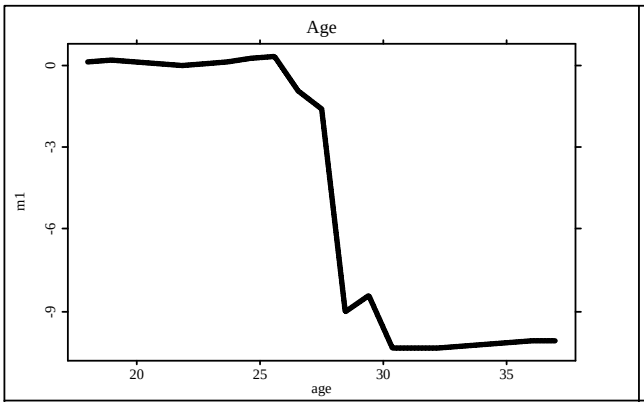
Consider probability of an apprentice to be unemployed directly after finishing his apprenticeship. This prob depends on schooling, etc.

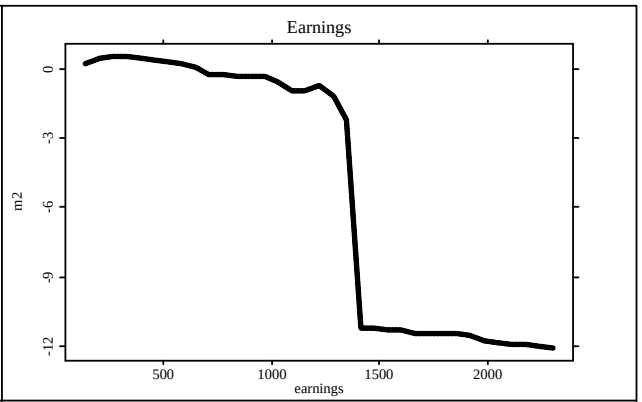


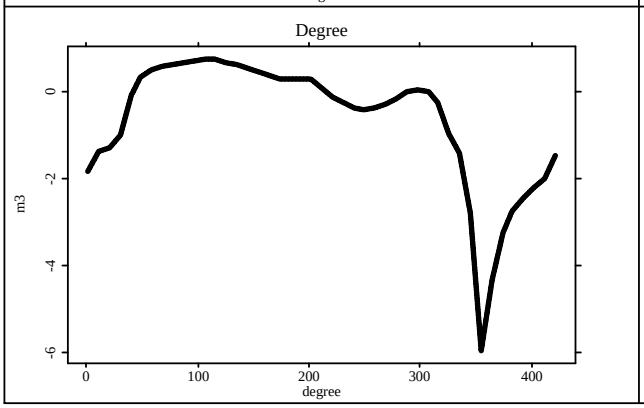
Variable	Definition/Comments
Sex	Sex of the respondent. It takes the value "1" if the respondent is female, "0" if male.
Age	Age of the respondent in the year the apprenticeship was completed.
Schooling	Years of schooling (9 - 13).
Earnings	Gross monthly earnings as an apprentice.
Big City	"1" if the city where the respondent lived at end of his apprenticeship has between 250 000 and 500 000 inhabitants.
Huge City	"1" if the city has more than 500 000 inhabitants.
Degree	Percentage of people apprenticed in a certain occupation, divided by the percentage of people employed in this occupation in the entire economy.
U-Rate	Unemployment rate in the state the respondent lived in during the year the apprenticeship was completed.

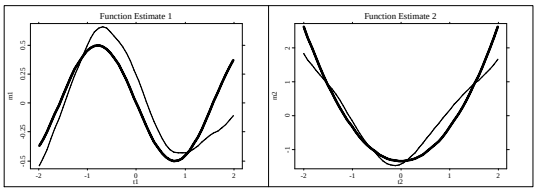
Table 5: *Explanatory variables.*










```
Linear Effects:
----------------------
Sex (female=1) -0.4742
Schooling       0.0616
Big City        1.0341
Huge City      -0.2497
U-Rate          0.2616
Constant       -4.3536
```

Estimates of the additive components in the reduced model and the coefficients for the linear part.