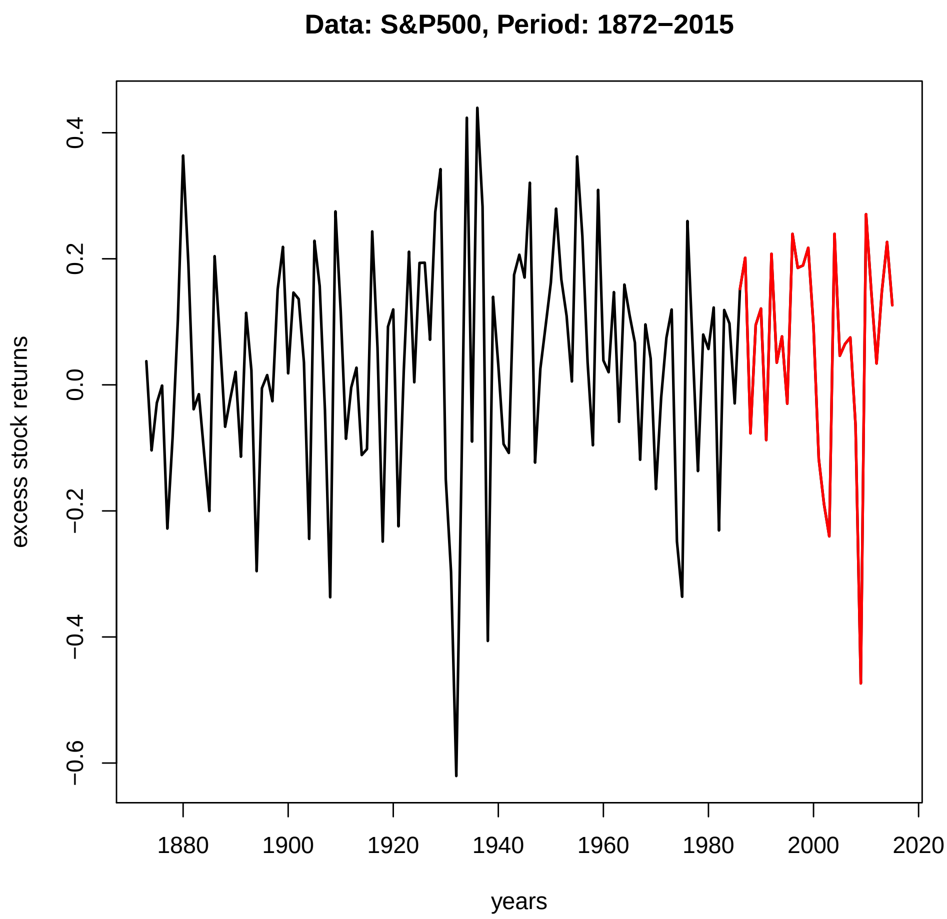
Dynamic long-term return models to be used for pension products

* based on joint work with
Stefan Sperlich, Jens Perch Nielsen, and Enno Mammen

Michael Scholz
Faculty of Statistics
TU Dortmund University

FASI – Seminar
London, May 2017


Data: S&P500, Period: 1872–2015
excess stock returns
0.4
0.2
0.0
−0.2
−0.4
−0.6
1880
1900
1920
1940
1960
1980
2000
2020
years

Objectives of the talk:

- Are **equity returns or premiums** predictable? Until mid-1980's: Predictability would contradict the **efficient markets paradigm**.
- Empirical research in the late 20th century and recent progress in asset pricing theory suggest that excess returns are **predictable**.
- We take the **long-term actuarial view** and base our predictions on **annual data** of the S&P500 from 1872 through 2015 on a one year horizon.
- **Our interests**:
  - Actuarial models of **long-term saving** and potential **econometric improvements** to such models.
  - **Market timing/compare assets** based on the Sharpe ratio (SR)

$$SR_t = \frac{\mathbb{E}(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}{\sqrt{Var(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}}$$

- Not many historical years in our records and **data sparsity** is an important issue.
- **Bias** might be of **great importance** when predicting yearly data. Classical trade-off of variance and bias depends on the horizon/frequency.
- **Advocate** for non- and semi-parametric methods in financial applications:
  - **Powerful data-analytic tools**: Local-linear kernel smoothing and wild bootstrap.
  - With **suitable modifications** those techniques can perform well in different economic fields.
  - Include **prior knowledge** in the statistical modelling process for **bias reduction** and to **avoid the curse of dimensionality** and other problems.

Prediction
Sharpe Ratio
R_V^2
Combined Estimator
Estimation Mean Fct.
Estimation Var Fct.
Simulation Study
Empirical Study
Outlook/ Summary

Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

- The **excess stock returns**:

$$S_t = \log\left(\frac{P_t + D_t}{P_{t-1}}\right) - r_{t-1}$$

with **dividends** D_t paid during period t, **stock price** P_t at the end of period t, and **short-term interest rate** $r_t = \log(1 + R_t/100)$ with **discount rate** R_t

- Consider **one-year-ahead predictions** ($T = 1$), but predictions over the **next *T*** periods are also easily included:

$$Y_t = S_t + \ldots + S_{t+T-1} = \sum_{i=0}^{T-1} S_{t+i}$$

but this would pose **greater statistical challenges**.

- One traditional equation for the **value of a stock** is

$$P_t = \sum_{j=1}^{\infty} (1+\gamma)^{-j}(1+g)^{j-1} D_t$$

with γ discount rate and g growth of dividend yields.

- Price of stocks depends on quantities such as **dividend yield**, **interest rate**, **inflation** (last two highly correlated with almost any relevant discount rate)

- Covariates $\boldsymbol{X}_t$ with predictive power: **dividend-price ratio**, **earnings-price ratio**, **interest rates**, ...

- Consider the model

$$Y_t = m(\boldsymbol{X}_{t-1}) + v(\boldsymbol{X}_{t-1})^{1/2}\varepsilon_t$$

where $\mathbb{E}(\varepsilon_t|\boldsymbol{X}_{t-1}) = 0$ and $Var(\varepsilon_t|\boldsymbol{X}_{t-1}) = 1$.

- A way to examine the performance of an investment by adjusting for its risk (**reward-to-variability ratio**)
- It measures the **excess return** (or **risk premium**) per **unit of deviation** in an investment asset or a trading strategy. Sharpe (JPM 1994):

$$SR_t = \frac{E(Y_t)}{\sqrt{Var(Y_t)}}$$

- Use in finance:
  - SR characterizes how well the **return of an asset** compensates the investor for the **risk taken**.
  - When comparing two assets vs. a common benchmark, the one with a **higher SR** provides **better return for the same risk** (the same return for a lower risk)

- In practice: **ex-post** SR used with **realized** rather than **expected returns**

$$SR_{t,simple} = \frac{mean(Y_t)}{sd(Y_t)}$$

- **Easy** to calculate but **depends on length** of observations, includes both **systematic and idiosyncratic risk**
- **Non-normality** of assets, effect of **covariates**, **predictions**?
- In our setting:

$$SR_t(\boldsymbol{x}_{t-1}) = \frac{\mathbb{E}(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}{\sqrt{Var(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}} = \frac{m(\boldsymbol{x}_{t-1})}{\sqrt{\nu(\boldsymbol{x}_{t-1})}}$$

and we get a **two-step** estimator for the **Sharpe ratio** as

$$\widehat{SR}_t = \frac{\hat{m}_t}{\sqrt{\hat{\nu}_t}}$$

Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

In detail:

- Consider $Y_t = \mu + \xi_t$ and $Y_t = g(X_t) + \zeta_t$
- μ estimated by the mean $\bar{Y}$ and g by local-linear kernel regression
- Nielsen and Sperlich (Astin Bull. 2003) define the **validated** R^2_V as

$$R^2_V = 1 - \frac{\sum_t \{Y_t - \hat{g}_{-t}\}^2}{\sum_t \{Y_t - \bar{Y}_{-t}\}^2},$$

  where the function g and the simple mean $\bar{Y}$ are predicted at point t without the information contained in t
- A **cross-validation criterion** to rank different models used for both **choice of bandwidth** and **model selection**.
- Cross-validation optimal also for **β-recurrent Markov processes** (Bandi et. al (2016))

Properties:

- Replacement of **total variation** and **not explained variation** in usual R^2 by its **cross validated** analogs.
- $R^2_V \in (-\infty, 1]$
- Measures how well a given model and estimation principle predicts compared to another (here: to the CV mean)
- If $R^2_V < 0$ we cannot predict better than the mean.
- CV punishes **overfitting**, i. e. pretending a functional relationship that is not really there (leads to $R^2_V < 0$)
- It is a (non-classical) **out-of sample** measure

Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

- Basic idea: **Combined estimator** Glad (ScanJStat 1998)

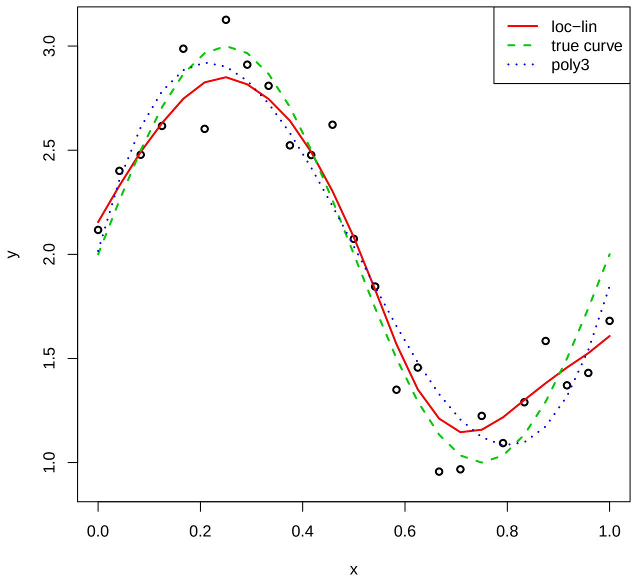
  Nonparametric estimator **multiplicatively** guided by, for example, parametric model

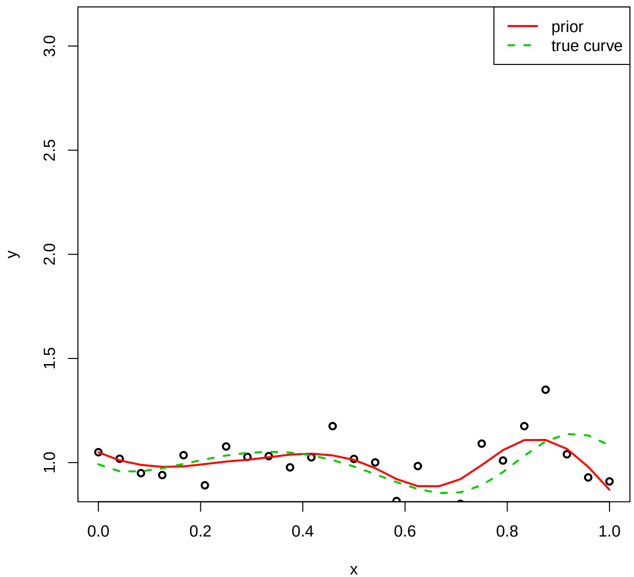
$$g(x) = g_\theta(x) \cdot \frac{g(x)}{g_\theta(x)}$$

- **Essential fact**:
  - Prior captures characteristics of **shape** of $g(x)$
  - **Correction factor** $g(x)/g_\theta(x)$ is **less variable** than original function $g(x)$
  - Nonparametric estimator gives **better results** with **less bias**
- Include **prior information** in analysis coming from
  - Good economic model
  - (Simple) empirical data analysis or statistical modelling (**shape**, **constructed variables**, **different frequencies**)

- **Local Problem**: Prior crosses x-axis
  - More robust estimates with suitable **truncation**:
    Clipping absolute value below $\frac{1}{10}$ and above 10
  - **Shift** by a distance c so that new prior strictly greater than zero and does not intersect the x-axis
- **Dimension reduction**: Use possible **overlapping** covariates $\boldsymbol{x}_1, \boldsymbol{x}_2$ for prior and correction factor

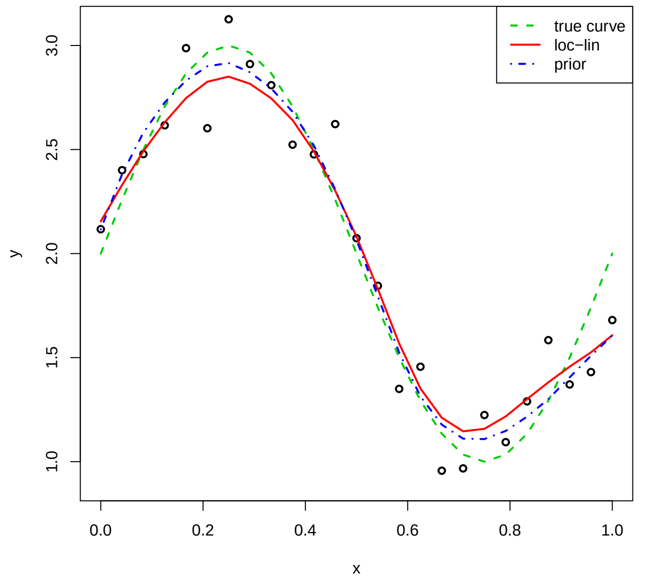
$$g(\boldsymbol{x}_1) = (g_\theta(\boldsymbol{x}_2) + c) \cdot \frac{g(\boldsymbol{x}_1)}{g_\theta(\boldsymbol{x}_2) + c}$$


loc−lin
true curve
poly3
y
x
0.0
0.2
0.4
0.6
0.8
1.0
1.0
1.5
2.0
2.5
3.0


prior
true curve
y
x
0.0
0.2
0.4
0.6
0.8
1.0
1.0
1.5
2.0
2.5
3.0

For $n = 25$ and $n.it = 300$ we get

$$mse_{prior} = 0.0879 + 0.2025, \; mse_{loc-lin} = 0.1001 + 0.2212, \; ratio = 0.9040$$



- Scholz et al. (IME 2015) applied the **combined estimator** to annual American data.
- A bootstrap test on the true functional form of the conditional expected returns **confirms predictability**.
- Including prior knowledge shows **notable improvements** in the prediction of excess stock returns compared to linear and fully nonparametric models.
- We will use their **best model** as starting point in our empirical analysis:
  - Prior: linear model with **risk-free** rate as predictive variable
  - Correction factor: fully nonparametric with **earnings by price** and **long term interest**
  - Gives $R_V^2 = 20.9$ (compared to a $R_V^2 = 14.5$ for a fully nonparametric model without prior)

- Four main approaches proposed in the literature: **direct method**, **residual based**, **likelihood-based**, and **difference-sequence method**
- **Direct method** is based on

$$Var(Y_t|\boldsymbol{X}_t = \boldsymbol{x}_t) = \mathbb{E}(Y_t^2|\boldsymbol{X}_t = \boldsymbol{x}_t) - \mathbb{E}(Y_t|\boldsymbol{X}_t = \boldsymbol{x}_t)^2$$

  where both parts are **separately** estimated. Result is **not nonnegative** and **not fully adaptive** to the mean fct. Härdle and Tsybakov (JoE 1997) with $\boldsymbol{X}_t = Y_{t-1}$.
- **Residual based** methods consist of two stages:
  1. estimate $\hat{m}$ and squared residuals $\hat{u}_t^2 = (Y_t - \hat{m}(\boldsymbol{X}_t))^2$
  2. estimate $\hat{v}$ from $\hat{u}_t^2 = v(\boldsymbol{X}_t) + \varepsilon_t$

- Different variants of residual based method (mostly) for 2nd step:
  - Fan and Yao (Biometrika 1998) apply **loc-lin** in both stages. Result is **not nonnegative** but **asymp. fully adaptive** to the unknown mean fct.
  - Ziegelmann (ET 2002) proposes the **local exponential estimator** to ensure **nonnegativity**

$$\sum_t \left(\hat{u}_t^2 - \Psi\left\{\alpha + \beta(X_t - x)\right\}\right)^2 K_h(X_t - x) \quad \Rightarrow \quad Min_{\alpha,\beta}$$

  - Mishra, Su, and Ullah (JBES 2010) propose the use of the **combined estimator** with a parametric guide. They ignore **bias** reduction in 1st step.
  - Xu and Phillips (JBES 2011) use a **re-weighted local constant** estimator (maximize the empirical likelihood s.t. a bias-reducing moment restriction).

Estimation of the conditional variance function III

- Yu and Jones (JASA 2004) use estimators based on a **localized normal likelihood** (standard **loc-lin** for estimating the mean m and **loc log-lin** for variance ν)

$$-\sum_t \left\{ \frac{(Y_t - m(X_t))^2}{\nu(X_t)} + \log(\nu(X_t)) \right\} K_h(X_t - x)$$

- Wang et al. (AnnStat 2008)
  - analyze the **effect of the mean on variance fct. estimation**
  - compare the performance of the residual-based estimators to a **first-order-difference-based** (**FOD**) **estimator**: loc-lin on

$$D_t^2 = \frac{(Y_t - Y_{t+1})^2}{2}$$

- Residual-based estimators use an **optimal** estimator for mean fct. in

$$\hat{v}(x) = \sum_t w_t(x)(Y_t - \hat{m}(x_t))^2,$$

works well if in $\hat{m}$ the **bias is negligible**.

- Bias **cannot be further reduced** in 2nd stage.
- FOD: crude estimator $\hat{m}(x_t) = Y_{t+1}$

Our strategy:

- Use **combined estimator with simple linear prior** in both stages
- Reasons:
  - FOD was not convincingly performing in **small samples**
  - We know that the mean fct. is rather **smooth**
  - But **bias reduction** is key due to **sparsity**
  - We cannot compare FOD and residual-based results in terms of R_V^2
  - Maybe FOD together with combined estimator?

Overview:

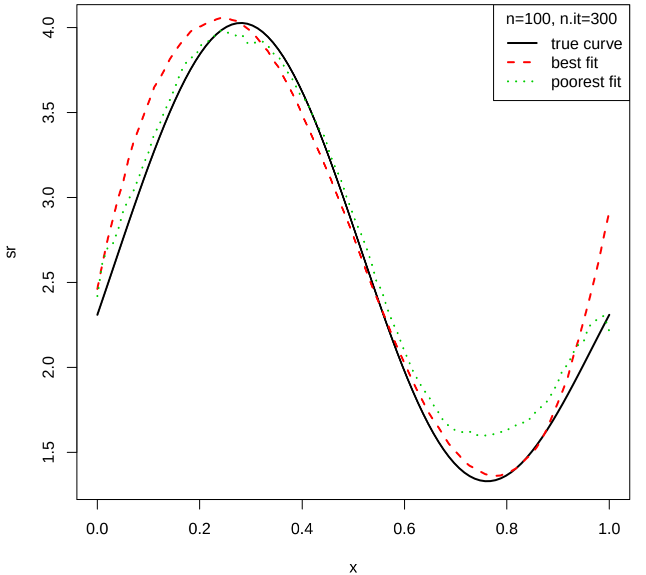
- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

- Which procedure gives reasonable results for estimation of Sharpe ratio?
- Simulate mean and variance function on $x \sim U[0,1]$ as

$$m(x) = 2 + \sin(a\pi x), \quad v(x) = x^2 - x + 0.75 \quad \text{and} \quad sr(x) = m(x)/v(x)^{1/2}$$

- $n = 100$
  - **Best model** in terms of **cross-validated mean square error** (0.180) uses **combined estimator** for mean with a poly3 prior and **combined estimator** for variance with a linear prior
  - **Poorest model** in terms of cross-validated mean square error (0.690) uses **combined estimator** for mean with a linear prior and **FOD estimator** for variance with a poly5 prior

- Averages of **best** and **poorest** model over 300 independent estimates ($n = 100$)

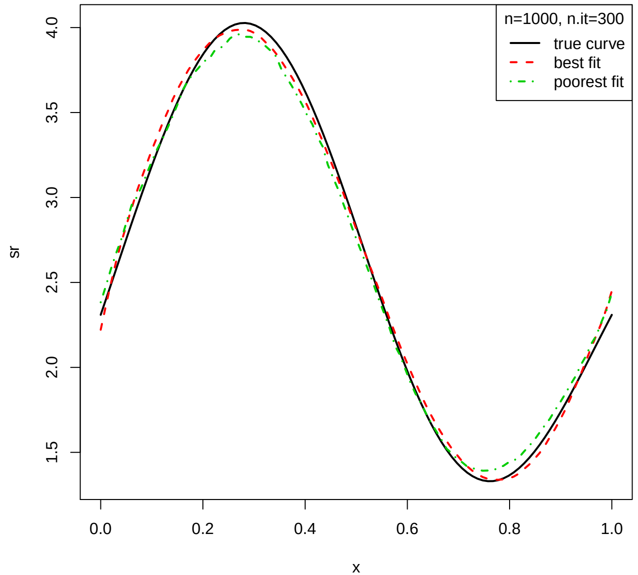


- Which procedure gives reasonable results for estimation of Sharpe ratio?
- Simulate mean and variance function on $x \sim U[0, 1]$ as

$$m(x) = 2 + \sin(a\pi x), \quad \nu(x) = x^2 - x + 0.75 \quad \text{and} \quad sr(x) = m(x)/\nu(x)^{1/2}$$

- $n = 1000$
  - **Best model** in terms of **cross-validated mean square error** (0.017) uses **combined estimator** for mean with a poly3 prior and **combined estimator** for variance with an exponential prior
  - **Poorest model** in terms of cross-validated mean square error (0.054) uses **combined estimator** for mean with a exponential prior and **FOD estimator** for variance without prior

- Averages of **best** and **poorest** model over 300 independent estimates ($n = 1000$)



- **Annual American Data**: Updated and revised version of Robert Shiller's dataset - Chapter 26 in Market volatility (1989)

Tabelle: US market data (1872-2015).

	Max	Min	Mean	Sd
Excess Stock Returns	0.44	-0.62	0.04	0.18
Dividend by Price	0.09	0.01	0.04	0.02
Earnings by Price	0.17	0.02	0.08	0.03
Short-term Interest Rate	17.63	0.19	4.61	2.84
Long-term Interest Rate	14.59	1.91	4.60	2.26
Inflation	0.17	-0.19	0.02	0.06
Spread	3.27	-5.06	-0.01	1.58

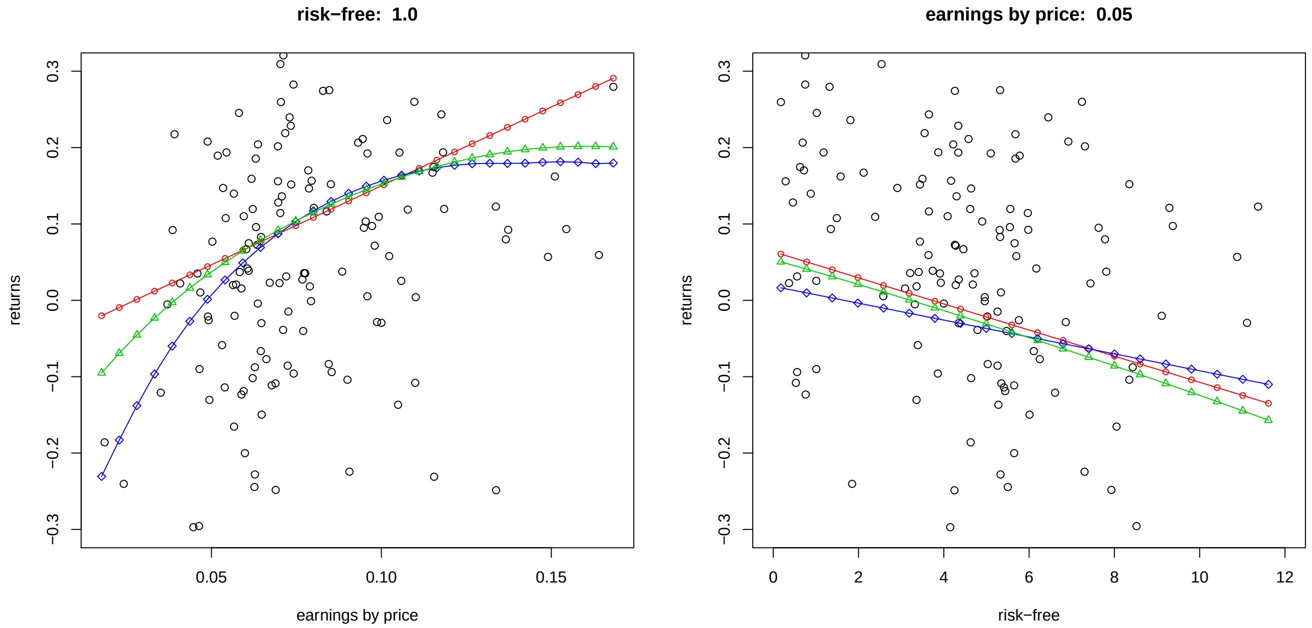
- **estimation of the mean fct.**:

$$m(\boldsymbol{x}_1) = (m_\theta(\boldsymbol{x}_2) + c) \cdot \frac{m(\boldsymbol{x}_1)}{m_\theta(\boldsymbol{x}_2) + c}$$

Tabelle: Predictive power (in percent)

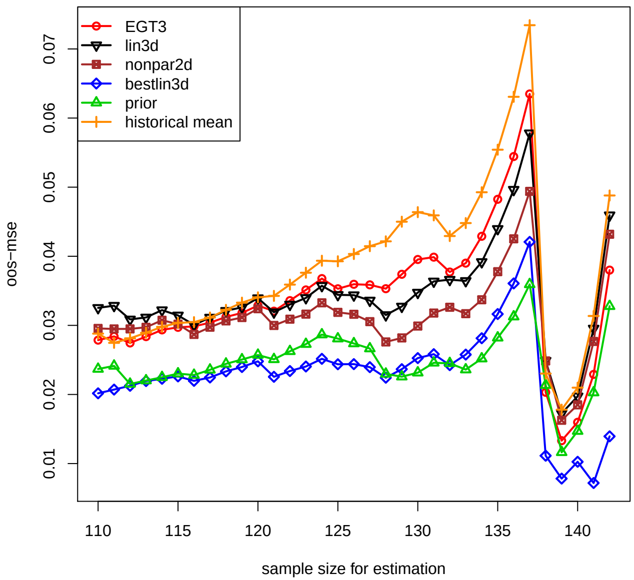
		prior						no prior
		S	d	e	r	L	inf	
corr.	e	8.5	8.4	9.1	16.5	10.9	11.2	11.5
fac-	e, L	9.8	12.8	14.4	20.9	13.0	13.2	14.5
tor	e, r	10.9	9.0	12.1	11.9	10.3	14.0	14.7

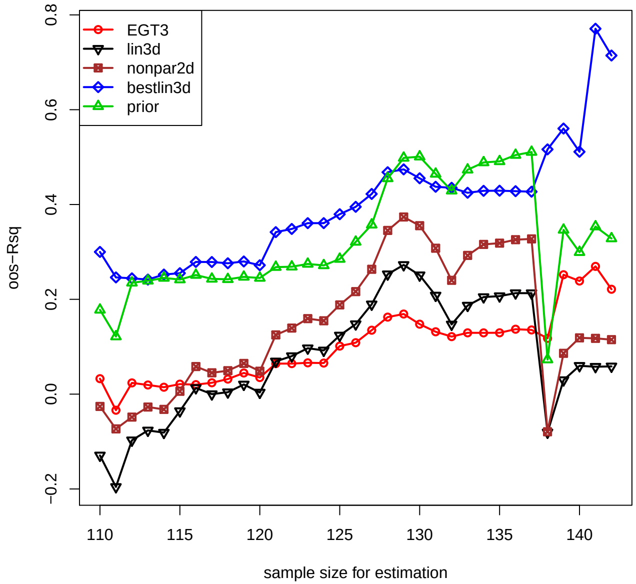
- Our choice: linear model with **risk-free** as prior and fully nonparametric with **earnings by price** and **long-term interest rate** for correction factor.

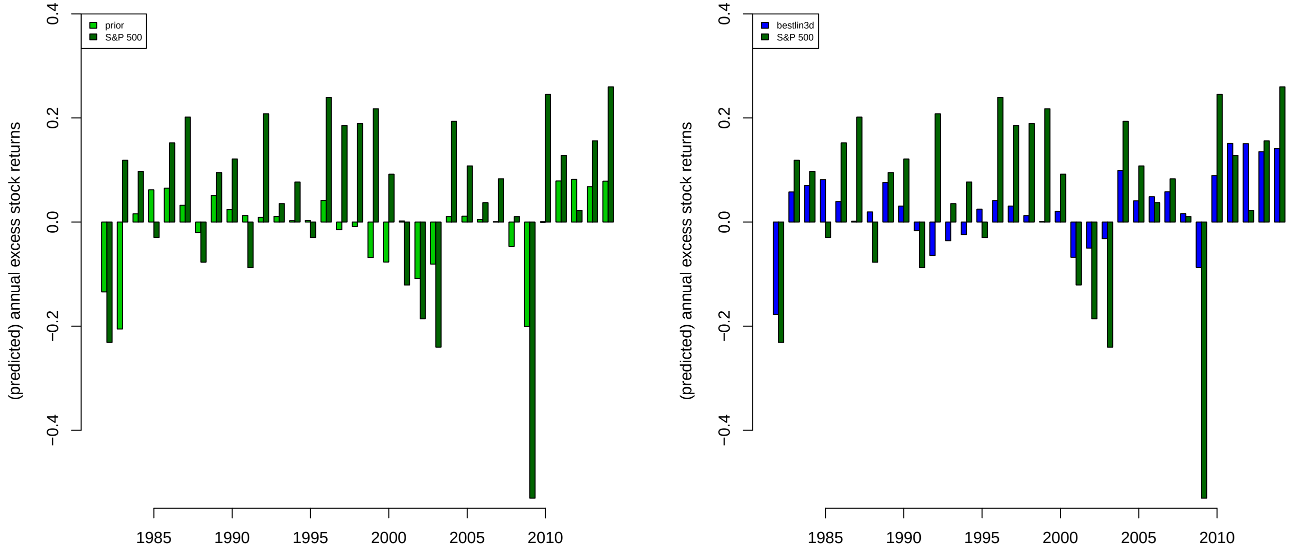

risk−free: 1.0
returns
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.05
0.10
0.15
earnings by price
earnings by price: 0.05
returns
0
2
4
6
8
10
12
risk−free

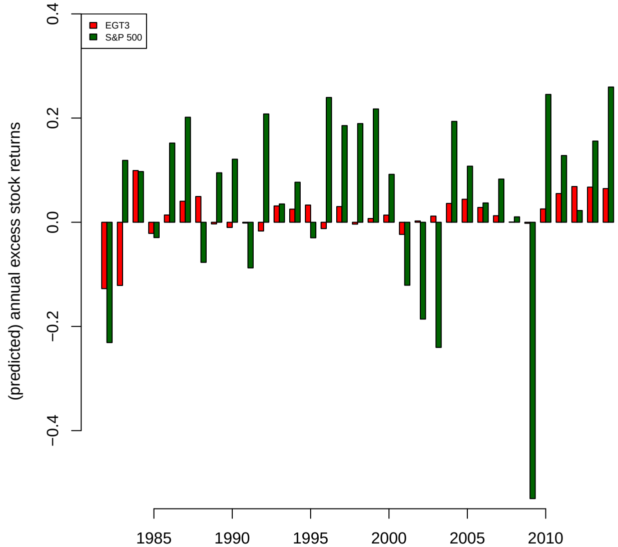
We compare **six different models**:

1. **EGT3** is based on the complete subset regression by Elliott et al. (2013), $k = 3$
2. **lin3d** is the linear model on $\{e, r, L\}$
3. **nonpar2d** is the fully nonlinear model on $\{e, L\}$
4. **bestlin3d** is the three-dimensional linear model that performs best (in terms of oos-mse) with hindsight at each period in time
5. **prior** is the model guided by prior with a linear prior on $\{r\}$ and nonparametric correction factor on $\{e, L\}$, as chosen with our validation criterion
6. **historical mean**


EGT3
lin3d
nonpar2d
bestlin3d
prior
historical mean
oos−mse
0.01
0.02
0.03
0.04
0.05
0.06
0.07
110
115
120
125
130
135
140
sample size for estimation


EGT3
lin3d
nonpar2d
bestlin3d
prior
oos−Rsq
−0.2
0.0
0.2
0.4
0.6
0.8
110
115
120
125
130
135
140
sample size for estimation


prior
S&P 500
(predicted) annual excess stock returns
0.4
0.2
0.0
−0.2
−0.4
1985
1990
1995
2000
2005
2010
bestlin3d
S&P 500
(predicted) annual excess stock returns
0.4
0.2
0.0
−0.2
−0.4
1985
1990
1995
2000
2005
2010


EGT3
S&P 500
(predicted) annual excess stock returns
0.4
0.2
0.0
−0.2
−0.4
1985
1990
1995
2000
2005
2010

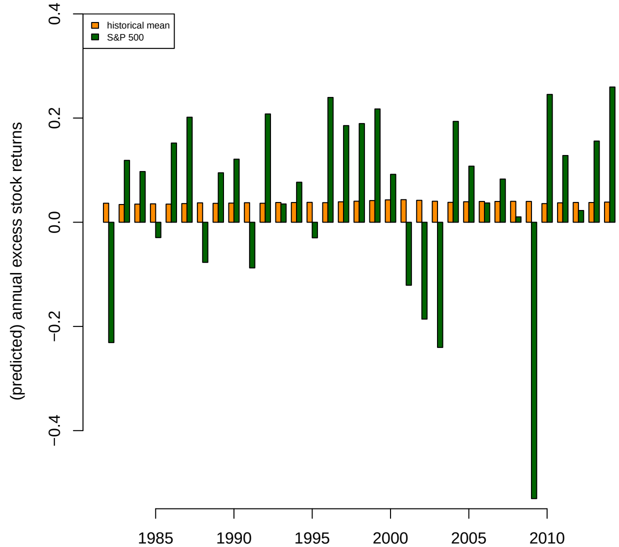

historical mean
S&P 500
(predicted) annual excess stock returns
0.4
0.2
0.0
−0.2
−0.4
1985
1990
1995
2000
2005
2010

Tabelle: Oos-mse of predictions along the business cylce, oos-period: 1982–2014

model	whole period	business-cycle peaks	business-cycle troughs
EGT3	0.028	0.021	0.074
lin3d	0.033	0.028	0.062
nonpar2d	0.030	0.026	0.052
bestlin3d	0.020	0.016	0.051
prior	0.024	0.022	0.033
historical mean	0.029	0.019	0.103

Tabelle: Subsample stability

	1927–1956		1956–1985		1985–2014	
model	oos-mse	oos-R^2	oos-mse	oos-R^2	oos-mse	oos-R^2
EGT3	0.055	10.6	0.018	34.5	0.028	1.9
lin3d	0.060	2.9	0.017	35.2	0.025	12.7
nonpar2d	0.052	16.3	0.024	12.0	0.030	-2.6
bestlin3d	0.049	21.4	0.016	41.6	0.024	16.3
prior	0.052	15.9	0.018	34.8	0.024	16.8
historical mean	0.062		0.027		0.029	

Tabelle: Portfolio performance: Compound annual growth rate (cagr) and maximum drawdown (mdd)

	1982–2014		2000–2014		2008–2014	
	cagr	mdd	cagr	mdd	cagr	mdd
(a) buy-and-hold	9.1	40.7	3.7	40.7	5.1	40.7
(b) simple strategy						
EGT3	10.9	30.5	11.2	30.5	17.2	0.0
bestlin3d	14.8	2.0	16.1	0.0	17.2	0.0
prior	12.5	5.5	11.3	5.5	16.7	0.0
historical mean	11.0	38.8	8.0	38.8	8.8	38.8
(c) conservative strategy						
EGT3	9.6	11.7	9.8	11.7	14.4	0.0
bestlin3d	12.6	0.0	13.8	0.0	16.8	0.0
prior	10.5	0.2	9.7	0.2	14.2	0.0
historical mean	9.0	18.1	6.9	18.1	7.3	18.1

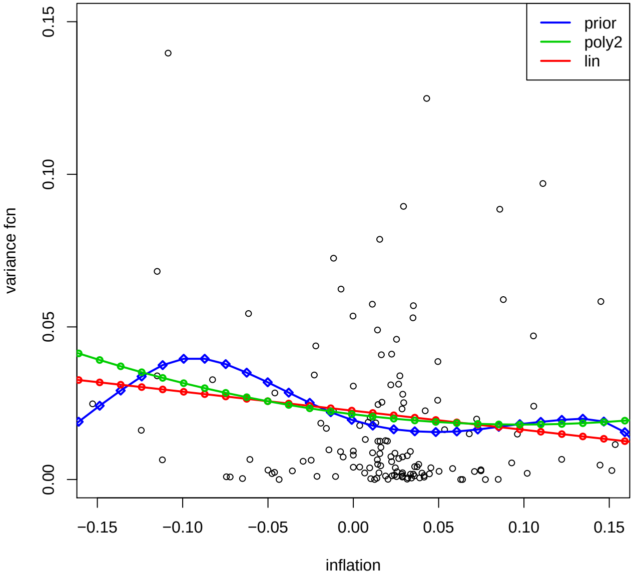
- **estimation of the variance fct.**:

$$\nu(\boldsymbol{x}_1) = (\nu_\theta(\boldsymbol{x}_2) + c) \cdot \frac{\nu(\boldsymbol{x}_1)}{\nu_\theta(\boldsymbol{x}_2) + c}$$

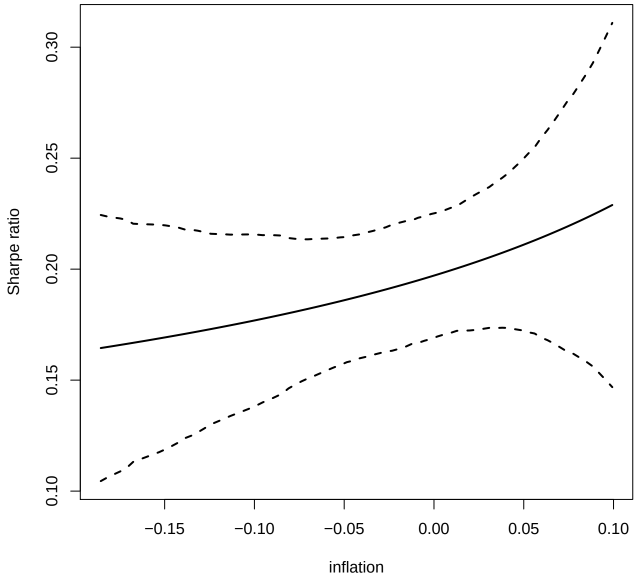
Tabelle: Predictive power (in percent)

		prior in mean&var		prior in mean	no prior
		R_V^2	prior	R_V^2	R_V^2
corr.	S	7.4	L	1.7	0.5
fac-	$S, spread$	5.9	L	1.1	-0.5
tor	$inf, spread$	10.7	L, inf	5.4	-1.2

- In all models **long-term interest** and/or **inflation** as prior
- Best model without any prior: $R_V^2 = 1.8$ with **spread**


prior
poly2
lin
variance fcn
0.00
0.05
0.10
0.15
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
inflation

Estimator of **Sharpe ratio** evaluated at mean values for other covariates than **inflation** with confidence intervals based on **wild bootstrap**



Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

Summary

- Estimator for the Sharpe ratio as a two stage estimator of conditional mean and variance function using a **combined estimator with parametric priors**.
- Include **prior knowledge** in the statistical modelling process. Improve this way smoothing due to **bias reduction**.
- An **ex-ante measure of asset performance** incorporating the risk taken.

Outlook

- **Market timing**, **Out-of-sample performance**, behavior during **declining markets**?
- Use of **generated regressors** as in Scholz et. all (2016) or **technical indicators** as in Neely et. all (2014)

Prediction
Sharpe Ratio
R_V^2
Combined Estimator
Estimation Mean Fct.
Estimation Var Fct.
Simulation Study
Empirical Study
Outlook/ Summary

Thank you for your attention!