Semiparametric Structure Guided By prior knowledge

with applications in economics

Dissertation

Presented for the Degree of Doctor rerum politicarum
at the Faculty of Economic Sciences
of the Georg-August-Universität Göttingen

by
Michael Scholz
from
Leisnig, Germany

Göttingen, 2011

First Examiner: Prof. Dr. Stefan Sperlich
Second Examiner: J.Prof Xiaohua Yu, Ph.D.

Day of oral exam: 08. April 2011

Contents

List of Figures

List of Tables

Abbreviations and Symbols

List of Symbols for Chapter 2:

$V(\mathbf{p}, x)$	indirect utility function
$\mathbf{p},\ \tilde{\mathbf{p}}$	vector of log-prices and transformed log-prices
$x,\ \tilde{x}$	log-total expenditure and transformed log-total expenditure
$\mathbf{w}$	vector of expenditure share functions
W_i^j	expenditure share for good j of individual i
P_i^j	log-price for good j of individual i
X_i	log-total expenditure of individual i
$\mathbf{A}$	parameter matrix of price coefficients
$\mathbf{f}$	vector of Engel-Curves
∇_x	derivative (gradient) with respect to x
$\mathbf{0}$	vector of zeros
$\boldsymbol{\alpha}$	vector of local level of $\mathbf{f}$
$\boldsymbol{\beta}$	vector of local derivative of $\mathbf{f}$
$\boldsymbol{\Gamma}$	matrix of local level of $\mathbf{A}$
$\boldsymbol{\Delta}$	matrix of local derivative of $\mathbf{A}$
$\boldsymbol{\Omega}$	weighting matrix
$\Delta_i,\ K_i,\ S_i,\ T_i$	abbreviations defined on p. 13
K	symmetric kernel function
$g,\ h,\ h_0$	smoothing bandwidth
$C_i^s,\ T_i^{s,-t}$	abbreviations defined on p. 15
$\mathbb{1}$	indicator function
$O,\ o$	Landau notation
$\varepsilon_i^j,\ \sigma_\varepsilon^j$	residuals and their standard deviation
$W_i^{j*},\ \varepsilon_i^{j*}$	bootstrap sample, bootstrap residual
u_i	standard normal random variable
χ_k^2	χ^2–distributed random variable with k degrees of freedom
$\lfloor z \rfloor$	largest integer not greater than z
M	number of price directions
$N,\ N_0$	sample size
$\langle\ \cdot, \cdot\ \rangle$	dot product

List of Symbols for Chapters 3 and 4:

$R^2,\ R^2_{adj},\ R^2_V$	measures of prediction quality (classical, adjusted, validated)
Y, Y^*, Y^{**}, Y^{***}	dependent stochastic variables
$X,\ \boldsymbol{X},\ v,\ w$	explanatory variables
$\varepsilon,\ \zeta,\ \xi$	mean zero error terms
μ	mean
$\bar{Y}$	sample mean
$g,\ p,\ \tilde{g}, m$	unknown functions
$h,\ h_0$	smoothing bandwidth
$\hat{g}_{-t},\ \bar{Y}_{-t}$	cross-validated values
S	excess stock return
P	stock price
D	dividend
R	discount rate
G	dividend growth
d	dividend by price
e	earnings by price
r	risk-free rate
L	long-term interest rate
b	bond yield
inf	inflation
T	prediction horizon
$b(x)$	deterministic part of predicted realisation
$\sigma(x)$	stochastic part of predicted realisation
$u_x,\ u^b$	random variable with mean zero and variance one
$\hat{m}_{NW},\ \hat{m}_{LL}$	Nadaraya-Watson and local-linear estimator
K	kernel function
$n,\ N,\ \mathcal{T}$	sample sizes
$f_X,\ f_{ij}$	densities
$\mathcal{F}_i^j$	σ-algebra
$\alpha(n)$	mixing coefficient
$\Delta,\ \tilde{\Delta}$	rate of convergence
$O,\ o$	Landau notation
τ	test statistic defined on page 65
$\hat{\varepsilon}^0$	residuals under the null hypothesis
F^*	empirical distribution function
$\beta,\ \boldsymbol{\beta}$	regression coefficients
g_θ	parametric function with parameter θ
$c,\ C$	generic constants

List of Abbreviations:

CSE	Copenhagen stock exchange
FED	Federal Reserve System
MSE	Mean squared error
OLS	Ordinary least squares
SUR	Seemingly unrelated regression
VAR	Vector autoregression
cf.	compare (Latin: confer)
e. g.	for example (Latin: exempli gratia)
i. e.	that is (Latin: id est)
i.i.d.	independent and identically distributed

Acknowledgements

It is a pleasure to thank those who made this thesis possible. I am especially grateful to my supervisor, Stefan Sperlich, Professor at the University of Geneva, whose encouragement, guidance, and academic support from the first ideas to the final version enabled me to develop a better understanding of the topic and my intuition for statistical and econometric problems. I am similarly grateful to Jens Perch Nielsen, Professor at Cass Business School, without whose knowledge and assistance this thesis would not have been successful. I am indebted to both of them for their practical advice, helpful comments, and the many insightful suggestions that they made in numerous meaningful debates.

Special thanks are also due to my colleagues at the University of Göttingen for the very enjoyable cooperation and invaluable assistance. In particular, I would like to thank Professor Xiaohua Yu for his willingness to act as examiner as well as for his detailed and patient course that gave me a deeper insight into consumer behavior and demand analysis. I gratefully acknowledge the support received from Professor Zucchini, whose door was always open, and from Professor Böker, whose clear and accurate approach sharpened my view for details.

Many thanks go to the co-authors, especially to Krishna Pendakur, Professor at the Simon Fraser University, who contributed his knowledge to the work on consumer behavior and demand analysis, and to Dietmar Ferger, Professor at the TU Dresden, who taught me basic and advanced mathematical statistics, and who guided me in the publishing process of my Diploma thesis.

Finally, I wish to express my love and gratitude to my family. Without their understanding, continuous support, and dedication throughout my studies, it would not have been possible for me to complete this thesis.

Göttingen, April 2011

Chapter 1

Overview

"October. This is one of the peculiarly dangerous months to speculate in stocks. The others are July, January, September, April, November, May, March, June, December, August, and February."

Mark Twain

Among many developments in statistical modelling in recent years, non- and semiparametric methods have proved to be a particularly powerful data-analytic tool. Nevertheless, there still exist justified doubts regarding there forecasting performance, for example in the context of financial time series. The aim of this thesis is to demonstrate that, by suitable modification, these techniques can perform well in different economic fields, like empirical demand analysis or prediction of stock returns, if they are adapted to the specific application under investigation.

The relationship between variables in many applications exhibit special features such as heteroscedasticity or nonlinear functional forms. Traditionally, the latter problem is addressed by making use of parametric models which are simple to apply. However, one is then faced to the problem of choosing among infinitely many different nonlinear forms, a choice that could be critical. The global structure for the underlying stochastic dynamics offered by a classical parametric approach is certainly useful for obtaining a first impression of the relationships of interest. But to find more detailed structures that help to better understand and, in particular, to better approximate the real world, more flexible techniques are required. Non- and semiparametric methods relax assumptions regarding the form of the regression function, thereby allowing for more flexibility in modelling the relationships in a more data-driven manner. An introduction to the concept of non- and semiparametric smoothing can be found, for example in Härdle et al. (2004). This thesis is based on local-polynomial modelling techniques which were constructed to solve a number of specific economic questions.

Certain problems in data analysis cannot be solved using the described techniques. In particular, the so called *curse of dimensionality* is a case in point; the higher the dimension of the problem, the more sparsely distributed are the observations, leading to a deterioration in performance, and of the accuracy of the estimates. For illustrations of this difficulty see Silverman (1986), Härdle (1990), Scott (1992), or Fan and Gijbels (1996). To circumvent this problem, the imposition of structure in the model is often proposed in the statistical literature. Thus numerous articles use additivity or separability (e. g. Stone (1985), Hastie and Tibshirani (1990), or Nielsen and Linton (1998)). In contrast, this thesis makes use of the semiparametric nature of economic problems to reduce dimensionality, and is based on the structure that is inherent in the economic process that generates the data. A key feature in this thesis is to show how prior knowledge can guide the modelling process. This is done either by directly applying economic theory (to suggest limiting behavior, monotonicity, etc.) or by examining simple parametric models to identify the coarse features of the relationships. The use of prior knowledge not only improves the plausibility of the model but also the interpretability of the results. Furthermore, it can be used to address some other well-known problems associated with fully nonparametric approaches. For example, the estimation accuracy on boundaries can be improved by appropriate transformations motivated by the economic context, or the bias can sometimes be reduced by applying a semiparametric approach.

Each chapter of this thesis is self-contained. It is possible to skip a part such that the rest still remains understandable. Chapter 1 sets the scene for this work and outlines the specific economic problems and their possible solutions. Chapter 2 is a contribution to the analysis of consumer expenditure and price micro-data, while Chapter 3 and 4 address the prediction of excess stock returns. The use of nonparametrically generated bond yields is proposed and prior information about the shape of the unknown conditional mean function is used in the estimation process. Finally, Chapter 5 concludes and gives a short outlook. Chapters 2 to 4 are based on separate papers. Consequently, this involves some replication especially in the introductions and the material related to non- and semiparametric techniques. To give an overview of the topics treated in this work, we now highlight the main ideas and results, as well as the contributions to the research.

In Chapter 2, a semiparametric model of consumer demand, defined as the relationship between quantity demands, prices and total expenditure, is considered. Since typical consumer demand micro-data have a large amount of variation in total expenditure across consumers, it might be possible to identify complex relationships between demands and expenditure. In this model, indirect utility is specified as a partially linear function with a nonparametric part for expenditure and a parametric part (with fixed- or varying-coefficients) for prices. Since the starting point is a model of indirect utility, rationality restrictions like homogeneity and Slutsky symmetry are easily imposed. The resulting model for expenditure shares, comprising functions of expenditures and prices, is given (locally) by a fraction whose numerator is partially linear, but whose denominator is nonconstant and given by the derivative of the numerator. The key idea is that, by using a local polynomial model for the numerator, the denominator is given by a *lower-order* local polynomial. The model can thus be estimated using modified versions of local polynomial modelling techniques. A new asymmetric version of the wild bootstrap is introduced for inference. It takes into account that expenditure shares lie in the interval $[0, 1]$. To achieve this it is necessary to draw the bootstrap residuals in a special way. A modest Monte Carlo study verified that the proposed techniques work and that the bootstrap procedure achieves an acceptable level of accuracy. Finally an empirical study is described in which the model is implemented on Canadian expenditure and price micro-data. Some of the expenditure share equations in this model exhibit remarkable degrees of nonlinearity. The approach proposed in Chapter 2 contributes to the methods available for addressing the *curse of dimensionality*, because the nonparametric part is reduced to a single dimension. In contrast to other semiparametric models, it is entirely based on observed variables, and does not require any numerical inversions to generate a latent regressor. Consequently, the algorithm is computationally efficient and numerically robust. Large data sets can be handled in acceptable time and the results are readily interpreted. Chapter 2 is based on Pendakur, Scholz, and Sperlich (2010) and, in this thesis, two extra sections are included: (a) the imposition of the Slutsky symmetry in Section 2.6 and (b) the restricted least square for a symmetric matrix in Section 2.7.

Chapter 3 is a contribution to the discussion that addresses the question of whether empirical models are able to forecast the equity premium more accurately than the simple historical mean. This problem is intensively debated in the financial literature. The low predictive power is disappointing, even when using nonparametric models that make use of typical predictor variables. Classical approaches are based on the well-known Gordon growth or dividend discount model, and interpret the price of a stock today as the discounted present value of future cash flows to the investor. In contrast, the so called FED model directly relates yields on stocks to yields on bonds, but fails in predicting stock returns. Motivated by the co-movement of bond and stock returns, one could pose the question of whether expected returns on stocks and bonds are driven by the same information, and to what extend they move together. Chapter 3 proposes the inclusion of the current bond yield in a prediction model, which results in a notable improvement of the prediction of stock returns, as measured by the *validated* R^2. This way, the bond captures the perhaps most important part of the stock return, namely the part related to the change in long-term interest rate. Since the current bond yield is unknown, it is nonparametrically predicted in a prior step. The essential point is that the inclusion of the generated bond can be seen as a kind of dimension reduction that imposes more structure in an appropriate way that circumvents the *curse of dimensionality and complexity.* Since nonparametrically generated regressors are included in a nonparametric prediction approach of *dependent* time series data, this chapter also provides a theoretical justification for the use of constructed variables in the nonparametric regression. In an empirical part the proposed method is implemented on Danish stock and bond market data. The inclusion of predicted bond yields greatly improves the prediction quality of stock returns. The best prediction model (for one-year stock returns) not only outperforms the simple historical mean, it also results in an increase of the prediction quality by a factor of almost 5 compared to the best model without constructed bonds. Chapter 3 is based on Scholz, Sperlich, and Nielsen (2011) and presents additional tables of results in Section 3.7.

Chapter 4, which is based on Scholz, Nielsen, and Sperlich (2011), also investigates whether equity returns (or premiums) can be predicted by empirical models. While many authors favor the historical mean, or other simple parametric methods, this part of the thesis focuses on nonlinear relationships. A fully nonparametric approach serves as starting point and allows a flexible nonlinear form of the conditional mean function. A straightforward bootstrap-test confirms that non- and semiparametric techniques yield better forecasts than do parametric models. It establishes that the proposed techniques work and yield significantly better results. In contrast to the previous chapter, a new approach is proposed to include prior knowledge in the forecasting procedure of excess stock returns. Economic theory directly guides the modelling process in an innovative way. In consequence of this approach a dimension and bias reduction is achieved, both to impose more structure to circumvent the *curse of dimensionality.* It can be shown that certain boundary and bandwidth difficulties are thereby overcome using a *single* idea. The available prior information is included in a semiparametric fashion, where the nonparametric smoother is multiplicatively guided by the prior. Here, the direct application of economic theory, or the examination of standard parametric models, lead to the necessary prior. The potential of the proposed method is illustrated in an empirical part using annual American stock market data. The bootstrap test shows that non- and semiparametric models are more appropriate than linear regressions, and that the inclusion of prior knowledge greatly improves the prediction quality. The results show that the proposed approach outperforms the simple historical mean. Its predictive power is 35% higher than that of the best fully nonparametric model.

Chapter 2

Semiparametric Indirect Utility and Consumer Demand

2.1 Introduction

The specification and estimation of consumer demand systems, defined as the relationship between quantity demands, prices and total expenditures, represent many long-standing problems in econometric theory. Recent work has focused on the inclusion of highly nonlinear relationships between quantity demands (or expenditure shares) and total expenditures into empirical models of consumer demand. Since typical consumer demand micro-data have a large amount of variation in total expenditures across consumers, it might be possible to identify complex relationships between demands and expenditure. Consumer demand models must satisfy a set of nonlinear cross-equation rationality restrictions (see, for example, Deaton and Muellbauer (1980), or Varian (1978)), known as the Slutsky symmetry restrictions. Such complex relationships have been hard to incorporate into semi- and nonparametric approaches.

This chapter presents a semiparametric approach to the consumer demand problem which allows for the imposition of the Slutsky symmetry restrictions. We use a flexible nonparametric estimation method in the total expenditure direction, where the data provide a lot of information, to get demands which are arbitrarily flexible in total expenditure (i. e., arbitrarily flexible Engel curves). However, in the price directions, where the data are less rich, we propose a parametric structure.

Like most models for consumer demand, our model uses the vector of expenditure shares commanded by each good as the dependent variable. In this chapter, we introduce a wild bootstrap that accounts for the fact that expenditure shares lie in the interval [0,1]. The idea is to draw bootstrap residuals from a local adaptive distribution that respects the boundaries via asymmetry.

Nonparametric approaches to consumer demand started by considering Engel curves, defined as the relationship between expenditure shares and the total expenditures of the consumer at a fixed vector of prices. In these models only 1 nonparametric direction is considered while the others are fixed. Work by Blundell, Duncan, and Pendakur (1998) or Blundell, Chen, and Kristensen (2003) revealed considerable complexity in the shapes of Engel curves. A fully nonparametric approach, which considers both price and expenditure directions together and which allows for the imposition of rationality restrictions, has been developed by Haag, Hoderlein, and Pendakur (2009). In their article, the shapes of the demand equations are not restricted, but the *curse of dimensionality* is a case in point: with M price directions and 1 expenditure direction, the researcher is confronted with a $M + 1$ dimensional problem. Even if homogeneity, another rationality condition, is imposed, the dimensionality of the problem reduces only to M, which is still very high in typical applications.

Parametric approaches like the popular Almost Ideal (Deaton and Muellbauer, 1980), dynamic Almost Ideal (Mazzocchi, 2006), Translog (Jorgensen, Lau, and Stoker, 1980) and Quadratic Almost Ideal (Banks, Blundell, and Lewbel, 1997) demand models typically impose strict limits on the functional complexity of Engel curves. In these cases, they must be linear, nearly linear, or quadratic, respectively, in the log of total expenditure. This lack of complexity is driven by the need for these parametric models to satisfy the Slutsky symmetry restrictions.

A major use of consumer demand systems is in policy analysis: demand systems are used to assess whether or not indirect tax changes are desirable, and are used to assess changes in the cost-of-living. In this regard, lack of complexity has costs: in particular, if the Engel curve is wrong, then all consumer surplus calculations (including cost-of-living calculations) are also wrong. For example Banks, Blundell, and Lewbel (1997) or Lewbel and Pendakur (2009) show that the false imposition of linear and quadratic Engel curves can lead to very misleading estimates of behavioural and welfare responses to indirect tax changes.

In between the fully nonparametric and the fully parametric approaches, we have the realm of semiparametric econometrics. Two recent papers have explored this area. Lewbel and Pendakur (2009) propose a fully parametric approach which satisfies rationality restrictions and for which Engel curves can be arbitrarily complex. Because their model allows for arbitrarily complex Engel curves but parametrically restricted dependence of expenditure shares on prices, it may be interpreted as semiparametric. However, their approach relies critically on a particular interpretation of the *error term* in the regression: it must represent unobserved preference heterogeneity, and thus cannot be measurement error or any other deviations from optimal choice on the consumer's part. Further, Lewbel and Pendakur (2009) do not allow for a varying-coefficients structure for price effects.

Pendakur and Sperlich (2010) propose a semiparametric model which allows for these latter interpretations of the role of the error term, does not restrict the shape of Engel curves, and incorporates price effects either parametrically or semiparametrically (through fixed- or varying-coefficients, cf. Sarmiento (2005)). Pendakur and Sperlich (2010) propose a model in which expenditure-shares are nonparametric in utility, an unobserved regressor, and (semi-)parametric in log-prices. The familiarity of this partially linear form makes the model appealing, but the unobserved regressor (utility) must be constructed under the model via numerical inversion of the (unknown) cost function. In the present approach, we propose a model in which utility is nonparametric in log-expenditure and parametric in log-prices. This results in a model of expenditure-shares which is locally nonlinear but has no unobserved or generated regressors. All of these semiparametric approaches address the *curse of dimensionality*: each of them has just 1 nonparametric dimension.

The local nonlinearity of our approach is based on the fact that we model indirect utility as a partially linear function. Since Roy's Identity (Roy (1947)) gives expenditure shares as the ratio of derivatives of indirect utility, expenditure shares in our model are also given by a ratio. This ratio has nonparametric functions in the numerator and their derivatives in the denominator. The key idea is that, by using a local polynomial model for the numerator, the denominator is given by a *lower-order* local polynomial that comprises the derivatives of the numerator. This fact suggests a natural iterative procedure to estimate the model. Our algorithm is computationally efficient and numerically robust. Large data sets can be handled in acceptable time and the results are readily interpreted.

In Section 2.2 we introduce the model. In Section 2.3 we discuss the basic estimation idea, give the associated algorithm and describe the bootstrap inference. The nonparametric part of the model is estimated with an univariate local linear smoother on transformed data, a method that can easily be applied in empirical research. For the parametric part of the model we use a restricted least squares estimator to satisfy the Slutsky symmetry restrictions. For inference we introduce an asymmetric version of the wild bootstrap. To fulfill the constraints that the (bootstrap) responses must be in the interval $[0, 1]$, we propose a local adaptive χ^2-distribution for the bootstrap errors. A nice feature of our approach is that confidence intervals created this way are narrower than those based on standard wild bootstrap.

In Section 2.4 we evaluate our proposed methods and the accuracy of the bootstrap procedure in a small simulation study. We also implement the model on Canadian price and expenditure data. In the empirical part we find that some expenditure share equations in this model exhibit remarkable degrees of nonlinearity. Section 2.5 concludes and discusses extensions.

2.2 A Semiparametric Model for Indirect Utility

Define the indirect utility function $V(\mathbf{p}, x)$ to give the maximum utility attained by a consumer when faced with a vector of log–prices $\mathbf{p} = (p^1, \dots, p^M)$ and log–total expenditure x. Let the expenditure share of a good be defined as the expenditure on that good divided by the total expenditure available to the consumer. Denote $\mathbf{w} = (w^1, \dots, w^M)$ as the vector of expenditure share functions and note that since expenditure shares sum to 1, $w^M = 1 - \sum_{j=1}^{M-1} w^j$. Let $\{W_i^1, \dots, W_i^M, P_i^1, \dots, P_i^M, X_i\}_{i=1}^N$ be a random vector giving the expenditure shares, log–prices and log–total expenditure of a population of N individuals. Note that, as commonly done in the literature of demand systems, we use the superscript notation for single elements of vectors or matrices, i. e. for single goods or commodities, and the subscript for individuals.

2.2.1 A Partial Linear and Varying-Coefficients Model

We consider two semiparametric specifications of the indirect utility function. First, we consider a partially linear (or, fixed-coefficients) specification of the form

$$V(\mathbf{p}, x) = x - \sum_{k=1}^{M} f^k(x) p^k - \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} a^{kl} p^k p^l, \tag{2.1}$$

or, in matrix notation,

$$V(\mathbf{p}, x) = x - \mathbf{f}(x)^\top \mathbf{p} - \frac{1}{2} \mathbf{p}^\top \mathbf{A} \mathbf{p}, \tag{2.2}$$

where $\mathbf{f} = (f^1, \dots, f^M)^\top$ are unknown differentiable functions of log–total expenditure and $\mathbf{A} = \{a^{kl}\}_{k,l=1}^M$ are parameters. We impose the normalisation that $a^{kl} = a^{lk}$, or, equivalently, $\mathbf{A} = \mathbf{A}^\top$. This is not a restriction: since $p^k p^l = p^l p^k$, there is a symmetric version of $\mathbf{A}$ that yields the same V as any asymmetric version. Second, we consider the varying-coefficients extension of this model:

$$V(\mathbf{p}, x) = x - \sum_{k=1}^{M} f^k(x) p^k - \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} a^{kl}(x) p^k p^l, \tag{2.3}$$

or, in matrix notation,

$$V(\mathbf{p}, x) = x - \mathbf{f}(x)^\top \mathbf{p} - \frac{1}{2} \mathbf{p}^\top \mathbf{A}(x) \mathbf{p}, \tag{2.4}$$

where $a^{kl}(x) = a^{lk}(x)$ for all k, l, or, equivalently, $\mathbf{A}(x) = \mathbf{A}(x)^\top$.

Expenditure shares are functions of total expenditure and all prices. Roy's Identity relates the expenditure share for good j, $w^j(\mathbf{p}, x)$, to derivatives of the indirect utility function: $w^j(\mathbf{p}, x) = -\left[\partial V(\mathbf{p}, x)/\partial p^j\right] / \left[\partial V(\mathbf{p}, x)/\partial x\right]$. Application of Roy's Identity to the fixed-coefficients model yields

$$w^j(\mathbf{p}, x) = \frac{f^j(x) + \sum_{k=1}^{M} a^{jk} p^k}{1 - \sum_{k=1}^{M} \nabla_x f^k(x) p^k},$$

with ∇_x indicating the derivative (here of $f^k(x)$) with respect to x; or, in matrix notation,

$$\mathbf{w}(\mathbf{p}, x) = \frac{\mathbf{f}(x) + \ \mathbf{A}\mathbf{p}}{1 - \nabla_x \mathbf{f}(x)^\top \ \mathbf{p}}.$$

For the varying-coefficients model we get

$$w^j(\mathbf{p}, x) = \frac{f^j(x) + \sum_{k=1}^{M} a^{jk}(x) p^k}{1 - \sum_{k=1}^{M} \nabla_x f^k(x) p^k - \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} \nabla_x a^{kl}(x) p^k p^l},$$

or, in matrix notation,

$$\mathbf{w}(\mathbf{p}, x) = \frac{\mathbf{f}(x) + \ \mathbf{A}(x)\mathbf{p}}{1 - \nabla_x \mathbf{f}(x)^\top \mathbf{p} - \frac{1}{2} \mathbf{p}^\top \nabla_x \mathbf{A}(x) \mathbf{p}}.$$

We describe how to estimate these expenditure share equations in Section 2.3.

The motivation for these models is as follows. In real-world applications, there is typically a large amount of observed variation in total expenditures, so one may reasonably hope to identify a nonparametric component in that direction. However, typical micro-data sources do not have nearly as much variation in the price directions, which suggests that partially linear modelling might describe these effects sufficiently well. If in addition, the researcher feels that more may be identified on the strength of observed price variation, the varying-coefficients model allows price effects in the model (2.3) to be different at different expenditure levels. This would seem to be a pure advantage of the varying-coefficients approach. However, in practise, this extension seriously increases the variance and computational cost of the estimates. In particular, the algorithm for model (2.3) is about five times slower than the one for model (2.1). The important feature here is that nonparametric dimensionality is 1 in both models.

2.2.2 Rationality Restrictions

Rationality is comprised of three conditions: homogeneity, symmetry and concavity. Here we will deal only with symmetry and homogeneity (concavity is a topic of its own, investigated, e. g. in Millimet and Rusty (2008)). Slutsky symmetry (see, for example, Mas-Colell, Whinston, and Green (1995)) relates to the fact that expenditure share equations are derived in terms of the derivatives of indirect utility V. Slutsky symmetry gives minimal restrictions under which expenditure share equations lead to a unique indirect utility function. In our context it is satisfied if and only if $\mathbf{A} = \mathbf{A}^\top$ in the expenditure share equations[1] (or, in the varying-coefficients case, if $\mathbf{A}(x) = \mathbf{A}(x)^\top$). In the indirect utility function the restriction of these matrices to symmetry is only a normalisation. However, in the expenditure share equations, this constraint is crucial. In particular, because each expenditure share equation could be estimated separately, the estimated matrix could be asymmetric. In our estimation section below we use an algorithm which maintains symmetry, and which is the semiparametric analog to a linearly restricted Seemingly Unrelated Regression (SUR) estimator[2].

Homogeneity is sometimes referred to as *no money illusion*. If consumers do not suffer from money illusion, then scaling prices and expenditures by the same factor cannot affect utility. This requires that indirect utility is homogeneous of degree zero in (unlogged) prices and expenditure. This can be achieved by dividing all prices and expenditure by the price of the M-th expenditure category. Note that we use logarithms, so we subtract p^M from each log-price and from log-expenditure in the indirect utility function. For the fixed-coefficients case, this yields

$$V(\mathbf{p}, x) = (x - p^M) - \sum_{k=1}^{M-1} f^k(x - p^M) \cdot (p^k - p^M) - \frac{1}{2} \sum_{k=1}^{M-1} \sum_{l=1}^{M-1} a^{kl}(p^k - p^M)(p^l - p^M),$$

in model (2.1) and analogously in model (2.3). The sums go only to $M-1$ because the M-th element of each sum (which multiplies $p^M - p^M$) is zero. Denoting $\tilde{x} = x - p^M$, $\tilde{p}^j = p^j - p^M$ and $\tilde{\mathbf{p}} = (\tilde{p}^1, \dots, \tilde{p}^{M-1})$ we may write this more compactly as

$$V(\tilde{\mathbf{p}}, \tilde{x}) = \tilde{x} - \sum_{k=1}^{M-1} f^k(\tilde{x}) \cdot \tilde{p}^k - \frac{1}{2} \sum_{k=1}^{M-1} \sum_{l=1}^{M-1} a^{kl} \tilde{p}^k \tilde{p}^l, \tag{2.5}$$

with a^{kl} depending on $\tilde{x}$ in the varying-coefficients case. We thus estimate only the first $(M-1)$ elements of $\mathbf{f}$ and $\mathbf{w}$, and the first $(M-1)$ rows and columns of $\mathbf{A}$. In matrix notation, this may be written with $\mathbf{f} = (f^1, \dots, f^{M-1})^\top$ and $\mathbf{A} = \{a^{kl}\}_{k,l=1}^{M-1}$ as

$$V(\tilde{\mathbf{p}}, \tilde{x}) = \tilde{x} - \mathbf{f}(\tilde{x})^\top \tilde{\mathbf{p}} - \tilde{\mathbf{p}}^\top \mathbf{A} \tilde{\mathbf{p}},$$

for the fixed-coefficients case and

$$V(\tilde{\mathbf{p}}, \tilde{x}) = \tilde{x} - \mathbf{f}(\tilde{x})^\top \tilde{\mathbf{p}} - \tilde{\mathbf{p}}^\top \mathbf{A}(\tilde{x}) \tilde{\mathbf{p}},$$

[1] For reasons of clarity and comprehensibility, we skip here the presentation of the imposition of Slutsky symmetry and defer it to the appendix in Section 2.6.

[2] For more details, cf. Section 2.7.

for the varying-coefficients case. Once again, since expenditures sum to 1 by construction, we have $w^M(\tilde{\mathbf{p}}, \tilde{x}) = 1 - \sum_{i=1}^{M-1} w^i(\tilde{\mathbf{p}}, \tilde{x})$, and we need only consider the first $(M-1)$ expenditure share equations.

As before, we get the expenditure share equations

$$\mathbf{w}(\tilde{\mathbf{p}}, \tilde{x}) = \frac{\mathbf{f}(\tilde{x}) + \mathbf{A}\tilde{\mathbf{p}}}{1 - \nabla_{\tilde{x}}\mathbf{f}(\tilde{x})^\top \tilde{\mathbf{p}}}, \tag{2.6}$$

for the fixed-coefficients model (2.1), and

$$\mathbf{w}(\tilde{\mathbf{p}}, \tilde{x}) = \frac{\mathbf{f}(\tilde{x}) + \mathbf{A}(\tilde{x})\tilde{\mathbf{p}}}{1 - \nabla_{\tilde{x}}\mathbf{f}(\tilde{x})^\top \tilde{\mathbf{p}} - \frac{1}{2}\tilde{\mathbf{p}}^\top \nabla_{\tilde{x}}\mathbf{A}(\tilde{x})\tilde{\mathbf{p}}}, \tag{2.7}$$

for the varying-coefficients model (2.3). Here, $\nabla_{\tilde{x}}\mathbf{f}(\tilde{x})$ is the $(M-1)$–dimensional vector of the derivatives of $\mathbf{f}(\tilde{x})$ with respect to $\tilde{x}$, and $\nabla_{\tilde{x}}\mathbf{A}(\tilde{x})$ is the $(M-1) \times (M-1)$ matrix function equal to the derivatives of $\mathbf{A}$ with respect to $\tilde{x}$.

These expressions for budget shares have a nice feature in comparison to Pendakur and Sperlich (2010). Whereas their model for expenditure shares uses a nonparametric function of a generated regressor which must be constructed under the model using numerical inversion of the unknown cost function, the expression above uses only observed regressors. However, in comparison to Pendakur and Sperlich (2010), which is a partially linear model, the above expression is partially linear only in the numerator. The presence of the denominator seems to complicate the development of an estimation algorithm. However, as we show below, with the use of local polynomials this problem becomes manageable.

2.3 Estimation of the Models

In the following sections we show how to estimate the $(M-1)$–dimensional vector $\mathbf{w}(\tilde{\mathbf{p}}, \tilde{x})$ under the model. These estimates satisfy *adding-up* by construction, since $w^M(\tilde{\mathbf{p}}, \tilde{x}) = 1 - \sum_{i=1}^{M-1} w^i(\tilde{\mathbf{p}}, \tilde{x})$. They satisfy homogeneity (*no money illusion*) also by construction due to the use of normalised prices and expenditures as regressors. Finally, they can satisfy Slutsky symmetry because $\mathbf{A}$ (or $\mathbf{A}(\tilde{x})$) is easily restricted to be a symmetric matrix (see, for example, Deschamps (1988)).

A more difficult question is the restriction of the estimated budget shares to be everywhere in the range $[0, 1]$. This problem is referred to as the *global regularity* problem in the literature on consumer demand. Roughly speaking, demand systems that are not homothetic (i. e. whose budget shares respond to total expenditure) cannot be globally regular without restricting either the domain of $\mathbf{p}, x$ or the domain of model error terms in ad hoc ways. See Pollack and Wales (1991) for a discussion of the former, and Lewbel and Pendakur (2009) for a discussion of the latter. We will judge our estimates in terms of *local regularity*, i. e. in terms of whether or not estimated budget shares are in the range $[0, 1]$ in a $\mathbf{p}, x$ domain of interest. In particular, under homogeneity and when $\mathbf{p} = \mathbf{0}_M$, in both the fixed-coefficients and varying-coefficients model, we have

$$\mathbf{w}(\mathbf{p}, x) = \mathbf{w}(\tilde{\mathbf{p}}, \tilde{x}) = \mathbf{f}(\tilde{x}) = \mathbf{f}(x).$$

The estimated functions $\mathbf{f}(x)$ thus characterise budget shares over a domain spanned by x with log-prices fixed at $\mathbf{0}_M$. If these estimated functions lie in the interval $[0, 1]$, then we say that our estimates are *locally regular* in this sense. Note also that the vast majority of the literature on estimating expenditure systems does not tackle this problem due to its complexity (an exception is Moral-Arce, Rodríguez-Póo, and Sperlich (2007)).

2.3.1 Basic Ideas

The basic idea of estimating the unknown nonparametric functions f^j and the (potentially varying) coefficients a^{jk}, $j, k = 1, \ldots, M-1$, consists of iteratively solving minimization problems, where the iteration is necessary only for the nonparametric part of the model. We use kernel smoothing for the nonparametric part, and least squares for the parametric coefficients in case of the fixed-coefficients model (2.1). Again, to obtain estimates that fulfill the condition of Slutsky symmetry, (linearly) restricted least squares are used for the parametric part[3].

Keeping the dependence on $\tilde{x}$, we may approximate

$$\mathbf{f}(t) \quad \approx \quad \mathbf{f}(\tilde{x}) + \nabla_{\tilde{x}}\mathbf{f}(\tilde{x})(t - \tilde{x}) \;\; \approx \;\; \boldsymbol{\alpha}(\tilde{x}) + \boldsymbol{\beta}(\tilde{x})(t - \tilde{x}), \tag{2.8}$$

where $\boldsymbol{\alpha}(\tilde{x})$ and $\boldsymbol{\beta}(\tilde{x})$ are the local level and derivative of $\mathbf{f}(t)$. Then, for the partial linear model the local problem is

$$\begin{aligned} & \min_{\boldsymbol{\alpha}(\tilde{x}),\boldsymbol{\beta}(\tilde{x}),\mathbf{A}} \sum_{i=1}^{N} \mathbf{e}_i^\top \boldsymbol{\Omega} \mathbf{e}_i, \text{ with} \\ \mathbf{e}_i \quad \equiv \quad & \mathbf{w}_i - \frac{\boldsymbol{\alpha}(\tilde{x}) + (\tilde{x}_i - \tilde{x})\,\boldsymbol{\beta}(\tilde{x}) + \mathbf{A}\tilde{\mathbf{p}}_i}{1 - \boldsymbol{\beta}(\tilde{x})^\top \tilde{\mathbf{p}}_i}, \end{aligned}$$

where $\boldsymbol{\Omega}$ is an $(M-1) \times (M-1)$ weighting matrix.

[3]More details can be found in Section 2.7.

Similarly, for the varying-coefficients model (2.7), the local problem in the neighbourhood of each given $\tilde{x}$ is

$$\min_{\boldsymbol{\alpha}(\tilde{x}),\boldsymbol{\beta}(\tilde{x}),\ \boldsymbol{\Gamma}(\tilde{x}),\boldsymbol{\Delta}(\tilde{x})} \sum_{i=1}^{N} \mathbf{e}_i^\top \boldsymbol{\Omega} \mathbf{e}_i, \text{ with}$$

$$\mathbf{e}_i \equiv \mathbf{w}_i - \frac{\boldsymbol{\alpha}(\tilde{x}) + (\tilde{x}_i - \tilde{x})\,\boldsymbol{\beta}(\tilde{x}) + \boldsymbol{\Gamma}(\tilde{x})\tilde{\mathbf{p}}_i + (\tilde{x}_i - \tilde{x})\,\boldsymbol{\Delta}(\tilde{x})\tilde{\mathbf{p}}_i}{1 - \boldsymbol{\beta}(\tilde{x})^\top \tilde{\mathbf{p}}_i - \frac{1}{2}\tilde{\mathbf{p}}_i^\top \boldsymbol{\Delta}(\tilde{x})\tilde{\mathbf{p}}_i},$$

where $\boldsymbol{\Omega}$ is now a different $(M-1)\times(M-1)$ weighting matrix and $\boldsymbol{\Gamma}(\tilde{x})$ and $\boldsymbol{\Delta}(\tilde{x})$ are the local level and derivative, respectively, of the price coefficients.

Here, the imposition of homogeneity is done via the use of normalised prices and expenditures (i. e. $\tilde{x}$ instead of x etc.). The imposition of Slutsky symmetry is achieved by the restriction $\mathbf{A} = \mathbf{A}^\top$, or in the varying-coefficients case by $\mathbf{A}\,(x) = \mathbf{A}\,(x)^\top$, i. e. by restricting $\boldsymbol{\Gamma}(\tilde{x}) = \boldsymbol{\Gamma}(\tilde{x})^\top$ and $\boldsymbol{\Delta}(\tilde{x}) = \boldsymbol{\Delta}(\tilde{x})^\top$. This local linear approach could easily be extended to higher order local polynomials, but for this we would need stronger assumptions on the data and the model.

2.3.2 The Estimation Algorithm

Denote $\Delta_i = \tilde{X}_i - \tilde{x}$, $K_i = K((\tilde{X}_i - \tilde{x})/h)/h$, where K is some symmetric kernel function with the usual properties and h a bandwidth that controls the smoothness of the estimate. We omit an extra subscript h in K_i for the sake of notation.

Let us start with the minimization problem for the partial linear model (2.1). As above, the α^j are related to the functions f^j at point $\tilde{x}$ and the parameters β^j to its first derivatives, while the parameters a^{jk} are fixed for all $\tilde{x}$:

$$\min_{\alpha^j,\beta^j} \sum_{j=1}^{M-1}\sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + \sum_{k=1}^{M-1} a^{jk}\tilde{P}_i^k}{1 - \sum_{k=1}^{M-1} \beta^k \tilde{P}_i^k} \right)^2 K_i. \tag{2.9}$$

In order to minimize, we set the first derivative equal to zero. Taking the derivative of (2.9) with respect to α^j, and using the notations $S_i = 1 - \sum_{k=1}^{M-1} \beta^k \tilde{P}_i^k$ and $T_i^j = \sum_{k=1}^{M-1} a^{jk}\tilde{P}_i^k$, we solve

$$0 = \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i\beta^j + T_i^j}{S_i} \right) \frac{K_i}{S_i}. \tag{2.10}$$

This gives immediately (for $j = 1, \ldots, M-1$)

$$\alpha^j = \left[\sum_{i=1}^{N} W_i^j K_i / S_i - \beta^j \sum_{i=1}^{N} K_i \Delta_i / S_i^2 - \sum_{i=1}^{N} K_i T_i^j / S_i^2 \right] \left[\sum_{i=1}^{N} K_i / S_i^2 \right]^{-1}. \tag{2.11}$$

On the other hand, by differentiating (2.9) with respect to β^j (again for $j = 1, \ldots, M-1$), we get the equations

$$\begin{aligned} 0 \;=\; & \sum_{i=1}^{N} \left(W_i^1 - \frac{\alpha^1 + \Delta_i \beta^1 + T_i^1}{S_i} \right) K_i \cdot \frac{(\alpha^1 + \Delta_i \beta^1 + T_i^1)\tilde{P}_i^j}{S_i^2} + \cdots + \\ & \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + T_i^j}{S_i} \right) K_i \cdot \frac{\Delta_i S_i + (\alpha^j + \Delta_i \beta^j + T_i^j)\tilde{P}_i^j}{S_i^2} + \cdots + \\ & \sum_{i=1}^{N} \left(W_i^{M-1} - \frac{\alpha^{M-1} + \Delta_i \beta^{M-1} + T_i^{M-1}}{S_i} \right) K_i \frac{(\alpha^{M-1} + \Delta_i \beta^{M-1} + T_i^{M-1})\tilde{P}_i^j}{S_i^2}. \end{aligned}$$

This is equivalent to

$$\begin{aligned} 0 \;=\; & \sum_{k=1}^{M-1} \sum_{i=1}^{N} \left(W_i^k - \frac{\alpha^k + \Delta_i \beta^k + T_i^k}{S_i} \right) K_i \cdot \frac{(\alpha^k + \Delta_i \beta^k + T_i^k)\tilde{P}_i^j}{S_i^2} + \\ & \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + T_i^j}{S_i} \right) K_i \frac{\Delta_i}{S_i}. \end{aligned} \tag{2.12}$$

Certainly, we can not solve equation (2.12) analytically for β^j. But, for our iterative purpose it is enough to consider the following implicit representation:

$$\begin{aligned} \beta^j \;=\; & \Bigg[\sum_{k=1}^{M-1} \sum_{i=1}^{N} \left(W_i^k - \frac{\alpha^k + \Delta_i \beta^k + T_i^k}{S_i} \right) K_i \cdot \frac{(\alpha^k + \Delta_i \beta^k + T_i^k)\tilde{P}_i^j}{S_i^2} + \\ & \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + T_i^j}{S_i} \right) K_i \frac{\Delta_i}{S_i} \Bigg] \Bigg/ \sum_{i=1}^{N} \frac{K_i \Delta_i^2}{S_i^2}. \end{aligned} \tag{2.13}$$

We use the implicit representation (2.13) to calculate new values for β^j. With them we get new S_i, so that we can find new α^j:

$$\begin{aligned} \beta^j_{new} \;=\; & \Bigg[\sum_{k=1}^{M-1} \sum_{i=1}^{N} \left(W_i^k - \frac{\alpha^k_{old} + \Delta_i \beta^k_{old} + T^k_{i,old}}{S_{i,old}} \right) K_i \frac{(\alpha^k_{old} + \Delta_i \beta^k_{old} + T^k_{i,old})\tilde{P}_i^j}{S^2_{i,old}} \\ & + \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j_{old} + T^j_{i,old}}{S_{i,old}} \right) K_i \frac{\Delta_i}{S_{i,old}} \Bigg] \Bigg/ \sum_{i=1}^{N} \frac{K_i \Delta_i^2}{S^2_{i,old}}, \end{aligned} \tag{2.14}$$

$$S_{i,new} = 1 - \sum_{k=1}^{M-1} \beta^k_{new} \tilde{P}_i^k, \text{ and} \tag{2.15}$$

$$\alpha^j_{new} = \frac{\sum_{i=1}^{N} W_i^j K_i / S_{i,new} - \beta^j_{new} \sum_{i=1}^{N} K_i \Delta_i / S^2_{i,new} - \sum_{i=1}^{N} K_i T^j_{i,old} / S^2_{i,new}}{\sum_{i=1}^{N} K_i / S^2_{i,new}}.$$

We repeat these steps until convergence. The optimal **A** will be the symmetric matrix that minimizes the least squares problem. In practice, at the end of each iteration step, we solve the restricted least squares problem resulting from equation (2.6). With some algebra, the problem is given by

$$W_i^j \cdot (1 - \sum_{k=1}^{M-1} \beta_i^k \tilde{P}_i^k) - \alpha_i^j = \sum_{k=1}^{M-1} a^{jk} \tilde{P}_i^k. \tag{2.16}$$

The modification of the algorithm to take the varying coefficients $\mathbf{A}(\tilde{x})$ into account is carried out along ideas of Fan and Zhang (1999), though it is more complex in our context. With the same local linear approximation arguments as above, we get the local problem in the neighbourhood of $\tilde{x}$ as

$$\min_{\boldsymbol{\theta}} \sum_{j=1}^{M-1} \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + \sum\limits_{k=1}^{M-1} (\gamma^{jk} + \Delta_i \delta^{jk}) \tilde{P}_i^k}{1 - \sum\limits_{k=1}^{M-1} \beta^k \tilde{P}_i^k - \frac{1}{2} \sum\limits_{k=1}^{M-1} \sum\limits_{l=1}^{M-1} \delta^{kl} \tilde{P}_i^k \tilde{P}_i^l} \right)^2 K_i, \tag{2.17}$$

with $\boldsymbol{\theta}$ denoting $\alpha^j, \beta^j, \gamma^{jk}$ and δ^{jk}. Note that γ^{jk} and δ^{jk} are symmetric since we consider a symmetric matrix of functions $a^{kl}(\tilde{x})$. The minimization of (2.17) in the usual way gives the extended algorithm in analogy to the first step of 2.3.2. For α^j and β^j we proceed as before but with $S_i = 1 - \sum \beta^k \tilde{P}_i^k - 1/2 \sum \sum \delta^{kl} \tilde{P}_i^k \tilde{P}_i^l$ and $T_i^j = \sum (\gamma^{jk} + \Delta_i \delta^{jk}) \tilde{P}_i^k$. Furthermore, we obtain

$$\gamma^{st} = \frac{\sum_{i=1}^N \left[\left(W_i^s - \frac{C_i^s}{S_i} \right) \tilde{P}_i^t + \left(W_i^t - \frac{C_i^t}{S_i} \right) \tilde{P}_i^s \, \mathbb{1}_{s \neq t} \right] \frac{K_i}{S_i}}{\sum_{i=1}^N \left[(\tilde{P}_i^t)^2 + (\tilde{P}_i^s)^2 \mathbb{1}_{s \neq t} \right] \frac{K_i}{S_i^2}},$$

with $C_i^s = \alpha^s + \Delta_i \beta^s + T_i^s - \gamma^{st} \tilde{P}_i^t$ and, defining $T_i^{s,-t} = T_i^s - \Delta_i \delta^{st} \tilde{P}_i^t$,

$$\begin{aligned} \delta^{st} = & \left[\sum_{k=1}^{M-1} \sum_{i=1}^{N} \left(W_i^k - \frac{\alpha^k + \Delta_i \beta^k + T_i^k}{S_i} \right) \frac{K_i}{S_i^2} (\alpha^k + \Delta_i \beta^k + T_i^k) \tilde{P}_i^t \tilde{P}_i^s + \right. \\ & \left. \sum_{i=1}^{N} \left\{ \left(W_i^s - \frac{\alpha^s + \Delta_i \beta^s + T_i^{s,-t}}{S_i} \right) \tilde{P}_i^t + \left(W_i^t - \frac{\alpha^t + \Delta_i \beta^t + T_i^{t,-s}}{S_i} \right) \tilde{P}_i^s \right\} \frac{K_i \Delta_i}{S_i} \right] \\ & \times \left[\sum_{i=1}^{N} \left\{ (\tilde{P}_i^t)^2 + (\tilde{P}_i^s)^2 \mathbb{1}_{s \neq t} \right\} \frac{\Delta_i^2 K_i}{S_i^2} \right]^{-1} . \end{aligned}$$

2.3.3 Bootstrap Inference

The *wild bootstrap* draws bootstrap responses based on the estimated model (2.1) with given sample $\{W_i, \tilde{X}_i, \tilde{P}_i\}_{i=1}^N$ and estimates $\hat{\alpha}^j$, $\hat{\beta}^j$ and $\hat{a}^{jk}$, $k, j = 1, \ldots, M-1$. Denote a prior bandwidth g with $O(g) > O(h)$ (obeying the needs of asymptotic theory, cf. Härdle and Marron (1991)), and let h be the bandwidth giving the desired smoothness in the original sample. The basic idea is now to use the estimated residuals from an estimate with bandwidth g,

$$\hat{\varepsilon}_i^j = W_i^j - \frac{\hat{\alpha}^j(\tilde{X}_i) + \sum_{k=1}^{M-1} \hat{a}^{jk} \tilde{P}_i^k}{1 - \sum_{k=1}^{M-1} \hat{\beta}^k(\tilde{X}_i) \tilde{P}_i^k}, \tag{2.18}$$

to get wild bootstrap residuals ε_i^{j*}. Given them we create the bootstrap samples $\{W_i^*, \tilde{X}_i, \tilde{P}_i\}_{i=1}^N$ by

$$W_i^{j*} = \frac{\hat{\alpha}^j(\tilde{X}_i) + \sum_{k=1}^{M-1} \hat{a}^{jk} \tilde{P}_i^k}{1 - \sum_{k=1}^{M-1} \hat{\beta}^k(\tilde{X}_i) \tilde{P}_i^k} + \varepsilon_i^{j*}, \tag{2.19}$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, M-1$. Here, ε_i^{j*} are bootstrap residuals that replicate desired properties of the distribution(s) of $\hat{\varepsilon}_i^j$. The W_i^{M*} are generated using the adding-up restriction $\sum_{j=1}^M W_i^{j*} = 1$. Repeating this many times, we get estimates (for $\mathbf{f}$ and $\mathbf{A}$) for each bootstrap sample and can use the bootstrap quantiles to construct pointwise confidence bands for the estimates.

There exists several strategies to obtain bootstrap residuals ε_i^{j*}. Typically, when no restriction is faced, one may use $\varepsilon_i^{j*} = u_i \cdot \hat{\varepsilon}_i^j$, where u_i is a standard normal random scalar. Under the additional assumption of homoscedasticity, this can even be simplified to $\varepsilon_i^{j*} = u_i \cdot \hat{\sigma}_\varepsilon^j$, where σ_ε^j is estimated from the residuals (2.18).

In our case, one could argue that such bootstrap errors could cause the bootstrap values of W_i^{j*} to lie outside the admissible range of $[0,1]$ for budget shares. On the one hand, this may not matter because the estimation algorithm does not control the constraint that $\hat{W}_i^j \in [0,1]$. However, given that actual expenditure shares are bounded, the bootstrap residuals may poorly reflect the true error distribution and misrepresent the confidence intervals, for example putting them outside $[0,1]$.

To address the possibility that inference is hampered by bootstrap budget-shares lying outside the interval $[0,1]$, we introduce an alternative formulation of the wild bootstrap. Because there are many expenditure shares, the main bounding problem is the lower bound at 0, and this is the problem we deal with. Thus, we are faced with a conditionally asymmetric (to the right) error distribution. We thus consider an asymmetric distribution for ε_i^{j*} given $\hat{\varepsilon}_i^j$ as follows. Generate bootstrap errors via

$$\frac{\chi_k^2}{\sqrt{k}} \cdot \frac{|\hat{\varepsilon}_i^j|}{\sqrt{2}} - \frac{|\hat{\varepsilon}_i^j|}{\sqrt{2}} \cdot \sqrt{k} \leq |\hat{W}_i^j|, \tag{2.20}$$

where $k \leq \lfloor (W_i^j / \hat{\varepsilon}_i^j)^2 \cdot 2 \rfloor$. In the case that k is less than one we draw the bootstrap residual ε_i^{j*} from $\chi_1^2 \cdot |\hat{W}_i^j| - |\hat{W}_i^j|$. Note that this fulfills $E[\varepsilon_i^{j*}] = 0$ and $E[(\varepsilon_i^{j*})^2] =$

$E[(\hat{\varepsilon}_i^j)^2]$ for all i and j. From (2.20) we get that for positive $\hat{W}_i^j$ the bootstrap analog is also always positive. This method leads automatically to confidence bands that lie almost fully inside the interval $[0,1]$ and are consequently narrower than those based on a simple normal bootstrap described above. In the simulation study below, we present empirical evidence that the above introduced asymmetric bootstrap is accurate.

2.3.4 Practical Considerations

One issue in such iterative procedures is the question of adequate initial values for the nonparametric part. Here we have a convenient model feature to exploit: when we normalise prices in the sample such that $\tilde{\mathbf{p}}_i = (0,\dots,0)$ for some group of consumers, say N_0, equation (2.9) reduces to the well-known local linear case. Since for the denominator term we have $S_i = 1$, we get the objective function

$$\min_{\alpha^j,\beta^j} \sum_{j=1}^{M-1} \sum_{i=1}^{N_0} \left(W_i^j - \alpha^j + \Delta_i \beta^j \right)^2 K_i. \tag{2.21}$$

Solving this problem on the sample of consumers where $\tilde{\mathbf{p}}_i = (0,\dots,0)$ gives us consistent estimates (though with a possibly large variance depending on the subsample size N_0) which can be used as starting values for α^j and β^j. For the varying-coefficients model we also need starting values for the γ^{jk} and δ^{jk}. As a natural choice for the γ^{jk} we use the results of the algorithm in Section 2.3.2 and zero for all δ^{jk} (i. e. starting in the first iteration with a simpler model).

For the bandwidth choice we recommend the use of the same bandwidth h for all expenditure categories because: (a) the functions refer to the same expenditure data in all equations; and (b) the economic theory does not suggest that the shares of some goods would be smoother than those of others. Plug-in bandwidths could be derived from the asymptotics of $\widehat{\mathbf{f}}$, or, alternatively, one could construct a risk estimate similar to cross-validation but this time jointly for all elements of $\widehat{\mathbf{f}}$. It is clear that the first possibility depends on derivatives of the unknown $\widehat{\mathbf{f}}$ and the density of expenditures, while the second approach would be computationally costly. A rough idea of a bandwidth to start with may be derived from (2.21). Run a leave-one-out cross-validation for (2.21) with the subsample of individuals fulfilling $\tilde{\mathbf{p}}_i = (0,\dots,0)$, and correct the obtained bandwidth h_0 for the size of the full sample, i. e. $h = N^{-1/5} h_0 N_0^{1/5}$.

In the fixed-coefficients model (2.1) we recommend running the estimation algorithm twice: first with an undersmoothing bandwidth in the nonparametric part to keep the possible smoothing bias small. The resulting estimate for the coefficient matrix $\mathbf{A}$ is saved and used in the second run, which uses a larger bandwidth for the nonparametric part to get reasonably smooth $\widehat{\mathbf{f}}$. This is unnecessary in the varying-coefficients model (2.3), where we estimate only nonparametric functions.

Recall that the M^{th} equation and its Engel curve is simply a result of the homogeneity and the adding-up condition $\sum_{j=1}^{M} W_i^j = 1$. In practice one might choose the item with the least variation in expenditure shares across the households.

2.4 Empirical Analysis

2.4.1 A Simulation Study

First, to produce some artificial data we generated 33 distinct price vectors, normally distributed in each dimension, for each of 6 expenditure categories (i. e. we have 6 items with different prices in 33 regions). As in typically observed micro-data, we did not allow for a wide price variety, see Lewbel (2000). Summary values for these price vectors can be found in Table 2.1.

Table 2.1: *Summary of price vectors used in the simulation.*

	1	2	3	4	5
Min	3.905	3.449	3.763	0.880	2.794
Max	4.130	3.585	3.919	1.121	3.010
Mean	4.018	3.517	3.841	1.002	2.901
Std.	0.030	0.020	0.020	0.030	0.030

For 32 regions (i. e. price vectors) we uniformly draw 30 log–total expenditure values from the interval $[1, 2]$. For the reference region (number 33) we draw 40 uniformly distributed values between one and two. In total this yields $N = 1000$ observations. These are used to generate expenditure shares using the expenditure functions shown in Figure 2.1 (solid lines), price parameters given in Table 2.2, and normal error terms with mean zero and standard deviation 0.01. In order to get shares which fulfill the conditions $W^j \in [0, 1]$ and $\sum_j W^j = 1$ we applied the rejection method (we dropped and replaced values outside $[0, 1]$).

Next, we estimate the functions α^j and the price parameters using our estimation algorithm introduced in Section 2.3.2. This is repeated 250 times (using the same functions, price parameters, and range of log–total expenditure values) to get an idea

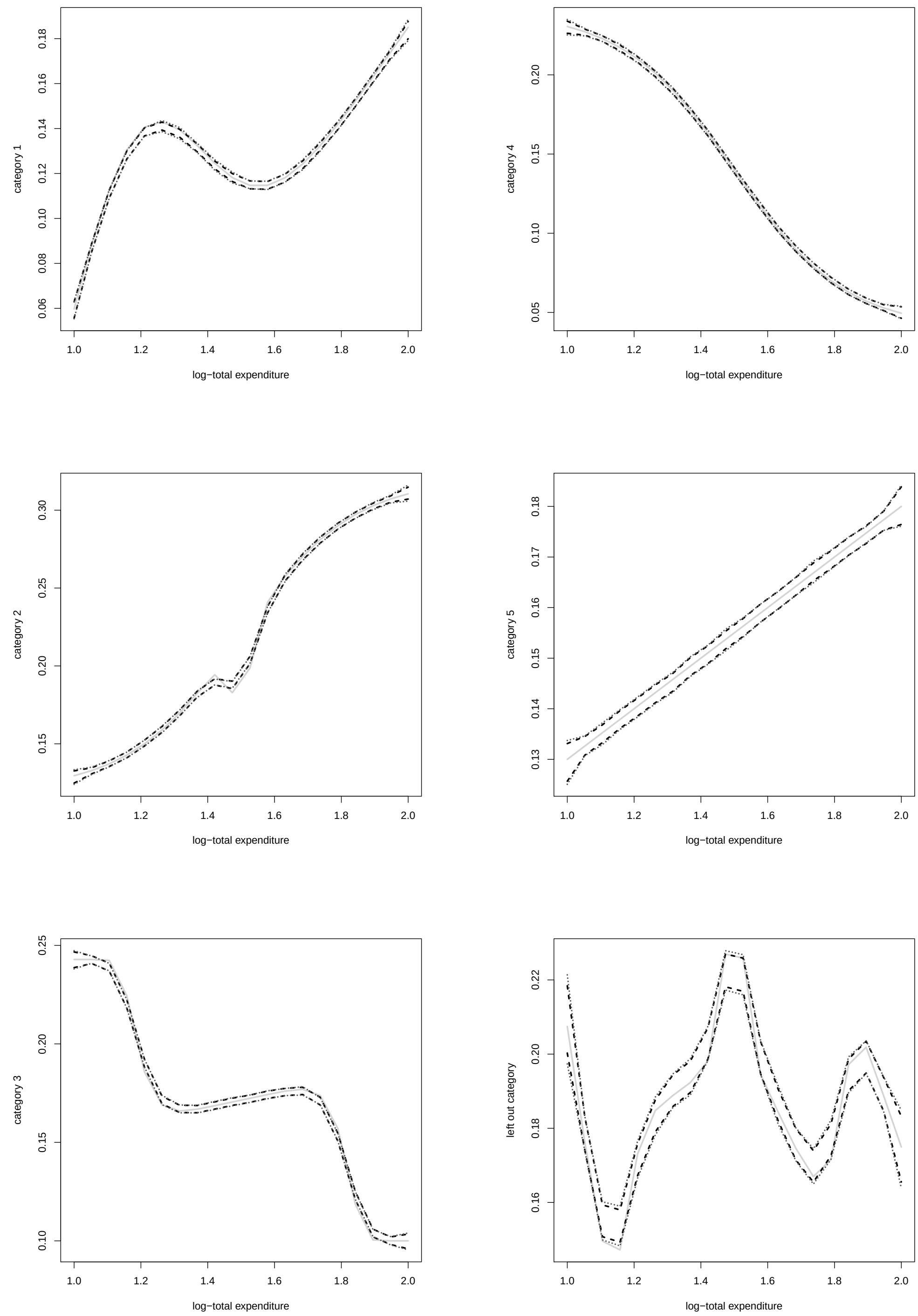


Figure 2.1: *Simulation of 6 different budget share functions (solid line) with 90% coverage probability (dotted lines) and asymptotic 90% bootstrap confidence intervals (dashed lines)*

Table 2.2: *Price parameters used in the simulation.*

	1	2	3	4	5
1	−0.150	−0.100	0.150	0.100	0.280
2		0.250	0.100	−0.250	0.170
3			0.320	−0.220	−0.190
4				−0.200	0.150
5					−0.180

of the mean squared errors of our estimators. For estimating $\mathbf{f}$ we used the Gaussian kernel with $h \approx 0.034$, the smallest bandwidth giving smooth estimates.

In Figure 2.1 we have plotted the true functions (solid lines) together with intervals of 90% coverage probabilities for the estimates (dotted lines) as a result of the 250 simulation runs. On the one hand, we see pretty narrow bands which accurately capture even those functions with flat plateaus in the intermediate range (category 3) and with bumps (category 2). Such functions are often hard to estimate in practise. We also see the limits of the method as for example boundary effects. Our smoother can estimate without any bias the linear function. In Table 2.3 we give the estimated parameter means, together with the standard deviations. The exactness of our simulation is demonstrated by the small total MSE of only $6.83 \cdot 10^{-6}$.

Table 2.3: *Estimated price parameters and standard deviations (in brackets).*

	1	2	3	4	5
1	−0.1515	−0.1006	0.1494	0.0997	0.2799
	(0.0140)	(0.0106)	(0.0103)	(0.0090)	(0.0087)
2		0.2494	0.0999	−0.2497	0.1699
		(0.0176)	(0.0129)	(0.0102)	(0.0102)
3			0.3200	−0.2208	−0.1891
			(0.0170)	(0.0096)	(0.0100)
4				−0.1992	0.1490
				(0.0117)	(0.0079)
5					−0.1803
					(0.0117)

To verify the functioning of our new bootstrap procedure we constructed 200 bootstrap samples (with $g = h$ as N is relatively small) along Section 2.3.3 for each of 100 simulation runs. We calculated 90% bootstrap confidence intervals around the estimator for each simulation. The mean of upper and lower bounds of these intervals are given in Figure 2.1 (dashed lines). The fact that the 90% coverage probability intervals and the means of the 90% bootstrap intervals almost coincide indicates that our bootstrap procedure is acceptably accurate.

2.4.2 Analysing Household Expenditures in Canada

In our empirical study we use the same Canadian data as in Lewbel and Pendakur (2009) and Pendakur and Sperlich (2010) which come from public sources (see also Pendakur (2002)). The price and expenditure data are available for 12 years in 5 regions: Atlantic, Quebec, Ontario, Prairies and British Columbia. This yields 60 distinct price vectors, where prices are normalised in a way that all prices of the categories from Ontario in 1986 are one, i. e. $\tilde{\mathbf{p}}_{O,86} = (0,\ldots,0)$. These 189 observations define the base price vector and we use them to get the starting values. Note further, to achieve homogeneity we subtracted p^M, the price of the left–out expenditure category, from all other prices and total expenditure.

We use 6952 observations of rental–tenure unattached individuals aged between 25 and 64 with no dependants to minimise demographic variation in preferences. Our analysis includes annual total–expenditure in nine categories: food–in, food–out, rent, clothing, household operations, household furnishing and equipment, private transportation, public transportation and personal care. The left–out category is personal care. We get thus a system of eight expenditure share equations which depend on eight (normalised) log–prices and (normalised) log–total expenditure. These expenditure categories account for about three-quarters of the current consumption of the households in the sample. Summary statistics of the observations are given in Table 2.4.

We note that this choice of commodities is arbitrary: one could divide these goods into subcategories, or aggregate them up into larger categories. We choose these categories because they offer the finest gradation consistent with largest possible time span for the price data (finer gradations of price data are available, but for shorter periods of time). Another advantage of this choice of commodities is that they are directly comparable with Pendakur and Sperlich (2010).

As noted above, when $\tilde{\mathbf{p}} = (0,\ldots,0)$ (as for the observations in Ontario 1986), the price effects in expenditure shares amount to zero, yielding

$$\mathbf{w}(\mathbf{p},x) = \mathbf{w}(\tilde{\mathbf{p}},\tilde{x}) = \mathbf{f}(\tilde{x}) = \mathbf{f}(x),$$

Table 2.4: *Canadian household expenditure data.*

		Min	Max	Mean	Std.
expenditure shares	food–in	0.00	0.63	0.17	0.09
	food–out	0.00	0.64	0.08	0.08
	rent	0.01	0.95	0.40	0.13
	operations	0.00	0.64	0.08	0.04
	furnishing	0.00	0.65	0.04	0.06
	clothing	0.00	0.53	0.09	0.06
	private trans	0.00	0.59	0.08	0.09
	public trans	0.00	0.34	0.04	0.04
log–prices	food–in	−0.39	0.07	−0.03	0.09
	food–out	−0.42	0.25	0.05	0.12
	rent	−0.35	0.14	−0.12	0.15
	operations	−0.28	0.10	−0.04	0.08
	furnishing	−0.16	0.21	−0.03	0.09
	clothing	−0.07	0.44	0.10	0.11
	private trans	−0.51	0.30	−0.09	0.18
	public trans	−0.59	0.40	0.01	0.25
log–total expenditure		3.03	6.26	4.61	0.45

which we will refer to as the vector of Engel curves. The estimated Engel curves of all expenditure categories can be found in Figures 2.2 and 2.3 as solid lines, where the horizontal axes refer to $\tilde{x}$, i. e. the log total expenditures minus p^M.

We include pointwise 90% confidence intervals which we calculated, as described in Section 2.3.3, with heteroscedastic error terms and 500 bootstrap iterations using our new conditionally asymmetric wild bootstrap procedure. To generate the bootstrap samples we used $g = h$ with $h = 0.17$. This is somewhat larger than $h_0(N_0/N)^{1/5}$ which would give wiggly estimates $\widehat{\mathbf{f}}$. The estimation algorithm was implemented with the Gaussian kernel, and it converged in our setting after about 15 iterations. In all figures, the resulting Engel curves are compared to those of Banks, Blundell, and Lewbel (1997) and those of Pendakur and Sperlich (2010) (assuming a partial linear cost function with Slutsky symmetry). In terms of local regularity, the estimated values of budget-shares lie entirely within the interval $[0, 1]$. Although we do not assess the global regularity of our estimates or estimator, it is notable that the estimated budget shares satisfy this condition locally.

Food-at-home and food-out are strong necessities and luxuries, respectively, with

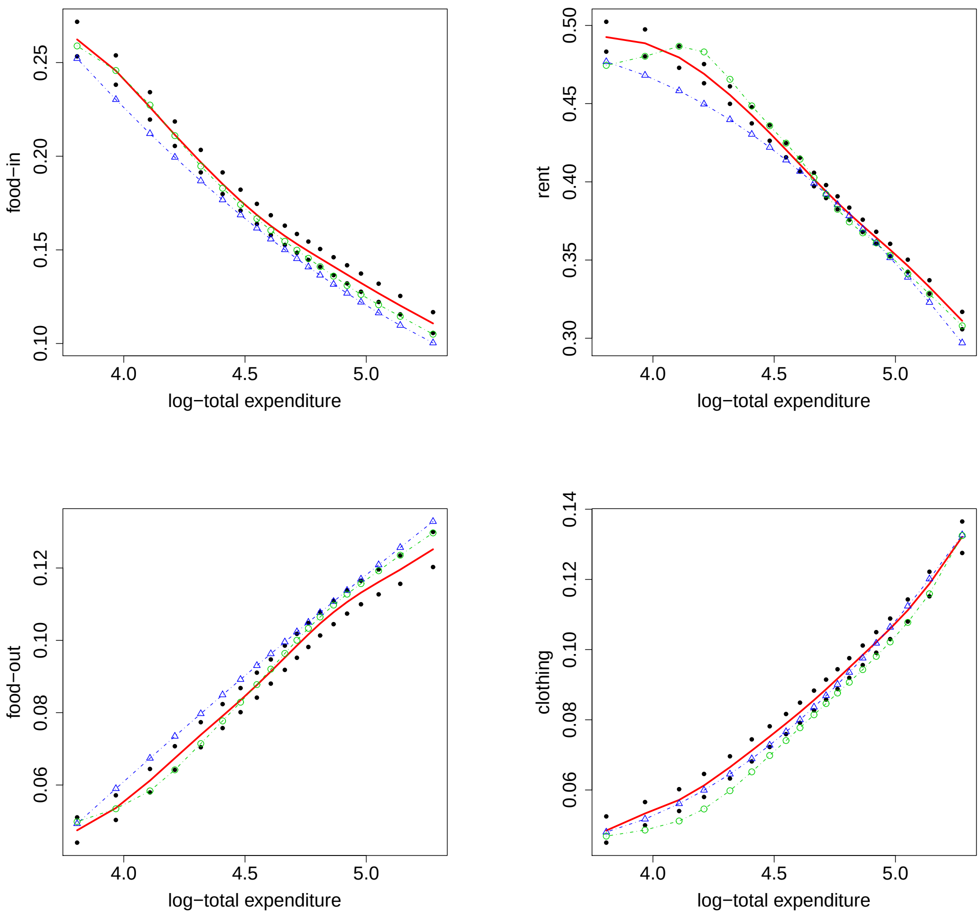


Figure 2.2: *Estimates of food-in, food-out, rent, and clothing (solid line) with 90% pointwise confidence bands (dotted), together with estimates using Banks, Blundell, and Lewbel (1997) (triangles) and Pendakur and Sperlich (2010) (circles).*

nearly linear Engel curves in both cases. The near-linearity of these Engel curves has been observed in a large number of empirical investigations, including Banks, Blundell, and Lewbel (1997). Some curvature is observed in the rent and clothing equations, especially near the bottom of the distribution. This curvature is noted in semiparametric work, such as Pendakur and Sperlich (2010) and Lewbel and Pendakur (2009). The most curvature is noted in smaller budget shares like household operation, private transportation and public transportation. The curvature in household operation seems quite strong, and that in private transportation seems decidedly non-quadratic. In the figures one can see that most of the expenditure-share equations are very similar between the present approach and the partially linear cost function approach of Pendakur and Sperlich (2010). However, two exceptions are the household-operation and public-transportation equations. These are estimated about

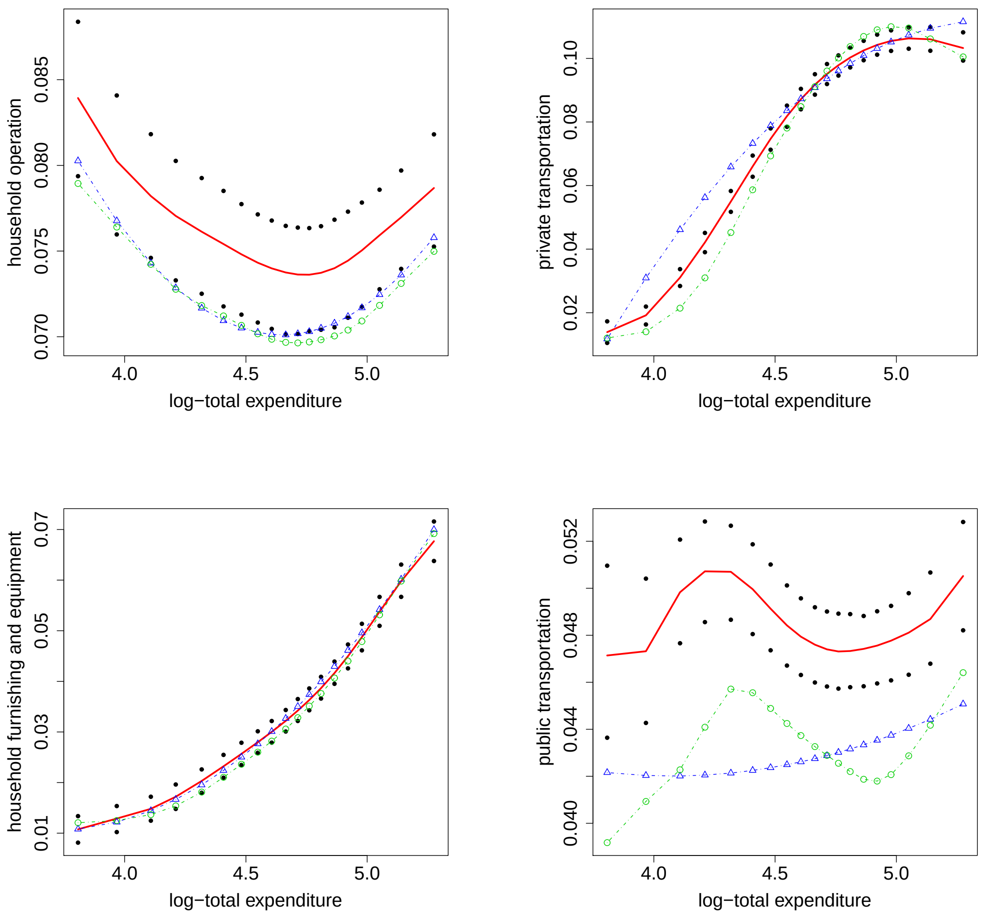


Figure 2.3: *Estimates of household operations, furnishing and equipment, private and public transportation (solid line) with 90% pointwise confidence bands (dashed), together with estimates using Banks, Blundell, and Lewbel (1997) (triangles) and Pendakur and Sperlich (2010) (circles).*

0.5 percentage points higher in the present approach. We note that this variation between estimated models is not overwhelmingly large. For example, in some expenditure shares equations Pendakur and Sperlich (2010) report a difference of more than 0.5 percentage points between symmetry-restricted and symmetry-unrestricted estimates of their partially linear cost function model. This highlights the fact that although the two models are similar in spirit, they are not identical in practise.

Table 2.5 gives the estimated symmetric price parameters and in brackets the bootstrapped standard deviations[4]. These estimated price effects are in the plausible range, and are similar to those found in Pendakur and Sperlich (2010).

[4]Note that we do not display estimated price elasticities in Tables 2.3 and 2.5 but only estimated

Table 2.5: *Estimated symmetric price effects a^{jk} (with bootstrap standard deviations in brackets).*

	food–in	food–out	rent	oper	furn	clothing	priv tr	pub tr
food–in	−0.026	0.013	−0.006	−0.008	0.009	0.006	0.037	−0.058
	(0.036)	(0.018)	(0.012)	(0.019)	(0.014)	(0.015)	(0.007)	(0.006)
food–out		−0.035	0.047	0.012	−0.002	−0.069	0.001	−0.045
		(0.014)	(0.007)	(0.012)	(0.010)	(0.009)	(0.005)	(0.005)
rent			0.186	0.023	−0.026	−0.021	−0.036	0.087
			(0.017)	(0.007)	(0.005)	(0.008)	(0.006)	(0.005)
oper				0.040	0.010	−0.016	−0.029	0.023
				(0.017)	(0.011)	(0.011)	(0.004)	(0.005)
furn					−0.038	0.026	−0.017	−0.024
					(0.016)	(0.009)	(0.004)	(0.004)
clothing						0.005	−0.002	−0.014
						(0.010)	(0.005)	(0.004)
priv tr							0.002	0.006
							(0.006)	(0.003)
pub tr								−0.011
								(0.003)

Thus, the estimated Engel curves are plausible and have some evidence of complexity beyond the quadratic form of Banks, Blundell, and Lewbel (1997). Compared to Pendakur and Sperlich (2010), the present approach has an important computational advantage: it is based entirely on observed regressors, and does not require any numerical inversions to generate a latent regressor. In comparison with Lewbel and Pendakur (2009), the present approach has an important interpretational difference: whereas Lewbel and Pendakur (2009) must interpret model error terms as unobserved preference heterogeneity parameters, the present approach is based on the more standard view of error terms as measurement or other non-behavioural error.

The varying-coefficients extension is similarly easy to implement. The estimated Engel curves are almost identical to those found in the fixed-coefficients case, with some deviations in the tails. Depending on the bandwidth, the estimated price parameters evaluated at median log-expenditure are statistically indistinguishable from those of the fixed-coefficients model, but their estimated variance is much greater. In particular, we got approximately twice the standard errors for estimated parameters evaluated at median expenditures relative to their fixed-coefficients counterparts. We found that the varying-coefficients estimates of $\mathbf{f}$ were very similar to the fixed-

parameters which can also be positive in the diagonal.

coefficient estimates, and so we do not present them here.

2.5 Conclusions

We propose an indirect utility model which is nonparametric in the expenditure direction and parametric (with fixed- or varying-coefficients) in the price directions. The utility function implies a consumer demand system that has parametric log–price effects and nonparametric log–total expenditure effects. We avoid the *curse of dimensionality*, typically associated in fully nonparametric estimation of consumer demand, since the nonparametric part of the model is only one dimensional. The model is easily restricted to satisfy the rationality conditions of homogeneity and Slutsky symmetry.

We provide a new wild bootstrap procedure that allows for conditional asymmetries and guarantees positive shares. We show the finite sample performance of our estimators in a simulation study, and finally apply our method to Canadian expenditure data.

The application of this model to Canadian price and expenditure data shows not only the potential of the model but also suggests that some expenditure shares are more complex than the linear ones in popular parametric demand models. The simulation study reveals further that it is also possible to capture shapes which are difficult to estimate (cf. Hastie and Tibshirani (1984)), such as those with flat plateaus in the intermediate range or with bumps.

2.6 Appendix A: Slutsky Symmetry

The imposition of Slutsky symmetry requires that $a^{lk} = a^{kl}$ ($a^{lk}(x) = a^{kl}(x)$) for all k, l, or equivalently, that $\mathbf{A} = \mathbf{A}^\top$ ($\mathbf{A}(x) = \mathbf{A}(x)^\top$). To see this, start with the definition of the Slutsky matrix, whose elements are given by

$$s_{ij} = \frac{\partial h^i}{\partial b^j} = \frac{\partial g^i}{\partial y} \cdot q^j + \frac{\partial g^i}{\partial b^j}, \tag{2.22}$$

where h^i denotes the Hicksian demand, g^i the Marshallian demand, y total expenditure, b^j the price, and q^j the quantity of good/category j. With $q^i = g^i = w^i \cdot y/b^i$, where g^i and w^i are functions of $y, b^1, \ldots, b^M$, we get

$$\frac{\partial g^i}{\partial y} = \frac{\partial w^i}{\partial y} \cdot \frac{y}{b^i} + \frac{w^i}{b^i} \quad \text{and} \quad \frac{\partial g^i}{\partial b^j} = \frac{\partial w^i}{\partial b^j} \cdot \frac{y}{b^i}. \tag{2.23}$$

Using the equations in (2.22) and (2.23) we can write the difference of $s_{ij} - s_{ji}$, for $i, j = 1, \ldots, M-1$, as

$$s_{ij} - s_{ji} = \left(\frac{\partial w^i}{\partial y} \cdot w^j - \frac{\partial w^j}{\partial y} \cdot w^i \right) \frac{y^2}{b^i b^j} + \left(\frac{\partial w^i}{\partial b^j} \cdot \frac{y}{b^i} - \frac{\partial w^j}{\partial b^i} \cdot \frac{y}{b^j} \right). \tag{2.24}$$

With the abbreviations $T^i = \sum a^{ik} \tilde{p}^k$, $S = 1 - \sum \partial f^k / \partial \tilde{x} \cdot \tilde{p}^k$ and $f^i = f^i(\tilde{x})$, we can rewrite (2.6) in the following way

$$w^i = \frac{f^i + T^i}{S}, \tag{2.25}$$

but note that w^i depends on $\tilde{x} = \log y - \log b^M$, log-total expenditure, and $\tilde{p}^j = \log b^j - \log b^M$ the log-prices for $j = 1, \ldots, M-1$. Now we can differentiate (2.25) w.r.t. total expenditure and the j-th price, and obtain with $U = \sum \partial^2 f^k / \partial \tilde{x}^2 \cdot \tilde{p}^k$

$$\frac{\partial w^i}{\partial y} = \frac{\frac{\partial f^i}{\partial \tilde{x}} \cdot S + (f^i + T^i) \cdot U}{y S^2} \quad \text{and} \quad \frac{\partial w^i}{\partial b^j} = \frac{a^{ij} S + (f^i + T^i) \cdot \frac{\partial f^j}{\partial \tilde{x}}}{b^j S^2}.$$

Plugging-in these results and equation (2.25) in (2.24), we get immediately that $s_{ij} - s_{ji} = 0$ if $a_{ij} = a_{ji}$.

2.7 Appendix B: Restricted Least Squares for a Symmetric Matrix

Recall that to estimate the symmetric parameters a^{jk}, $j, k = 1, \ldots, d$, $(d = M-1)$ we use equation (2.6) and get, with some algebra, for a single individual i

$$W_i^j \cdot (1 - \sum_{k=1}^{d} \beta_i^k \tilde{P}_i^k) - \alpha_i^j = \sum_{k=1}^{d} a^{jk} \tilde{P}_i^k. \tag{2.26}$$

Here, the parameters $\alpha_i^j = \alpha^j(\tilde{X}_i)$ are related to the functions f^j at the point $\tilde{X}_i$ and the parameters $\beta_i^j = \beta^j(\tilde{X}_i)$ to its first derivatives. Defining $(\boldsymbol{W})_{ij} := W_i^j \cdot (1 - \sum_{k=1}^{d} \beta_i^k \tilde{P}_i^k) - \alpha_i^j$, $(\boldsymbol{P})_{ik} := \tilde{P}_i^k$ and $(\boldsymbol{A})_{kj} := a^{kj}$ we can formulate equation (2.26) using matrix notation:

$$\boldsymbol{W} = \boldsymbol{P} \cdot \boldsymbol{A}, \tag{2.27}$$

where $\boldsymbol{W}$, $\boldsymbol{P}$ are $N \times d$ matrices and $\boldsymbol{A}$ a $d \times d$ symmetric matrix. Note that it is not necessary to start in the model description (2.1) with symmetric parameters a^{jk}. However, when we start with arbitrary parameters we will end nevertheless in (2.27) with a symmetric parameter matrix. More specific, we get for (2.26):

$$W_i^j \cdot (1 - \sum_{k=1}^{d} \beta_i^k \tilde{P}_i^k) - \alpha_i^j = 1/2 \sum_{k=1}^{d} (a^{jk} + a^{kj}) \tilde{P}_i^k$$

and in matrix notation

$$\boldsymbol{W} = \frac{1}{2}\,\boldsymbol{P}(\boldsymbol{A} + \boldsymbol{A}^\top).$$

Obviously, $\boldsymbol{A} + \boldsymbol{A}^\top$ is symmetric, and this fact would yield an identification problem for non-symmetric $\boldsymbol{A}$. In other words, even if symmetry is not required in the economic modelling process, only symmetry of $\boldsymbol{A}$ makes the estimation problem identifiable.

Next, to calculate the unknown matrix $\boldsymbol{A}$ in equation (2.27) we should not use the standard least square method but want directly make use of the symmetry of $\boldsymbol{A}$. Denote $w^{ij} = (\boldsymbol{W})_{ij}$ and w^j the j–th column of $\boldsymbol{W}$, $p^{ij} = (\boldsymbol{P})_{ij}$, p_i the i–th row and p^j the j–th column of the price–matrix $\boldsymbol{P}$ and a^j also the j–th row of $\boldsymbol{A}$. Note that the symmetric matrix $\boldsymbol{A}$ comprises only $(d^2 + d)/2$ different parameters which are found for example in the lower triangular part, including the diagonal elements

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & \dots & a_d \\ a_2 & a_{d+1} & \dots & a_{2d-1} \\ \dots & \dots & \dots & \dots \\ a_d & a_{2d-1} & \dots & a_{(d^2+d)/2} \end{pmatrix} = \begin{pmatrix} a^{11} & a^{12} & \dots & a^{1d} \\ a^{21} & a^{22} & \dots & a^{2d} \\ \dots & \dots & \dots & \dots \\ a^{d1} & a^{d2} & \dots & a^{dd} \end{pmatrix}.$$

Let $\boldsymbol{a}_p$ be the one–dimensional array formed by these parameters,

$$\boldsymbol{a}_p = (a_1, \dots, a_{(d^2+d)/2}),$$

then we have to find the vector $\boldsymbol{a}_p$ that minimises

$$S := \sum_{j=1}^{d}\sum_{i=1}^{N}(w_{ij} - \langle\, p_i, a^j\,\rangle)^2 \longrightarrow \min_{\boldsymbol{a}_p}. \tag{2.28}$$

We obtain by differentiation of (2.28) with respect to all elements of $\boldsymbol{a}_p$ the linear equation system $\boldsymbol{B}\boldsymbol{a}_p = \boldsymbol{c}$ which can be solved by standard methods. In detail, we construct the coefficient matrix $\boldsymbol{B}$ and the constant vector $\boldsymbol{c}$ in the following way. For the diagonal elements of $\boldsymbol{A}$ we get

$$\frac{\partial S}{\partial a^{ll}} = -2\sum_{i=1}^{N}(w^{il} - \langle\, p_i, a^l\,\rangle)p^{il} \;=\; 0$$

for $l = 1, \dots, d$. This is equivalent to

$$\begin{aligned} \sum_{i=1}^{N} w^{il}p^{il} \;&=\; \sum_{i=1}^{N}\langle\, p_i, a^l\,\rangle\, p^{il} \\ &=\; \sum_{i=1}^{N}\sum_{j=1}^{d} p^{ij}a^{jl}p^{il} = \sum_{j=1}^{d}\sum_{i=1}^{N} p^{ij}p^{il}a^{jl}, \end{aligned}$$

what gives

$$\langle w^l, p^l \rangle = \sum_{j=1}^{d} \langle p^j, p^l \rangle a^{jl}. \tag{2.29}$$

For the off–diagonal elements we obtain

$$\frac{\partial S}{\partial a^{kl}} = -2\sum_{i=1}^{N}(w^{il} - \langle p_i, a^l \rangle)p^{ik} - 2\sum_{i=1}^{N}(w^{ik} - \langle p_i, a^k \rangle)p^{il} = 0$$

for $k, l = 1, \ldots, d$ and $k > l$. This is equivalent to

$$\langle w^l, p^k \rangle + \langle w^k, p^l \rangle = \sum_{j=1}^{d} \langle p^j, p^k \rangle a^{jl} + \sum_{j=1}^{d} \langle p^j, p^l \rangle a^{jk}. \tag{2.30}$$

The t–th entry in $\boldsymbol{c}$ we get now by the left hand side of (2.29), when $\boldsymbol{a}_p(t)$ corresponds to a diagonal element of $\boldsymbol{A}$, or by the left hand side of (2.30) otherwise. Note that $t = k + (2d - l)(l - 1)/2$ for $k \geq l$ and $t = 1, \ldots, (d^2 + d)/2$. The t–th row of $\boldsymbol{B}$ we obtain by the right hand side of (2.29) or (2.30), where obviously the factors of the a^{jk} are our searched coefficients. An explicit solution is thus available and no iteration is necessary for estimating $\mathbf{A}$ in the partial linear model case. This is exactly the reason for both the much smaller variance of the resulting estimates in practice and the much higher speed of the algorithm for estimating model (2.1) compared to the varying-coefficients model (2.3).

Example For the simple case $d = 3$ we get the linear equation system $\boldsymbol{B}\boldsymbol{a}_p = \boldsymbol{c}$ with $\boldsymbol{a}_p = (a^{11}, a^{21}, a^{31}, a^{22}, a^{32}, a^{33})$, the coefficient matrix $\boldsymbol{B}$

$$\begin{pmatrix}
\langle p^1, p^1 \rangle & \langle p^2, p^1 \rangle & \langle p^3, p^1 \rangle & 0 & 0 & 0 \\
\langle p^1, p^2 \rangle & \langle p^2, p^2 \rangle + \langle p^1, p^1 \rangle & \langle p^3, p^2 \rangle & \langle p^2, p^1 \rangle & \langle p^3, p^1 \rangle & 0 \\
\langle p^1, p^3 \rangle & \langle p^2, p^3 \rangle & \langle p^3, p^3 \rangle + \langle p^1, p^1 \rangle & 0 & \langle p^2, p^1 \rangle & \langle p^3, p^1 \rangle \\
0 & \langle p^1, p^2 \rangle & 0 & \langle p^2, p^2 \rangle & \langle p^3, p^2 \rangle & 0 \\
0 & \langle p^1, p^3 \rangle & \langle p^1, p^2 \rangle & \langle p^2, p^3 \rangle & \langle p^3, p^3 \rangle + \langle p^2, p^2 \rangle & \langle p^3, p^2 \rangle \\
0 & 0 & \langle p^1, p^3 \rangle & 0 & \langle p^2, p^3 \rangle & \langle p^3, p^3 \rangle
\end{pmatrix}$$

and vector $\boldsymbol{c}$ as

$$\begin{pmatrix}
\langle w^1, p^1 \rangle \\
\langle w^1, p^2 \rangle + \langle w^2, p^1 \rangle \\
\langle w^1, p^3 \rangle + \langle w^3, p^1 \rangle \\
\langle w^2, p^2 \rangle \\
\langle w^2, p^3 \rangle + \langle w^3, p^2 \rangle \\
\langle w^3, p^3 \rangle
\end{pmatrix}.$$

Chapter 3

Nonparametric Prediction of Stock Returns with Generated Bond Yields

3.1 Introduction and Motivation

For a long time predicting asset returns was a main objective in the empirical finance literature. It started with simple regressions of independent predictor variables on stock market returns. Typically, valuation ratios are used that primarily characterize the stock, like the dividend price ratio, dividend yield, earnings price ratio or the book-to-market ratio (cf. Wolf (2000)). Other variables that are related to the interest rate, for example treasury-bill rates and long-term bond yield, or macroeconomic indicators like inflation are often incorporated to improve prediction; see also Guo and Savickas (2006). For a detailed overview we refer to the examples in Goyal and Welch (2008), the references in Campbell and Thompson (2008) or Campbell and Diebold (2009).

The apparent predictability found by many authors was controversially discussed. As Lettau and Nieuwerburgh (2008) note, correct inference is problematic due to the high persistence of financial ratios, which have only poor out-of sample forecasting power that moreover shows significant instability over time. Therefore, the question of whether empirical models are able to forecast the equity premium more accurately than the simple historical mean was intensively debated in the financial literature. For example Goyal and Welch (2008) fail to provide benefits of predictive variables compared to the historical mean. Recently, Rapach, Strauss, and Zhou (2010) recommend a combination of individual forecasts, including this way the information provided from different variables and reducing forecast volatility.

Several authors report that long-horizon returns can be better predicted than short-horizon returns, see, for example, Valkanov (2003). Among others, Campbell, Lo, and MacKinlay (1997) document that the power to forecast stock returns increases with the horizon. In contrast, Boudoukh, Richardson, and Whitelaw (2008) criticize this findings as an illusion based on the fact that the classical R^2 of the typically used linear model is roughly proportional to the considered horizon.

A classical approach to valuate the price of a stock is the well-known Gordon growth or dividend discount model which expresses the dividend price ratio, d_t, in terms of the long-term discount rate, R, and the long-term growth rate of dividends, G, both hold constant, $d_t = R - G$. Allowing for time-varying discount rates, for example Campbell and Shiller (1988b) introduce a dividend ratio model, where the price of a stock today is seen as the discounted present value of future cash flows to the investor; compare also Campbell and Diebold (2009) or Wolf (2000). A well-known fact is the high correlation of any relevant discount rate to inflation and interest rate. Many authors conclude that a decrease in discount rate is related to an increase in the stock return, and point to the high correlation with an increase in the bond yield.

A direct comparison of stocks and bonds, mostly used by practitioners, makes the so-called FED model, which relates yields on stocks, as ratios of dividends or earnings to stock prices, to yields on bonds. Asness (2003) shows the empirical descriptive power of the model, but notes also that it fails in predicting stock returns. One of his criticisms is the comparison of real numbers to nominal ones. Actually, most studies discuss separately the predictability in stock and bond markets. However, Engsted and Tanggaard (2001) pose the interesting question of whether expected returns on stocks and bonds are driven by the same information and to what extent they move together. In their empirical setting, they find that excess stock and bond returns are positively correlated, but they also note that simple present value models cannot explain this finding. Already ten years before, Shiller and Beltratti (1992) analyzed the relation between stock prices and changes in long-term bond yields.

Our article is based on the ideas from present value relations of stocks and bonds that expected returns are associated with variables related to longer-term aspects of business conditions, as mentioned in Campbell (1987) or Campbell and Diebold (2009). Consequently, we include in a fully nonparametric prediction model of excess stock returns the bond yield of the same year. This way, the bond captures the perhaps most important part of the stock return, the one related to the change in long-term interest rate. We do this nonparametrically due to the promising findings of Nielsen and Sperlich (2003). They improved significantly the prediction power by allowing for nonlinearities and interactions. Certainly, the use of nonparametric methods is not undisputed, see, for example, the discussion in Qi (1999) or Racine (2001). We use local linear kernel regression to nest the linear model without bias. For the purpose of bandwidth selection and to measure the quality of prediction, we use the *validated* R^2 of Nielsen and Sperlich (2003). This allows us to directly compare the proposed cross-validated model with the cross-validated mean.

An obvious problem is that the current bond yield is unknown. Thus, we have to predict it in a first step. This raises the question why it is necessary to use a two-step procedure. One could directly include the variables used for the bond prediction when forecasting stock returns. The problem is that such a model would suffer from the *curse of dimensionality and complexity* in several aspects: the dimension of the covariates, possible over-fitting and the interpretability. In nonparametrics it is well known that the import of structure is an adequate way to circumvent these problems. This Chapter is based on the structure that is inherent in the economic process that generates the data, resulting in the inclusion of bond yields when predicting stock returns. In other words, one may think of the inclusion of predicted bond yields as a kind of dimension (or say, complexity) reduction. Additionally, Park et al. (1997) showed that an appropriate transformation of the predictors can significantly improve nonparametric prediction. Here, we use the additional knowledge about structure to improve the prediction of stock returns. To our knowledge we are the first who include nonparametrically generated regressors for nonparametric prediction of time series data. Therefore we also have to develop the theoretical justification for the use of constructed variables in nonparametric regression when the data are *dependent.*

For the empirical part of our work we use the annual Danish stock market data from Lund and Engsted (1996). This has been done for reasons of comparability and reproducibility as these were also the data used in the above mentioned papers of Engsted and Tanggaard (2001) or Nielsen and Sperlich (2003). It will be shown that the inclusion of predicted bond yields improves greatly the prediction quality of stock returns in terms of the *validated* R^2. With our best prediction model for one-year stock returns we not only beat the simple historical mean but we also obtain an impressive *validated* R^2 of 28.3 compared to 5.9 from the best model without constructed bonds. We also include in our empirical analysis the prediction of the ratio of stock returns and dividend yields getting similar results.

Section 3.2 describes the prediction framework: the measure used for quantifying the quality of prediction and the model in mathematical terms, Section 3.3 the appertaining theoretical proof. Section 3.4 presents our findings from the empirical study, and Section 3.5 concludes. Some of the technical parts of the proof are deferred to the appendix in Section 3.6.

3.2 The Prediction Framework

The most used measures of prediction quality in the financial and actuarial literature are traditional in-sample approaches like the classic R^2, the adjusted R^2, goodness-of-fit or testing methods. In various articles the authors construct tests to check whether the apparent prediction power is only due to high persistency. In our study, we use the *validated* R^2 – short R^2_V – which was introduced as a measure for prediction power. It measures how well the model predicts in the future compared to the sample mean. The classical R^2 is often used in empirical finance because it is easy to calculate and has a straight forward interpretation. But it should be known that it can hardly be used for prediction nor for comparison issues as it always prefers here the most complex model. An other problem appears when the estimator is not consistent and consequently the R^2, as shown by Valkanov (2003). For comparison often the adjusted R^2 is applied, which penalizes complexity via a degree of freedom adjustment. This gets meaningless when methods are applied for which it is not clear what degrees of freedom are. This is unfortunately the case in a nonparametric setting, see Sperlich, Linton, and Härdle (1999).

In prediction, we are not interested in how well the considered model explains the variation inside the sample – the interpretation of the R^2 or R^2_{adj} – but we would like to know how well the estimate works outside the considered sample. The idea of the R^2_V is to replace total variation and not explained variation by their cross-validated analogues. Note that cross-validation is a quite common in the nonparametric time series context, see Györfi et al. (1990). More formally, consider the two models

$$Y_t = \mu + \varepsilon_t \qquad \text{and} \qquad Y_t = g(X_t) + \zeta_t,$$

where μ is estimated by the sample mean $\bar{Y}$ and the unknown function g by local linear kernel regression. We suppress a subscription for the chosen smoothing parameter h, since we always apply the bandwidth h that maximizes the R^2_V. It is defined as

$$R^2_V = 1 - \frac{\sum_t \{Y_t - \hat{g}_{-t}\}^2}{\sum_t \{Y_t - \bar{Y}_{-t}\}^2}. \tag{3.1}$$

Note that in (3.1) cross-validated values $\hat{g}_{-t}$ and $\bar{Y}_{-t}$ are used, i. e. the function g and the mean $\bar{Y}$ are predicted at t without the information contained in this point in time. In fact, it is a kind of out-of-sample measure. We apply the *leave-(2s+1)-out* version of cross-validation, with s depending on the context. For time series this means that when one wants to predict stock returns over a four year horizon, it is important to exclude the t-th observation and the three years before and after year t, i. e. $s = 3$. This is the modified cross-validation for mixing data along Chu and Marron (1991). Following Györfi et al. (1990), we use it also to find the optimal (prediction) bandwidth.

Notice some basic properties of the R_V^2: it is independent of the amount of parameters and takes it values inside $(-\infty, 1]$. As it measures how well a given model and estimation principle predicts compared to the cross-validated mean, a negative R_V^2 indicates that one predicts worse than the cross-validated mean. In practical prediction it is well known, and confirmed in our empirical study, that it is hard to find models with important explanatory variables which beat even this cross-validated mean. Nielsen and Sperlich (2003) mention that complexity is one of the worst enemies of a good prediction. Therefore, the R_V^2 punishes overfitting (pretending a functional relationship that is not really there) resulting in $R_V^2 < 0$.

We analyze excess stock returns defined as

$$S_t = \log\{(P_t + D_t)/P_{t-1}\} - r_{t-1},$$

where D_t denotes the (nominal) dividends paid during year t, P_t the (nominal) stock price at the end of year t, and r_t the short-term interest rate, which is

$$r_t = \log(1 + R_t/100)$$

using the discount rate R_t. For prediction, we consider $Y_t = \sum_{i=0}^{T-1} S_{t+i}$, the excess stock return at time t over the next T years.

Based on the motivations from the introduction, we include the same years bond yield as a single regressor or together with further lagged covariates in the model equation, i. e. we consider the model

$$Y_t = g(\hat{b}_t, v_{t-1}) + \varepsilon_t, \tag{3.2}$$

with the unknown function g, the constructed bond yield $\hat{b}_t$, a vector of further regressors v_{t-1} and error terms ε_t, i. e. mean zero variables given the past. As mentioned above, the problem which occurs is that the current bond yield is unknown. Therefore, we must predict them in a prior step, i. e. we construct the bond yield with the fully nonparametric model

$$b_t = p(w_{t-1}) + \zeta_t, \tag{3.3}$$

where p is an unknown function, w_{t-1} is a vector of explanatory variables as for example, lagged interest rates or bond yields, and ζ_t an error. Both, model (3.2) and (3.3), are estimated with a local linear kernel smoother using cross-validation. For the choice of the bandwidth, we basically have two possibilities. Either we treat each model separately, determining first the best (in terms of R_V^2) bond model and using this in the second step, or we choose the bandwidth in both steps according to the best R_V^2 for the stock return prediction.

As discussed, not only economic intuition motivates the inclusion of the constructed bond yields but also statistical arguments. In Theorem 3.3.7 we will develop the mathematical justification for the use of constructed variables in the case of dependent data. In other words, when we nonparametrically estimate stock returns using a generated regressor, we asymptotically obtain the same function as had we observed the real bond yield. Basically, the bias of the final estimate is enlarged by an additive factor which is proportional to the bias of the predicted variable from the first step. A similar relationship holds for the variance which is increased by an additive term proportional to the variance of the constructed regressor. This relates the prediction of bonds to the prediction of stocks. For simplification ignore for a moment v_{t-1} in (3.2), and call the function containing the actual bonds $\tilde{g}$. A closer look to the prediction error ε_t gives

$$\begin{aligned} Y_t - g(\hat{b}_t) &= Y_t - \tilde{g}(b_t) + \tilde{g}(b_t) - g(\hat{b}_t) \\ &\simeq \tilde{\varepsilon}_t + g'(\hat{b}_t)(b_t - \hat{b}_t). \end{aligned} \tag{3.4}$$

The gain in our two-step procedure comes from the fact that the second term in (3.4) is quite predictable, as we confirm in the empirical part 3.4.2, especially documented in Table 3.2. An other idea would be the following: first, estimate g with the available bond data b_{t-1}, and second, evaluate $\hat{g}$ at the constructed $\hat{b}_t$. Since, however, this procedure did not improve the stock forecasts, we skip it from further considerations.

On could directly use the variables in the vector w_{t-1} as regressors in model (3.2). But the model would suffer from complexity and dimensionality in several aspects: the dimension of the covariates as well as their interplay. In the nonparametric literature, typically two strategies are proposed to circumvent this; either semiparametric modelling or additivity, both to import structure. Nielsen and Sperlich (2003) showed that additive models fail to improve the prediction of stock returns due to a non-ignorable interaction between the predictors. We believe that the imposition of additional structure, which is inherent in the underlying financial process, yields better results. We think of the same years bond yield as an important factor which captures some of the relevant features for the expected stock returns. In other words, the inclusion of bond yields, when predicting stock returns nonparametrically, can be seen as a kind of complexity and dimension reduction due to the import of more structure.

To see if it is possible to further improve the predictive power in our setting, we will also analyze the model (3.2) with a different dependent variable. We consider the ratio between current stock returns and, first, dividend yield, i. e. $Y_t^* = Y_t/d_t$, second, long-term interest rate, L_t, $Y_t^{**} = Y_t/L_t$, and third, risk-free rate, r_t, $Y_t^{***} = Y_t/r_t$; see Section 3.4.

3.3 Mathematical Justification

We have to prove that asymptotically the function estimate which makes use of constructed variables will coincide with the real one, as it is the case for non predicted but observed covariates. For the prediction in the time series context, we follow the steps from Ferraty, Núñez-Antón, and Vieu (2001) and combine them with Sperlich (2009). More technical parts of the proofs can be found in Section 3.6.

Let us consider a sample of real random variables $\{(X_i, Y_i), i = 1, \ldots, n\}$ which are not necessarily independent. We want to estimate the unknown function

$$m(x) = E(Y|X = x), \quad x \in \mathbb{R},$$

that should always exist. For a given time series $\{(Z_i), i \in \mathbb{N}\}$ a k-step ahead forecast is included in a natural way in the given context using $Y_i = Z_{i+k}$ and $X_i = Z_i$. We concentrate only on the case of an auto-regression function of order one. Since we face constructed realisations for X, we assume a predictor[5] with an additive bias and a stochastic error:

$$\hat{x} = x + b(x) + u_x \sigma_u(x), \tag{3.5}$$

where the random variables u_x are independent, with conditional mean zero and conditional variance one. This is rather general as it holds for almost all nonparametric predictors. For technical reasons, we further assume finite higher moments for u_x. Then, for example the Nadaraya-Watson estimator can be defined with

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{\hat{X}_i - x}{h}\right) \quad \text{and} \quad \hat{q}(x) = \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{\hat{X}_i - x}{h}\right)$$

as

$$\hat{m}_{NW}(x) = \frac{\hat{q}(x)}{\hat{f}(x)}, \tag{3.6}$$

where K denotes some kernel function with bandwidth h.

To measure the strength of dependence in the time series, we limit us to the *strong-* or *α-mixing*[6] defined in Doukhan (1994) or Fan and Yao (2003) as

$$\lim_{n \to \infty} \alpha(n) = 0,$$

for the *mixing coefficient*

$$\alpha(n) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{n}^{\infty}} |P(A)P(B) - P(AB)|,$$

[5] We do not specify a particular one but we will need some assumptions on it. The used sample of size N, for instance, consists of some instruments $\boldsymbol{Z} \in \mathbb{R}^{\delta}$. In the following, we have N=n, since we use the same series in both the prediction and final step.

[6] The weakest of the usually defined mixing conditions.

where $\mathcal{F}_i^j$ is the σ-algebra generated by $\{X_k, i \leq k \leq j\}$. We further assume, that the sequence $\{(X_i, Y_i), i = 1, \ldots, n\}$ is *algebraic* α-mixing, i. e. that for some real constants $a, c > 0$ we have $\alpha(n) \leq cn^{-a}$.

To calculate the asymptotic properties in the context of strong mixing, we make use of an exponential inequality of the Fuk-Nagaev type (cf. Rio (2000)).

Lemma 3.3.1 *For an algebraic α-mixing sequence of random variables $\{(Z_i), i \in \mathbb{N}\}$, with $s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\mathrm{cov}(Z_i, Z_j)|$ and $||Z_i||_\infty < \infty$ for all i, holds for some $\varepsilon > 0$ and $r > 1$*

$$P\left(\left|\sum_{i=1}^n Z_i\right| > 4\varepsilon\right) \leq 4\left(1 + \frac{\varepsilon^2}{rs_n^2}\right)^{-\frac{r}{2}} + 2ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{a+1}.$$

Furthermore, we need the Billingsley inequality from Bosq (1998) to bound from above the covariance of two elements of a strong-mixing time series.

Lemma 3.3.2 *For an α-mixing sequence of random variables $\{(Z_i), i \in \mathbb{N}\}$, with $||Z_i||_\infty < \infty$ for all $i \neq j$, holds*

$$|\mathrm{cov}(Z_i, Z_j)| \leq 4||Z_i||_\infty ||Z_j||_\infty \alpha(|i - j|).$$

To prove the asymptotic behavior of a kernel regression estimator (3.6), we make some common assumptions. As noted above, we analyse an algebraic α-mixing sequence of real random variables $\{(X_i, Y_i), i = 1, \ldots, n\}$. We suppose that for all $i \neq j$ the joint density f_{ij} for the pair (X_i, X_j) exists and that $|Y| < C < \infty$ almost surely.

Also for the unobservables X_i, we assume a density function f_X which is bounded and has a continuous second derivative. At the fix $x \in \mathbb{R}$ we suppose $f_X(x) > 0$.

Since we use kernel-based estimators, let the kernel K be integrable, bounded, with compact support and continuous second derivative. It fulfills $\int K(s)ds = 1$ and $\int sK(s)ds = \int K'(s)ds = \int K''(s)ds = 0$.

For both, the deterministic and the stochastic part of the predicted realisations $\hat{x}$ in (3.5), we assume that $b(x)$ and $b'(x)$ are of order $O(h_0^2)$ uniformly, and $\sigma_u^2(x)$ of order $O((nh_0^\delta)^{-1})$. Here, h_0 is a smoothing parameter tending to zero when the sample size n goes to infinity and δ refers to the dimension of the used instruments in the prediction step. Let further be $b(\cdot)$ and $\sigma_u(\cdot)$ Lipschitz-continuous.

To simplify our calculations, we further suppose that $h_0^2 h^{-1}$ and $(nh_0^\delta h)^{-1}$ go to zero, and use the usual assumption that nh and nh_0^δ go to infinity as $n \to \infty$.

Before we state the main result of the section, we collect some important facts. First, we define the following variables for $l \in \{0, 1\}$

$$Z_i = Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right) - E\left[Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right)\right] \tag{3.7}$$

and analyse the asymptotic behavior of

$$s_n^{2*} = \sum_{i=1}^{n} \sum_{j \neq i} |\mathrm{cov}(Z_i, Z_j)|.$$

Proposition 3.3.3 *Under the above assumptions,*

$$s_n^{2*} = o(nh) + O(n^2 \alpha(\tilde{\Delta})),$$

where $\tilde{\Delta}$ has the same order like the slowest from

$$\left\{\frac{1}{h \log n}, \frac{n h_0^\delta h}{\log n}, \frac{(n h_0^\delta h)^2 h}{\log n}\right\}.$$

Note, when $(nh_0^\delta)^{-1} = O(h^2)$ the above proposition reduces to $s_n^{2*} = o(nh) + O(n^2 \alpha((h \log n)^{-1}))$ as in the case without predicted realisations.

Proposition 3.3.4 *Under the given assumptions and if an $\varepsilon > 0$ exists such that*

$$\Delta^{a-1} = O(n^{-1-\varepsilon}), \tag{3.8}$$

with Δ from $\{h, (nh_0^\delta h)^{-1}, ((nh_0^\delta)^2 h^3)^{-1}\}$, it holds that

$$s_n^{2*} = o(n\Delta).$$

Proposition 3.3.5 *Under the stated assumptions, and with Δ as in Proposition 3.3.4 we have*

$$\mathrm{var}(Z_1) = O(\Delta).$$

A direct conclusion of these Propositions is

$$s_n^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} |\mathrm{cov}(Z_i, Z_j)| = n \cdot \mathrm{var}(Z_1) + s_n^{2*} = O(n\Delta), \tag{3.9}$$

with Δ from Proposition 3.3.4. Before we can directly specify the result about the convergence of the estimator (3.6) we need a further proposition.

Proposition 3.3.6 *Under the above assumptions, and with* Δ *from Proposition 3.3.4 that verifies the condition*

$$c_1 n^{\frac{3-a}{a+1}+\theta} \leq \Delta \leq c_2 n^{\frac{1}{1-a}-\theta}, \tag{3.10}$$

with existing $c_1,\ c_2,\ \theta > 0$*, holds for* ν *and* $\varepsilon > 0$ *with* $\psi = g$ *or* $\psi = f$

$$P\left(\left|E\hat{\psi}(x) - \hat{\psi}(x)\right| > \varepsilon\sqrt{\frac{\log n}{nh^2}}\ \Delta\right) = O(n^{-1-\nu}). \tag{3.11}$$

In the following, we can state without problems the main theorem of this section. For continuous (around x) functions m and f we find the quasi complete convergence (cf. Serfling (1980)) of the Nadaraya-Watson estimator with a constructed regressor.

Theorem 3.3.7 *Under the given assumptions and* (3.10)*,*

$$|\hat{m}_{NW}(x) - m(x)| \ \longrightarrow\ 0$$

quasi completely.

The extension to the local linear estimator is almost straightforward. With

$$s_j(x) = \sum_{i=1}^{n} K\left(\frac{\hat{X}_i - x}{h}\right)(\hat{X}_i - x)^j \quad \text{and} \quad t_j(x) = \sum_{i=1}^{n} K\left(\frac{\hat{X}_i - x}{h}\right)(\hat{X}_i - x)^j Y_i,$$

for $j = 0, 1, 2$, we can define

$$\hat{m}_{LL}(x) = \frac{t_0(x)s_2(x) - t_1(x)s_1(x)}{s_0(x)s_2(x) - s_1^2(x)}.$$

Basic algebra leads to

$$\hat{m}_{LL}(x) = \frac{\sum\limits_{i=1}^{n} C\left(\frac{\hat{X}_i - x}{h}\right) Y_i}{\sum\limits_{i=1}^{n} C\left(\frac{\hat{X}_i - x}{h}\right)}, \tag{3.12}$$

with $C\left(\frac{\hat{X}_i - x}{h}\right) = \sum\limits_{j \neq i} K\left(\frac{\hat{X}_i - x}{h}\right)(\hat{X}_j - \hat{X}_i)K\left(\frac{\hat{X}_j - x}{h}\right)(\hat{X}_j - x)$ as a discretization of $C(u) = \int K(u-v)vK(u)udu$[7]. Since equation (3.12) is of the same form like (3.6) and the kernel C fulfills the same conditions as K, the application of Theorem 3.3.7 leads to

[7] Note that C is a bimodal kernel. Since it puts more weight to points close to x, except if they are too close to x, than to points far from x, it would be a natural and desirable choice in the case of strong mixing data, as noted in Kim et al. (2009).

Corollary 3.3.8 *Under the assumptions of Theorem 3.3.7, quasi completely*

$$|\hat{m}_{LL}(x) - m(x)| \longrightarrow 0.$$

For mean square convergence, asymptotic normality and higher order polynomials, one could directly extend the work of Masry and Fan (1997) to the case of predicted regressors.

3.4 Empirical Evidence

Note that we interpret our method presented so far as a two stage model selection approach. Coming from economic motivation that the bond of the same year captures an important part of the stock return we search in the first step the optimal prediction model for the bond. Afterwards, as we have seen in Theorem 3.3.7, we can consistently predict stock returns using the nonparametrically constructed bond yields.

3.4.1 Data Description

We consider the annual Danish stock and bond market data for the period 1923 – 1996 from Lund and Engsted (1996). In the appendix of their work, a detailed description of the data can be found. We use a stock index based on a value weighted portfolio of individual stocks chosen to obtain maximum coverage of the market index of the Copenhagen Stock Exchange (CSE). Notice that the CSE was open during the second world war. When constructing the data, corrections were made for stock splits and new equity issues below market prices. Table 3.1 presents summary statistics of the available variables. In the following, we use the dividend price ratio, d, the stock return, S, the long-term interest rate, L, the short-term interest rate, r, and the bond yields, b, as explanatory variables.

3.4.2 Bond Prediction

We start with model (3.3) and try to find the best bond prediction model – no stocks yet. As we can see in Table 3.1, the bond data are less volatile than the stock market data. Furthermore, Table 3.2 indicates that we can predict the current bond yield quite well since the table shows large *validated* R^2_V values of some of the analyzed models. This is not surprising because it agrees with the in Section 3.1 established present value theory that the bond is basically driven by the interest rate and indeed all advisable models include the interest rate as an explanatory variable.

Table 3.1: *Danish stock and bond market data (1923-1996).*

	Min	Max	Mean	Sd
CSE Stock Price Index	64.78	3177.88	511.07	662.99
Dividend Accruing to Index	2.99	44.76	15.11	11.81
Stock Returns	-42.44	72.10	2.10	17.19
Bond Yield	-13.70	60.30	8.61	12.07
Dividend Yield	0.01	0.08	0.04	0.01
Long-term Interest Rate	3.80	19.45	8.24	4.29
Short-term Interest Rate	2.50	17.86	6.96	3.46

Note that in Table 3.2 also results of a simple linear regression are available. There are no large differences to the results from the nonparametric model, but in the most cases the R_V^2 values of the latter approach are larger than that one from the parametric counterpart. The fact that also this simple parametric approach gives a good prediction of bonds confirms our strategy. In fact, we are in the position to forecast the current bond yield in an adequate way and can include this produced information in the more interesting second step – the stock prediction.

Table 3.2: *R_V^2-values for bond model* (3.3).

w_{t-1}	S	L	r	d, L	d, r	S, L
par.	11.6	24.0	22.3	21.9	19.4	31.9
nonpar.	16.3	23.9	26.8	23.2	19.3	31.6

w_{t-1}	S, r	L, b	r, b	S, L, b	S, r, b	L, r, b
par.	33.1	29.2	35.2	31.9	37.4	30.9
nonpar.	33.0	29.2	35.5	31.7	37.4	31.8

Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.

3.4.3 Stock Prediction

For the presentation of our results we concentrate in the following on a forecast horizon of one year, i. e. $Y_t = S_t$. For larger horizons we got analog results but without further insight. First, we estimate model (3.2) without the constructed bond variable again with a simple parametric regression and a fully nonparametric kernel based method. The results are summarized in the first two lines of Table 3.3 and show that all parametric models produce negative *validated* R_V^2 values. It means that with a simple regression approach we cannot better forecast one-year stock returns than the simple mean. A more sophisticated technique is needed. In fact, our so far best nonparametric model[8] uses lagged bond yields, b_{t-1}, and gives a R_V^2 of 5.9. Second, we follow our in Section 3.2 proposed procedure and generate the

Table 3.3: R_V^2*-values for stock model* (3.2).

w_{t-1}	d	S	L	r	b	d,S	d,L	d,r	d,b
nonpar	-1.4	1.8	-4.2	-3.6	5.9	5.5	-6.0	-7.4	3.1
par	-1.3	-1.8	-4.2	-5.7	-4.0	-3.5	-5.8	-7.2	-6.2
$\hat{b}_t$	8.3	1.3	-3.5	1.4	10.6	-1.5	-3.8	2.9	10.1
$\hat{b}_t, v_{t-1}$	13.9	5.1	9.1	16.3	8.9	-1.6	28.3	21.6	3.8

w_{t-1}	S,L	S,r	S,b	L,r	L,b	r,b	d,S,L	d,S,r	d,S,b
nonpar	-3.5	-7.1	4.6	-9.4	0.8	0.5	-2.9	-6.7	3.3
par	-6.8	-7.9	-6.6	-9.3	-7.5	-8.6	-8.6	-9.8	-8.8
$\hat{b}_t$	-1.1	-3.1	3.9	-0.6	-0.9	-3.6	1.3	-3.5	8.9
$\hat{b}_t, v_{t-1}$	1.6	13.5	-1.3	15.8	15.6	20.3	10.8	14.2	0.8

w_{t-1}	d,L,r	d,L,b	d,r,b	S,L,r	S,L,b	L,r,b	S,r,b
nonpar	-11.2	-3.8	1.0	-11.0	0.3	-4.4	-1.6
par	-10.9	-9.9	-11.2	-12.5	-11.1	-13.0	-11.9
$\hat{b}_t$	-2.8	-1.0	-3.7	-2.1	1.8	-3.6	1.6
$\hat{b}_t, v_{t-1}$	17.5	16.6	20.4	10.0	1.6	15.9	5.1

In each panel: nonparametric and parametric stock prediction model with w_{t-1} as covariates but without the constructed bond $\hat{b}_t$ (first and second line), nonparametric model only with $\hat{b}_t$ as regressor and bandwidth chosen in the final step (third line), nonparametric model with $\hat{b}_t$ and the same variables w_{t-1} as in the bond prediction step; bandwidth selection in the final step (fourth line); bond $\hat{b}_t$ constructed with model (3.3) *and the variables w_{t-1}. Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.*

[8] Nielsen and Sperlich (2003) report in an analog setting a R_V^2 value of 5.5 for a fully nonparametric two-dimensional model with dividend-price ratio, d_{t-1}, and lagged excess stock returns, S_{t-1}, as explanatory variables, but do not use bond yields in their analysis.

current bond yield with model (3.3) but include only this as regressor in the final step (3.2). Since we use nonparametric methods in both parts, we have two possibilities to choose the appropriate bandwidth. On the one hand we can separately treat both models, i. e. we use the best bond model and include the so constructed variable into the stock prediction. The best model following this procedure[9] uses lagged bond yields, b_{t-1}, to construct the current bond in the first step and gives a R_V^2 value of 9.0. On the other hand we could also choose the bond model in the final step, not including the best bond prediction model but that one that yields the largest R_V^2 value when we predict stock returns. The predictive power of this method is again greater as can be seen from line three[10] of Table 3.3. Our so far best model again uses lagged bond yields, b_{t-1}, in the first step and has a R_V^2 value of 10.6.

Third, we construct the current bond as before, but we use for stock return prediction this regressor in model (3.2) together with any combination of lagged variables from the predictor set $\{d, S, L, r, b\}$. The two highest R_V^2 were achieved by $\hat{g}(\hat{b}_t, d_{t-1}, S_{t-1}, L_{t-1})$ where $R_V^2 = 30.3$ for $\hat{b}_t = \hat{p}(d_{t-1})$ and $R_V^2 = 28.9$ for $\hat{b}_t = \hat{p}(d_{t-1}, L_{t-1})$. Note that for an increasing set of regressor variables the corresponding multidimensional bandwidth grid, on which we looked for the best predicting bandwidth combination, had to be reduced for numerical reasons. Consequently, lower dimensional models have the tendency to be slightly favored in our study. The full set of results for the 25 times 25 combinations of $\{d, S, L, r, b\}$ is not shown here for the sake of presentation, but available in Tables 3.6 to 3.8 in the appendix 3.7. An interest finding is that for each variable set the diagonal of all results, given in the last line of Table 3.3, seems to be among the best prediction models. So here we have $w_{t-1} = v_{t-1}$ for models (3.3) and (3.2). We see now clearly that our proposal greatly improves the predictive power for stock returns. For the best model in Table 3.3 – we construct the bond and predict stock returns with the dividend yield, d_{t-1}, and long-term interest rate, L_{t-1} – we find an impressive R_V^2 value of 28.3, an increase of the prediction quality by a factor of almost 5 compared to the model without constructed bonds. This finding again indicates that the bond captures and provides the most important part of the stock return which is related to the change in long term interest rate.

The last part of our empirical study concentrates on the change of the dependent variable. Up to now, we used the excess stock return but for the following we divide this value by the dividend yield, the short-term or the long-term interest rate, i. e.

[9]Since the bandwidth choice in the final step yields better results, the R_V^2 values for this approach are deferred to the appendix (see Table 3.5).

[10]Note that some R_V^2 of this method are smaller compared to the R_V^2 of the sequential procedure. The reason is that the computational burden growths exponentially. Hence, we use only a grid of 30×30 bandwidths compared to 100 different bandwidths in the sequential procedure.

we use $Y_t^* = Y_t/d_t$, $Y_t^{**} = Y_t/L_t$, and $Y_t^{***} = Y_t/r_t$. Table 3.4 summarizes our findings for Y^*. The results for Y^{**} and Y^{***} are deferred in Tables 3.9 and 3.10 of the appendix in 3.7 because the same models appear but with somewhat lower R_V^2-values.

Table 3.4: *R_V^2-values for model (3.2) with $Y_t^* = Y_t/d_t$ as dependent variable.*

w_{t-1}	d, L	d, r	L, r	L, b	r, b	d, L, r	d, r, b	L, r, b
par.	-8.1	-10.4	-11.2	-4.5	-6.8	-15.0	-11.2	-12.7
nonpar.	-8.3	-10.5	-11.3	1.3	11.4	-15.2	2.4	12.3
final	39.9	35.4	41.4	31.3	35.8	39.5	42.6	43.5

Parametric and nonparametric prediction model for Y^* with w_{t-1} as covariate but without the constructed bond $\hat{b}_t$ (first and second line), nonparametric model with $\hat{b}_t$ and the same variables w_{t-1} as in the bond prediction step; bandwidth selection in the final step (third line); bond $\hat{b}_t$ constructed with model (3.3) and the variables w_{t-1}. Lagged explanatory variables: d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.

The first line of Table 3.4 refers again to the parametric version of model (3.2) and the second line to the fully nonparametric method, both without constructed bonds. Almost all of the parametric models have negative R_V^2 values and also only a small number of nonparametric models beat the simple mean. In contrast, when we include the constructed bond in the nonparametric prediction, a large increase of the *validated* R_V^2 can be observed. For example, the model which uses long- and short-term interest rate, and lagged bond yields for both the bond generation and following stock prediction, has a R_V^2 value (43.5%) that is over three and a half times larger than the value of the best model without constructed bond (12.3%). We see again that the same years bond yield includes the change in interest rate in the stock estimation procedure what results in the strongest improvements.

3.5 Concluding Remarks and Outlook

Motivated by economic theory and statistical arguments, we include the same years bond yield in the fully nonparametric prediction approach for excess stock returns. Since the current bond yield is unknown, we propose to construct it in a prior step using again nonparametric techniques. The bandwidths should be chosen in such a way that they maximize the R_V^2 of the final step. The empirical study demonstrates that this two-step approach can enormously improve the stock return prediction.

Moreover, we prove the consistency of our method and derive the asymptotic behavior of our final predictor. We illustrate the improvement due to our method using annual Danish stock and bond market data which were studied in detail in former articles by different authors. The results confirm our motivation of including the same years bond yield, that it captures the most important part of the stock return, namely the part related to the change in long-term interest rate. This actually holds not only for stock returns but also for transformed variables, as for example returns divided by dividend yields.

The statistical key points are the following. It is clear that we face a regression model that exhibits high complexity and dimensionality. An obvious remedy would be the imposition of structure. Since it has been shown that separability is inappropriate because of unknown interactions, we make use of financial theory to exploit the inherent structure of stock returns. Alternatively, one could interpret the first stage as an optimal nonparametric transformation that maps, for example, the long-term interest rate to the current bond yield, $L_{t-1} \to \hat{b}_t$. The subsequent nonparametric smoother of the transformed variable is than characterized by less bias. Here, we present a practical example in the spirit of the somewhat theoretical method proposed by Park et al. (1997) which improves nonparametric regression with simple transformation techniques. Note the difference of our approach to their work, namely that we use an additional variable for the transformation whereas Park et al. (1997) estimate on the original scale.

3.6 Appendix A: Proofs

Proof of Proposition 3.3.3. Since the variables Z_i in (3.7) are centered, we calculate for $i \neq j$

$$\begin{aligned}|EZ_iZ_j| =& \Big|EY_i^lY_j^lK\Big(\frac{\hat{X}_i - x}{h}\Big)K\Big(\frac{\hat{X}_j - x}{h}\Big) \\ &- EY_i^lK\Big(\frac{\hat{X}_i - x}{h}\Big)EY_j^lK\Big(\frac{\hat{X}_j - x}{h}\Big)\Big|. \end{aligned} \tag{3.13}$$

First, we analyse the second term in the last equation and use the assumption that all Y_i are bounded.

$$EY_i^lK\Big(\frac{\hat{X}_i - x}{h}\Big) \leq C\int\int K\Big(\frac{u - x + b(u) + v\sigma(u)}{h}\Big)f(u,v)dudv.$$

A simple Taylor-expansion of the kernel leads to

$$\begin{aligned}&C\int\int \{K\Big(\frac{u-x}{h}\Big) + K'\Big(\frac{u-x}{h}\Big)\Big(\frac{b(u)+v\sigma(u)}{h}\Big)+ \\ &K''\Big(\frac{u-x}{h} + \kappa\,\frac{b(u)+v\sigma(u)}{h}\Big)\frac{(b(u)+v\sigma(u))^2}{2h^2}\}f(u,v)dudv,\end{aligned}$$

where $\kappa \in (0,1)$. With the common substitution $s = (u-x)h^{-1}$ we get

$$EY_i^l K\Big(\frac{\hat{X}_i - x}{h}\Big) = O(h + (nh_0^\delta h)^{-1}).$$

Analog steps lead to

$$E\Big[Y_i^l Y_j^l K\Big(\frac{\hat{X}_i - x}{h}\Big) K\Big(\frac{\hat{X}_j - x}{h}\Big)\Big] = O(h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2}),$$

and thus the covariance $|\text{cov}(Z_i, Z_j)|$ for $i \neq j$ is of the same rate.

On the other hand, we can directly make use of Lemma 3.3.2 because all Y_i and K are bounded so that $||Z_i||_\infty < \infty$. It follows

$$|\text{cov}(Z_i, Z_j)| \leq C\alpha(|i-j|).$$

The idea is now to combine both results. When the indices of the two variables Z_i and Z_j are close[11] to each other we use the first result, and when they are far from each other the second one. To control this, we introduce a sequence of integers a_n and obtain

$$\begin{aligned} s_n^{2*} &= \sum_{i=1}^{n} \sum_{j \neq i} |\text{cov}(Z_i, Z_j)| \\ &\leq C\Big[\sum \sum_{0<|i-j|\leq a_n} \{h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2}\} + \sum \sum_{|i-j|>a_n} \alpha(|i-j|)\Big]. \end{aligned}$$

Since $i \neq j$, the largest possible term for a_n is $0 < |i-j| \leq a_n \cong n-1$ and the smallest $|i-j| > a_n \cong 1$. Furthermore, the maximum number of elements in s_n^{2*} is $n^2 - n$, and we obtain for Δ that it is of the same order like the slowest term in $O(h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2})$

$$\frac{\sum \sum_{0<|i-j|\leq a_n} \Delta}{\Delta n a_n} \leq \frac{n^2 - n}{n(n-1)} \quad \Longleftrightarrow \quad \sum \sum_{0<|i-j|\leq a_n} \Delta = O(\Delta n a_n)$$

and

$$\frac{\sum \sum_{|i-j|>a_n} \alpha(|i-j|)}{n^2 \alpha(a_n)} \leq \frac{n^2 - n}{n^2} \quad \Longleftrightarrow \quad \sum \sum_{|i-j|>a_n} \alpha(|i-j|) = O(n^2 \alpha(a_n)).$$

This means that

$$s_n^{2*} = O(\{h^2 + (nh_0^\delta)^{-1} + (nh_0^\delta h)^{-2}\} n a_n + n^2 \alpha(a_n)).$$

Choosing a_n from $\{\frac{1}{h \log n}, \frac{nh_0^\delta h}{\log n}, \frac{(nh_0^\delta)^2 h^3}{\log n}\}$ proves the Proposition. □

[11] Since we use time series data, this means that the two events are close in time.

Proof of Proposition 3.3.4. Using the algebraic mixing condition and Proposition 3.3.3

$$s_n^{2*} = o(nh) + O(n^2\tilde{\Delta}^{-a}),$$

with $\tilde{\Delta}$ from

$$\left\{\frac{1}{h\log n}, \frac{nh_0^\delta h}{\log n}, \frac{(nh_0^\delta h)^2 h}{\log n}\right\}.$$

Using the assumption (3.8) and noting that $\frac{(\log n)^a}{n^\varepsilon} \to 0$ for $n \to \infty$ closes the proof. □

Proof of Proposition 3.3.5. We use again that Y_1 is bounded so that remains to analyse

$$E\Big[K\Big(\frac{\hat{X}_1 - x}{h}\Big)\Big]^2 = \int\int K\Big(\frac{u - x + b(u) + v\sigma(u)}{h}\Big)^2 f(u,v)dudv.$$

With a Taylor-expansion and analog steps like in the proof of Proposition 3.3.3 we get

$$= \int\int \{K\Big(\frac{u-x}{h}\Big) + K'\Big(\frac{u-x}{h}\Big)\Big(\frac{b(u)+v\sigma(u)}{h}\Big) +$$
$$K''\Big(\frac{u-x}{h} + \kappa\,\frac{b(u)+v\sigma(u)}{h}\Big)\frac{(b(u)+v\sigma(u))^2}{2h^2}\Big\}^2 f(u,v)dudv,$$

where $\kappa \in (0,1)$, and find that

$$E\Big[K\Big(\frac{\hat{X}_1 - x}{h}\Big)\Big]^2 = O(h + (nh_0^\delta h)^{-1} + (nh_0^\delta)^2 h^3)^{-1})$$

what proves the Proposition. □

Proof of Proposition 3.3.6. Using $l = 0$ for $\psi = f$ and $l = 1$ for $\psi = q$, respectively, we directly get with (3.7)

$$\begin{aligned}
|E\hat{\psi}(x) - \hat{\psi}(x)| &= \left| E\left(\frac{1}{nh}\sum_{i=1}^n Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right)\right) - \frac{1}{nh}\sum_{i=1}^n Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right)\right| \\
&= \frac{1}{nh}\left|\sum_{i=1}^n \left\{Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right) - E\left(Y_i^l K\left(\frac{\hat{X}_i - x}{h}\right)\right)\right\}\right| \\
&= \frac{1}{nh}\left|\sum_{i=1}^n Z_i\right|.
\end{aligned}$$

Therefore, applying Lemma 3.3.1 we obtain

$$\begin{aligned}
P\left(|E\hat{\psi}(x) - \hat{\psi}(x)| > \delta\right) &= P\left(\Big|\sum_{i=1}^n Z_i\Big| > nh\delta\right) \\
&\le 4\left(1 + \frac{\delta^2 n^2 h^2}{16 r s_n^2}\right)^{-\frac{r}{2}} + 2ncr^{-1}\left(\frac{8r}{nh\delta}\right)^{a+1}.
\end{aligned}$$

Since we have seen in (3.9) that $s_n^2 = O(n\Delta)$, with Δ from Proposition 3.3.4, with $\delta = \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}$ we get

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big)$$
$$\leq 4\left(1 + \frac{\varepsilon^2 \log n}{16r}\right)^{-\frac{r}{2}} + 2ncr^{-1}\left(\frac{8r}{\varepsilon}\right)^{a+1}(n\Delta\log n)^{-\frac{a+1}{2}}. \qquad (3.14)$$

Now, we can choose $r > 1$ such that $\log n = o(r)$, and use the limit definition

$$\exp(x) = \lim_{z\to\infty}\left(1 + \frac{x}{z}\right)^z,$$

with $z = -r/2$. For the first term of the right hand side of (3.14), we obtain for $z \to \infty$

$$\left(1 + \frac{\varepsilon^2\log n}{16r}\right)^{-\frac{r}{2}} = \left(1 - \frac{\varepsilon^2\log n}{32z}\right)^{z} \longrightarrow \exp\left(-\frac{\varepsilon^2\log n}{32}\right) = n^{-\frac{\varepsilon^2}{32}}.$$

Noting that $C(\log n)^{-(a+1)/2} \leq C$ for $n > 2$ and a constant C, (3.14) can be expressed as

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big) \leq Cn^{-\frac{\varepsilon^2}{32}} + C\varepsilon^{-(a+1)}n^{1-\frac{a+1}{2}}r^a\Delta^{-\frac{a+1}{2}}.$$

With $r = n^b$ for $b > 0$, i. e. $\log n = o(r)$, and the left hand side of the assumption (3.10),

$$n^{1+ab-\frac{a+1}{2}}\Delta^{-\frac{a+1}{2}} \leq n^{1+ab-\frac{a+1}{2}-\frac{3-a}{2}-\theta\frac{a+1}{2}} = n^{-1-\theta\frac{a+1}{2}+ab} = n^{-1-\nu}.$$

Thus, we obtain for a sufficiently small b that

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big) \leq Cn^{-\frac{\varepsilon^2}{32}} + C\varepsilon^{-(a+1)}n^{-1-\nu}.$$

Finally, for a sufficiently large ε, we get that exist ν, $\varepsilon > 0$ such that

$$P\Big(|E\hat{\psi}(x) - \hat{\psi}(x)| > \varepsilon\sqrt{\frac{\log n}{nh^2}\,\Delta}\Big) \leq Cn^{-1-\nu},$$

what proves the assertion. □

Proof of Theorem 3.3.7. From Proposition 3.3.6 follows directly

$$E\hat{q}(x) - \hat{q}(x) \longrightarrow 0, \quad \text{and} \quad E\hat{f}(x) - \hat{f}(x) \longrightarrow 0, \qquad (3.15)$$

both quasi completely. With the first part of the proof of Proposition 3.3.3 we obtain[12]

$$E\hat{f}(x) = \frac{1}{h}EK\left(\frac{\hat{X} - x}{h}\right) = f(x) + B_f(x) + o(h_0^2 + h),$$

[12] A similar result can be found in Theorem 2.1 (i) in Sperlich (2009).

with $B_f(x) = h^2/2f''(x)\mu_2(K) + \{b(x)f'(x) + b'(x)f(x)\}\mu_1(K')$, and thus

$$E\hat{f}(x) - f(x) \longrightarrow 0. \tag{3.16}$$

The analog can be shown for $E\hat{q}(x)$. With

$$E\hat{q}(x) = \frac{1}{h} EYK\left(\frac{\hat{X} - x}{h}\right),$$

and taking the conditional expectation for $X = x$, we get

$$E\hat{q}(x) = \frac{1}{h}\int\int m(u)K\left(\frac{u - x + b(u) + v\sigma(u)}{h}\right) f(u,v)dudv.$$

Repeating the same steps as in the first part of the proof of Proposition 3.3.3, using $q = m \cdot f$ as well as that the function q is continuous over the compact support of the kernel K, i. e. that $q(x + hs) \longrightarrow q(x)$ uniformly in s, we obtain

$$E\hat{q}(x) - q(x) \longrightarrow 0. \tag{3.17}$$

Furthermore, from (3.16) and (3.11) follows the quasi complete convergence of $\hat{f}(x)$ to $f(x)$, i. e. for all $\varepsilon > 0$, it holds that

$$\sum_{n=1}^{\infty} P(|\hat{f}(x) - f(x)| > \varepsilon) < \infty.$$

Since $f(x) > 0$, we can define $\delta = \varepsilon = f(x)/2$ and get for $\delta > 0$

$$\sum_{n=1}^{\infty} P(\hat{f}(x) \leq \delta) < \infty. \tag{3.18}$$

Note, that with (3.6) and $q = f \cdot m$ we can state

$$\hat{m}_{NW}(x) - m(x) = \frac{\hat{q}(x) - q(x)}{\hat{f}(x)} + (f(x) - \hat{f}(x))\,\frac{m(x)}{\hat{f}(x)}, \tag{3.19}$$

and thus with (3.15) – (3.19) follows the assertion. □

3.7 Appendix B: Tables of Additional Results

Table 3.5: R_V^2*-values for stock model* (3.2) *using the constructed bond; bandwidths chosen separately in each step.*

w_{t-1}	d	S	L	r	b	d,S	d,L	d,r	d,b
$\hat{b}_{t,\text{sep}}$	2.2	0.1	-4.2	-10.7	9.0	-3.4	-5.2	-4.3	6.5
$\hat{b}_{t,\text{sep}}, v_{t-1}$	0.4	2.3	9.7	9.6	4.1	-2.0	13.0	19.7	3.0
w_{t-1}	S,L	S,r	S,b	L,r	L,b	r,b	d,S,L	d,S,r	d,S,b
$\hat{b}_{t,\text{sep}}$	-3.7	-4.0	6.4	-4.1	-2.4	-3.9	-0.2	-3.7	8.6
$\hat{b}_{t,\text{sep}}, v_{t-1}$	1.6	12.3	-36.0	4.0	10.2	18.6	1.9	-16.2	-28.5
w_{t-1}	d,L,r	d,L,b	d,r,b	S,L,r	S,L,b	L,r,b	S,r,b		
$\hat{b}_{t,\text{sep}}$	-3.7	-3.0	-4.1	-3.9	-2.0	-4.1	-3.6		
$\hat{b}_{t,\text{sep}}, v_{t-1}$	13.8	-3.4	14.2	-19.5	-11.2	13.2	5.1		

In each panel: nonparametric stock prediction model only with the constructed bond $\hat{b}_t$ as regressor (fist line), nonparametric model with $\hat{b}_t$ and the same covariates w_{t-1} as in the bond prediction step (second line); bond $\hat{b}_t$ constructed with model (3.3) *and the variables w_{t-1}; bandwidth selection in each step (in contrast to Table 3.3). Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.*

Table 3.6: *R_V^2-values for stock model (3.2) using the constructed bond and further covariates; bandwidths chosen in the final step (part I).*

	d	S	L	r	b
d	13.9	10.1	4.6	3.5	8.1
S	-0.4	5.1	-3.4	-6.6	0.0
L	-5.1	-3.3	9.1	-11.0	1.3
r	1.1	8.0	7.5	16.3	5.5
b	9.6	8.6	7.8	7.0	8.9
d, S	5.7	4.9	-4.3	-10.5	1.4
d, L	-5.8	-3.7	21.2	-9.8	2.5
d, r	-3.7	-3.8	3.8	16.6	4.4
d, b	5.8	9.8	6.2	4.5	8.4
S, L	-3.7	-1.5	-6.1	-7.1	4.6
S, r	-5.5	-4.8	-8.1	-8.4	1.7
S, b	1.9	2.5	-0.7	-1.4	6.1
L, r	1.2	-2.4	18.1	1.9	5.7
L, b	-4.8	2.5	10.7	5.1	6.1
r, b	-6.1	-2.1	-7.3	6.2	5.9
d, S, L	-4.5	-1.5	-3.9	2.9	4.1
d, S, r	-5.5	-3.7	-10.4	-4.0	0.6
d, S, b	6.3	2.9	5.6	1.7	4.9
d, L, r	-4.5	-4.0	14.2	4.2	3.6
d, L, b	-4.6	1.3	-3.1	3.8	5.9
d, r, b	-6.1	-3.0	-10.3	-4.7	1.9
S, L, r	-2.9	-2.8	-7.8	1.5	1.6
S, L, b	-1.1	2.5	1.2	-0.9	4.6
S, r, b	-2.3	-2.0	-1.7	4.2	1.7
L, r, b	-5.9	-2.5	-9.4	8.6	4.5

The columns refer to different regressors in the stock model and the rows correspond to different variables in the bond model. Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.

Table 3.7: *R_V^2-values for stock model (3.2) using the constructed bond and further covariates; bandwidths chosen in the final step (part II).*

	d,S	d,L	d,r	d,b	S,L	S,r	S,b	L,r	L,b	r,b
d	23.6	13.7	11.8	12.6	4.8	2.5	6.6	0.2	3.7	3.4
S	3.8	-5.9	-8.4	-1.1	1.3	0.3	4.1	-9.7	0.8	-3.3
L	-1.1	10.6	-13.3	-1.5	6.1	-13.2	1.0	11.0	13.0	-7.3
r	7.7	4.1	17.6	4.2	4.4	12.5	4.1	14.5	8.6	20.0
b	8.2	6.2	5.2	8.4	5.2	4.1	5.4	0.8	5.6	4.9
d,S	-1.6	-2.8	-10.1	0.4	0.3	-2.8	8.0	-13.7	-3.2	-6.2
d,L	-2.0	28.3	-10.5	-2.4	23.5	-12.3	0.7	19.3	21.4	-3.6
d,r	-3.2	-0.2	21.6	2.4	0.6	14.3	2.5	13.7	4.5	22.5
d,b	6.5	-0.3	-1.1	3.8	6.3	3.1	6.8	-2.9	3.5	2.7
S,L	-1.0	-2.4	-6.2	1.1	1.6	-8.3	1.9	-15.5	-5.3	-2.3
S,r	-3.2	-7.8	-6.0	1.4	1.8	13.5	-1.1	-16.5	-2.4	0.1
S,b	1.7	-2.8	-4.9	4.4	-1.2	-3.8	-1.3	-3.6	4.9	2.2
L,r	4.3	16.1	10.2	7.0	16.9	-1.8	5.8	15.8	20.6	9.1
L,b	1.6	1.2	2.9	1.7	0.0	-5.0	1.9	8.9	15.6	0.0
r,b	1.5	-5.1	11.7	6.1	-7.9	-4.1	0.2	2.2	5.0	20.3
d,S,L	0.7	-0.4	-1.8	1.7	9.3	-8.2	0.7	-13.0	-3.5	2.9
d,S,r	2.1	-8.1	-2.0	-1.2	-2.3	14.3	-1.3	-7.0	-7.1	-4.3
d,S,b	0.3	1.9	-2.7	1.6	-1.2	-5.2	3.0	-7.9	2.3	-0.9
d,L,r	-1.5	11.0	7.3	0.4	14.9	0.9	2.1	20.3	15.2	13.5
d,L,b	1.0	1.3	3.8	1.1	-2.7	-2.7	1.4	-0.9	8.7	3.1
d,r,b	0.2	-5.5	9.4	3.6	-9.7	-8.7	-1.3	-2.6	2.9	17.5
S,L,r	4.5	-6.8	1.5	0.9	7.0	2.1	-0.1	-11.5	-3.0	0.3
S,L,b	0.6	2.8	0.9	2.7	-2.5	-5.3	2.1	-5.1	5.3	-1.4
S,r,b	2.2	-6.6	11.2	5.2	-5.9	-2.3	1.3	2.7	6.3	24.7
L,r,b	-2.5	-2.5	6.6	-1.2	-7.5	-1.9	-0.9	2.5	-1.0	2.3

The columns refer to different regressors in the stock model and the rows correspond to different variables in the bond model. Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.

Table 3.8: R_V^2*-values for stock model* (3.2) *using the constructed bond and further covariates; bandwidths chosen in the final step (part III).*

	d,S,L	d,S,r	d,S,b	d,L,r	d,L,b	d,r,b	S,L,r	S,L,b	S,r,b	L,r,b
d	30.3	26.1	18.3	7.0	17.5	15.2	-1.5	1.7	4.5	0.7
S	-0.4	-0.9	3.4	-12.4	-3.6	-5.2	2.1	4.5	-5.6	3.3
L	9.9	-11.2	-1.8	10.6	13.0	-6.5	6.1	6.1	13.7	-10.4
r	7.1	14.4	3.6	14.6	2.8	17.8	8.3	5.2	14.2	13.4
b	2.5	2.2	4.5	-3.1	1.7	2.5	-5.1	1.8	-1.7	0.6
d,S	-6.3	-7.9	-0.9	-15.4	-6.4	-7.7	-7.4	2.8	-7.4	-1.0
d,L	28.9	-10.0	-2.8	26.8	26.9	-5.8	18.6	23.0	18.3	-6.4
d,r	-3.3	19.7	0.5	19.2	-1.3	23.1	9.9	0.9	16.5	19.0
d,b	-0.1	-1.1	3.6	-8.1	-2.1	-3.2	-6.0	1.4	-5.0	-0.8
S,L	5.8	-7.6	-1.8	-9.1	-5.8	5.1	0.7	1.9	-11.0	-5.6
S,r	-1.8	14.6	-2.4	-10.9	-9.1	0.3	9.6	1.2	-4.6	13.5
S,b	-4.0	-6.5	-0.9	-7.3	0.8	-0.2	-5.0	-4.4	-1.7	-8.7
L,r	15.4	11.5	7.3	14.5	15.8	13.4	12.7	17.4	17.3	6.8
L,b	-3.7	-6.3	-0.2	-0.3	15.8	4.6	-7.6	7.8	13.1	-5.5
r,b	-3.8	-3.4	-0.5	6.2	0.7	19.6	-8.5	-2.0	14.2	13.0
d,S,L	10.8	-8.9	-1.8	-7.2	-4.9	4.2	2.8	7.9	-1.6	-2.1
d,S,r	-3.5	14.2	-1.6	-9.2	-10.5	4.5	11.8	-2.5	-0.7	15.3
d,S,b	-4.4	-8.1	0.8	-6.3	-2.5	-6.2	-7.0	0.5	-2.9	-2.2
d,L,r	9.0	8.4	-0.6	17.5	9.4	13.4	19.6	15.3	23.6	12.0
d,L,b	-3.6	-6.0	-1.9	2.7	16.6	2.7	-6.4	7.4	6.2	-0.8
d,r,b	-4.4	-6.6	-1.4	3.2	-0.7	20.4	-11.6	-2.9	17.3	12.5
S,L,r	4.2	9.2	-1.6	-5.3	-8.0	-1.4	10.0	5.7	-2.3	8.0
S,L,b	-4.7	-6.4	-1.8	0.0	3.7	3.9	-9.0	1.6	-1.4	-6.5
S,r,b	-3.3	-1.6	2.1	3.7	-1.3	19.2	-5.5	2.1	15.9	18.6
L,r,b	-6.2	0.9	-2.4	4.7	-7.4	6.0	-2.9	-2.6	-3.6	5.1

The columns refer to different regressors in the stock model and the rows correspond to different variables in the bond model. Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.

Table 3.9: *R^2_V-values for model (3.2) with different dependent variables, $Y^*_t = Y_t/d_t$, $Y^{**}_t = Y_t/L_t$, and $Y^{***}_t = Y_t/r_t$ (part I).*

w_{t-1}	d	S	L	r	b	d,S	d,L	d,r	d,b	S,L	S,r	S,b	L,r	L,b	r,b
Y^*_t	-5.0	1.4	-2.8	-6.0	-2.3	-3.5	-8.1	-10.4	-6.4	-2.2	-4.5	-2.5	-11.2	-4.5	-6.8
	-5.1	2.9	-2.5	0.9	5.5	0.4	-8.3	-10.5	-0.8	-1.2	-3.2	6.6	-11.3	1.3	11.4
	28.2	7.4	13.0	18.4	16.7	-6.1	39.9	35.4	7.6	4.0	21.3	9.2	41.4	31.3	35.8
Y^{**}_t	-2.4	-1.6	-2.0	-1.8	-1.9	-4.0	-4.5	-4.3	-4.5	-3.7	-3.2	-4.2	-2.7	-4.0	-4.0
	-2.5	1.0	-2.0	-1.8	3.2	0.9	-4.6	-4.4	-1.6	-2.3	-2.7	2.1	-2.7	0.4	1.9
	8.1	2.0	3.8	7.6	3.2	-2.1	12.6	9.2	0.1	-0.3	4.0	-0.9	9.6	4.2	11.8
Y^{***}_t	-1.9	-2.6	-2.1	-1.4	-2.2	-4.6	-4.0	-3.6	-4.4	-4.7	-3.8	-5.6	-2.9	-4.4	-3.8
	-1.9	2.5	-2.1	-1.4	1.8	2.3	-4.1	-3.7	-2.3	-1.2	-2.3	0.7	-2.9	-0.8	0.5
	3.8	3.4	1.8	8.4	2.3	0.8	10.4	5.1	-0.7	-0.9	1.4	-2.2	3.7	2.4	8.5

In each panel: parametric and nonparametric prediction model with w_{t-1} as covariates but without the constructed bond $\hat{b}_t$ (first and second line), nonparamertic model with $\hat{b}_t$ and the same variables w_{t-1} as in the bond prediction step; bandwidth selection in the final step (third line). Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.

Table 3.10: R_V^2*-values for model* (3.2) *with different dependent variables,* $Y_t^* = Y_t/d_t$, $Y_t^{**} = Y_t/L_t$, *and* $Y_t^{***} = Y_t/r_t$ *(part II).*

w_{t-1}	d,S,L	d,S,r	d,S,b	d,L,r	d,L,b	d,r,b	S,L,r	S,L,b	L,r,b	S,r,b
Y_t^*	-7.3	-9.2	-6.7	-15.0	-8.6	-11.2	-10.4	-5.8	-12.7	-7.7
	-4.8	-3.2	0.8	-15.2	-5.5	2.4	2.8	2.2	12.3	3.1
	7.8	23.3	-5.3	39.5	19.7	42.6	29.8	6.0	43.5	2.1
Y_t^{**}	-6.3	-5.9	-6.8	-6.0	-6.8	-6.9	-5.4	-6.7	-5.6	-6.3
	-5.3	-5.5	-1.6	-6.1	-5.1	-3.7	-4.0	-1.8	0.4	-1.0
	0.7	2.3	-4.9	10.1	9.0	11.2	5.9	-0.1	10.4	2.9
Y_t^{***}	-6.8	-6.3	-7.7	-5.9	-6.7	-6.3	-6.7	-8.1	-5.7	-7.0
	-4.4	-5.3	-2.1	-6.0	-5.4	-4.1	-5.1	-2.9	-2.0	-2.4
	1.1	0.3	-3.2	4.7	8.0	7.1	-0.5	-0.5	3.9	2.0

In each panel: parametric and nonparametric prediction model with w_{t-1} as covariates but without the constructed bond $\hat{b}_t$ (first and second line), nonparamertic model with $\hat{b}_t$ and the same variables w_{t-1} as in the bond prediction step; bandwidth selection in the final step (third line). Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, b bond yield.

Chapter 4

Nonparametric Prediction of Stock Returns Guided by Prior Knowledge

4.1 Introduction and Overview

One of the most studied questions in economics and finance is whether equity returns or premiums are predictable. Until the mid-1980's, the view of financial economists was that returns are not predictable, at least not in an economically meaningful way, see for example Fama (1970), and that stock market volatility does not change much over time. Tests of predictability were motivated by *efficient capital markets* and it was common to assume that predictability would contradict to constant expected returns, the efficient markets paradigm.

However, the empirical research in the late twentieth century suggests that excess returns (over short-term interest rates) are predictable, especially over long horizons, as pointed out in Cochrane (1999). For example, Fama and French (1988b) or Poterba and Summers (1988) take only past returns in an univariate mean-reverting sense into account and find rather weak statistical significance, which seems stronger by the inclusion of other predictive variables. In the vast literature, among others, short term interest rates (Fama and Schwert (1977) or Campbell (1991)), yield spreads (Keim and Stambaugh (1986), Campbell (1987), or Fama and French (1989)), stock market volatility (French, Schwert, and Stambaugh (1987) or Goyal and Santa-Clara (2003)), book-to-market ratios (Kothari and Shanken (1997) or Ponti and Schall (1998)), and price-earnings ratios (Lamont (1998) or Campbell and Shiller (1988a)) are proposed. There are also numerous articles which examine the predictive power of the dividend yield and, particularly, the dividend ratio on excess stock returns over different horizons. The most influential of them are Fama and French (1988a, 1989), Campbell and Shiller (1988a,b), and Nelson and Kim (1993). For the economic interpretation and the question what drives this predictability, we refer to the discussion in Rey (2004).

By the recent progress in asset pricing theory and the still growing number of publications that report empirical evidence of return predictability it seems that the paradigm of constant expected returns was abandoned. In this spirit, conditional and dynamic asset pricing models (e. g. Campbell and Cochrane (1999)) as well as models that analyze the implications of return predictability on portfolio decisions, when expected returns are time-varying (e. g. Campbell and Viceira (1999)), are proposed. But, certain aspects of the empirical studies cast doubt on the predicting ability of price-based variables and should be considered with caution. While, for example, Fama and French (1988a), Campbell (1991) or Cochrane (1992) find that the aggregate dividend yield strongly predicts excess returns, with even stronger predictability at longer horizons, in contrast, Boudoukh, Richardson, and Whitelaw (2008) criticize these findings as an illusion based on the fact that the R^2 of the model is roughly proportional to the considered horizon. Also Ang and Bekaert (2007) find only short-horizon predictability. On the other hand, Rapach, Strauss, and Zhou (2010) recommend a combination of individual forecasts, including this way the information provided from different variables and reducing forecast volatility. Goyal and Welch (2008) favor the historical average in forecasting excess stock returns, which gives better results than predictive regressions with different variables, but then again Campbell and Thompson (2008) respond that many of them beat the historical mean by imposing weak restrictions on the signs of coefficients and return forecasts, or by imposing restrictions of steady-state valuation models. Thus, the evidence for stock market predictability is still controversial debated and less transparent than previous work may have suggested.

The most popular model in the economic and financial literature is the discounted-cash-flow or present-value model, which relates the price of a stock to its expected future cash flows, namely, its dividends, discounted to the present value using a constant or time-varying discount rate (e. g. Rozeff (1984), Campbell and Shiller (1987, 1988a,b) or West (1988)). The model assumes the efficient market paradigm of constant expected returns and is based on the well-known discrete-time perfect certainty model (Gordon growth model[13]) and its dynamic generalization. Hence,

[13] In this model, the stock price at time t is $P_t = \sum_{j=1}^{\infty} D_t(1+G)^j/(1+R)^j = D_{t+1}/(R-G)$, with dividend per share D, growth at constant rate G, price P, and interest rate R $(G < R)$. Thus, the dividend yield is the interest rate minus the dividend growth, $D_{t+1}/P_t = R - G$.

stock prices are high when dividends are discounted at a low rate[14] or when dividends are expected to grow rapidly. Limitations of this linear model like the apparently exponential growth of stock prices or dividends over time makes it less appropriate than a nonlinear model which can better capture the properties of returns over time as mentioned by Chen and Hong (2009). For example, Froot and Obstfeld (1991) introduce a dividend model with intrinsic bubbles which are nonlinearly driven by exogenous fundamental determinants of asset prices. An other possible extension to the simple model is the use of a log-linear approximation of the present-value relation, see, for example, in Campbell (1991) or Ang and Bekaert (2007). Thus, the asset price behavior can be modeled without imposing restrictions on expected returns. Following these studies and their results that expected asset returns and dividend ratios are time-varying and highly persistent, it is important to model the relationships between equity returns and dividend ratios, interest rates, excess returns, or cash flows in a nonlinear fashion.

In the most empirical studies, the linear predictive regression[15] is applied. Even though this type of model is rather simple, the econometric problems that appear in forecasting asset returns, in testing predictability, and in evaluating the predictive power of the model are numerous. First, the fact that several predictor variables like valuation ratios are highly persistent might cause the found predictability to be spurious. Stambaugh (1999) points out that, although an OLS estimate would be consistent, it is biased and has sampling distributions that differ from those in the standard setting. Also Nelson and Kim (1993) mention that biases affect inference and should be accounted for in practice when studying predictability. These problems become even more serious if data-mining is used. Ferson, Sarkissian, and Simin (2003) show that spurious regression and data-mining effects reinforce each other such that many regressions, based on single predictor variables, may result in spurious conclusions. Possible solutions can be found in Amihud and Hurvich (2004), where an augmented regression[16] is used, in Chiquoine and Hjalmarsson (2008), where an jackknifing procedure is proposed, or in Jansson and Moreira (2006), where inference in a bivariate regression is conducted. Second, an additional source of bias in predictive regressions is the *error-in-variables* problem coming from the fact that, for example, yields contain forecasts of future returns and dividend growth (cf. the discussion in Fama and French (1988a), Goetzmann and Jorion (1995) or Lettau and Ludvigson (2005)), and thus, the explanatory variable is not properly exogenous. Kothari and Shanken (1992) examine the extent to which aggregated stock

[14]In the original model the discount rates are assumed to be constant at the initial levels.

[15]Typically, the simple specification $Y_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1}$ is used, where Y_{t+1} is the log excess return (over a certain horizon) and X_t the predictive variable, which follows a first-order autoregressive process, $X_{t+1} = \gamma + \delta X_t + \xi_{t+1}$.

[16]They add a proxy for the errors in the autoregressive model.

return variation is explained by variables, chosen to reflect revisions in expectations of future dividends, and provide evidence that the error-in-variables problem is a major one[17]. Third, the main concern in long-horizon predictive regression follows from the use of overlapping data such that error terms are caused to be strongly serially correlated, particularly when the time horizon is relatively large compared to the sample size. Hodrick (1992) examines the statistical properties of different methods for conducting inference in long-horizon regression and his simulations indicate that the test statistics can be substantially biased, but he still concludes with some predictability for U.S. stock market returns. Also Nelson and Kim (1993) analyze small-sample biases in their simulations of a VAR system for returns and dividend yields. Under the null hypothesis of no predictability, they find that the simulated distributions of t-statistics are biased upward by an amount that increases with the horizon and, nevertheless, report predictability of post-war U.S. stock returns. In another simulation, Goetzmann and Jorion (1993) use a bootstraping approach to illustrate how inference may be affected and report only marginal evidence of predictability. More recently, Wolf (2000) uses subsampling for finding reliable confidence intervals—for regression parameters in the context of dependent and possibly heteroscedastic data—and does not find convincing evidence for the predictability of stock returns. Valkanov (2003) shows that, in finite samples where the forecasting horizon is a nontrivial fraction of the sample size, the *t*-statistics do not converge to a well-defined distribution, and reports only weak predictive power of the dividend yield. Also Ang and Bekaert (2007) find that, at long-horizons, excess return predictability by the dividend yield is not statistically significant using a structural model of equity premiums and accounting for small sample properties. Alternative econometric methods or new statistical tests for conducting valid inference and bias correction can be found in the literature[18]. These studies emphasize that the usual corrections to standard errors are only valid asymptotically and pose the question whether *asymptotic* should be measured in terms of years, decades, or centuries, particularly for long-horizon forecasts. Fourth, Rey (2004) notes that recent theoretical econometric results indicate that these methods fail to provide an asymptotically valid inference when the predictive variable has a near unit root. Lewellen (2004), Torous, Valkanov, and Yan (2004) or Campbell and Yogo (2006) show that incorporating information about the order of integration can result in large efficiency gains and therefore have a significant effect on inferences. Fifth, while previous studies usually review the inclusion of financial and macroeconomic variables in the linear

[17] Also Fama (1990) finds a substantial increase in R^2 including measures of future industrial production as further regressors in his analysis.

[18] See, for example, Cavanagh, Elliot, and Stock (1995), Mark (1995) Kilian (1999), Lewellen (2004), Campbell and Yogo (2006), Polk, Thompson, and Vuolteenaho (2006).

regression framework[19], the functional form of the regression is not verified. Chen and Hong (2009) mention that, for example, a VAR model cannot fully capture the nonlinear dynamics of dividend yields implied by the present value model. Thus, for a linear regression, one cannot conclude that the null hypothesis of no predictability holds, because there may exist a disregarded nonlinear relationship[20]. But, more and more articles in the literature address this topic. For example, Abhyankar, Copeland, and Wong (1997) provide a summary of evidence of nonlinearity, Qi (1999) uses a neural network to examine U.S. stock market return predictability, or Perez-Quiros and Timmermann (2000) apply a Markov switching model for returns of large and small U.S. firms. However, for all of them the functional form is known, while McMillan (2001) examines the relationship between U.S. stock market returns and various predictive variables with a model-free nonparametric estimator. Also Harvey (2001) or Drobetz and Hoechle (2003) analyze conditional expectations of excess returns with nonparametric techniques, but fail to improve forecasts. In contrast, Nielsen and Sperlich (2003) obtain improvements compared to parametric models using a local-linear kernel-based estimator and Danish stock market data. Sixth, different authors, for example, Goyal and Welch (2003, 2008), Butler, Grullon, and Weston (2005) or Campbell and Thompson (2008), criticize that most linear predictive regressions have often performed poorly out-of-sample[21]. It is well-known that useful information on possible misspecified models can be revealed by in-sample diagnostics, while in this way overfitting can be caused or spurious predictability captured. Out-of-sample evaluation could be a possibility to solve these problems and capture the true predictability of a model or a data generating process[22]. For example, Clark (2004) shows with Monte Carlo simulations that out-of-sample forecast comparisons can help prevent overfitting, but in contrast, Inoue and Kilian (2004) conclude that results of in-sample tests of predictability will be more credible due to more power than results of out-of-sample tests. Thus, an overall assessment of return predictability remains difficult, and the question, whether the reason for poor out-of-sample performance of linear prediction models is due to possible nonlinear relations or due to the unpredictability of returns, persists unclear. Numerous studies that use

[19]Hodrick (1992), Campbell and Shiller (1988a,b), or Stambaugh (1999) use a finite-order VAR system.

[20]Campbell and Shiller (1998) point out that it is quite possible that the true relation between valuation ratios and long-horizon returns is nonlinear. In this case a linear regression forecast might be excessively bearish.

[21]Particularly, during the bull market of the late 1990's, low valuation ratios predicted extraordinarily low stock returns that did not materialize until the early 2000's (Campbell and Shiller (1998)).

[22]In the literature, it is often argued that the data snooping bias (and the associated ill effects, see, for example, Lo and MacKinlay (1990) or White (2000)) can be alleviated, if not completely eliminated, by out-of sample evaluation. Also, unforeseen structural changes or regime shifts can cause poor forecasts from models with good in-sample fits.

out-of-sample tests have focused on valuation ratios. While, for Fama and French (1988a), the out-of-sample performance of the dividend yield has been a success, Bossaerts and Hillion (1999) discover that even the best prediction models have no out-of-sample forecasting power. Torous and Valkanov (2000) study predictive regressions with a small signal-noise ratio and find that in this case spurious regression is unlikely to be a problem. They further argue that the excessive noisy nature of returns, relative to the explanatory variables, can explain both the apparent in-sample predictability as well as the failure to find out-of-sample forecasting power. Rapach and Wohar (2006) test stock return predictability with a bootstrap procedure and find that certain financial variables display significant in-sample and out-of-sample forecasting ability. Goyal and Welch (2008) systematically analyze the in-sample and out-of-sample performance of mostly linear regressions and find that the historical average return almost always gives better return forecasts. However, Campbell and Thompson (2008) show that most of the variables used by Goyal and Welch (2008) perform better out-of-sample than the forecast produced with the historical average return, if weak restrictions on the signs of coefficients and return forecasts are imposed. Despite the small out-of-sample explanatory power, they conclude that it is still economically meaningful for investors.

In this chapter, we propose a new way to include prior knowledge in the prediction of stock returns. Economic theory directly guides the modelling process. The immediate consequence of that is a dimension and bias reduction, both to import more structure as a proper way to circumvent the *curse of dimensionality*. First, we start with a fully nonparametric approach which allows the modeling of nonlinearities and interactions of predictive variables. Here, we estimate the model by a local-linear kernel regression smoother which already improves the predictive power in contrast to simple linear versions of the model. The long-lasting popularity of simple predictive regression models justifies the usefulness of the linear method for stock return prediction. However, a model (statistical or from financial theory) can only be an approximate to the real world and thus a linear model can only be seen as a first step in the representation of the unknown relationship in mathematical terms. Second, we include in a semiparametric fashion the available prior information, where the former nonparametric estimator is multiplicatively guided by the prior. This could be, for example, a standard regression model or likewise a good economic model provided by the clever economist. This approach helps to reduce bias in the nonparametric estimation procedure and thus to improve again the predictive power. An economist might provide an economic model better than our structured one. A good economic model should then be validated along the lines of this chapter. A nonparametric smoother guided by this economic model might be an excellent predictor. Third, we propose a simple bootstrap test to evidence that our method works and does not give better results just by chance. Fourth, we apply the proposed technique to American data. For the empirical part of this chapter, we use the annual data provided by Robert Shiller[23] that include, among other variables, long term stock, bond and interest rate data since 1871 to examine long term historical trends in the US market. It is an updated and revised version of Chapter 26 from Shiller (1989), where a detailed description of the data can be found. Note further that the application to this data set is not meant as a comprehensive study rather as an illustration of the auspicious and potential use of the strategy developed in this chapter.

Our scope is to show that linear predictive regression models suffer from neglected nonlinear relationships and that the inclusion of prior information further improves out-of-sample performance of nonlinear prediction models. Moreover, we evidence that our predictor-based regression models beat the historical average excess stock return. For this purpose, we apply for all models the *validated* R^2 of Nielsen and Sperlich (2003). This quality measure of the prediction allows directly the comparison of the cross-validated proposed model with the cross-validated historical mean in an out-of-sample fashion. Note further that we also use this instrument to find the optimal bandwidth in non- and semiparametric regression as well as to select the best model.

Note that we do not control and thus allow for nonstationarity, i. e. unit roots, in the predictive variables. Here, we follow the arguments of Torous, Valkanov, and Yan (2004). They show that due to rational expectations nonstationarities in predictive variables as functions of asset prices, for example dividend by price or earnings by price, can occur.

For the American data we find that, due to our bootstrap test, nonlinear models are more adequate than linear regressions, and that the inclusion of prior knowledge greatly improves the prediction quality. With our best prediction model for one-year excess stock returns we not only beat the simple historical mean but we also obtain an essentially improved *validated* R^2 of 18.5, a relative increase of 35% compared to the best nonparametric model without prior, or a relative increase of 131% compared to the simple regression.

The remainder of the chapter is structured as follows. Section 4.2 describes the prediction framework and the used measure of validation. Furthermore, the bootstrap test is introduced and first results of linear and nonlinear models are provided. Section 4.3 considers the nonparametric prediction that is guided in a new way by prior knowledge. Among others, the dimension reduction approach is evolved. Finally, Section 4.4 outlines wider results, summarizes the chapter and gives a short outlook.

[23]Downloadable from `http://www.econ.yale.edu/~shiller/data.htm`.

4.2 Preliminaries and First Steps

We consider excess stock returns defined as

$$S_t = \log\{(P_t + D_t)/P_{t-1}\} - r_{t-1},$$

where D_t denotes the (nominal) dividends paid during year t, P_t the (nominal) stock price at the end of year t, and r_t the short-term interest rate, which is

$$r_t = \log(1 + R_t/100)$$

using the discount rate R_t. In this chapter, we concentrate on forecasts over the one-year horizon, but also longer periods can easily be included with $Y_t = \sum_{i=0}^{T-1} S_{t+i}$, the excess stock return at time t over the next T years.

In the following, we study the prediction problem[24]

$$Y_t = g(X_{t-1}) + \xi_t, \tag{4.1}$$

where we want to forecast excess stock returns Y_t using lagged predictive variables X_{t-1}, like the dividend-price ratio, d_{t-1}, earnings by price, e_{t-1}, the long-term interest rate, L_{t-1}, the risk-free rate, r_{t-1}, inflation, inf_{t-1}, the bond, b_{t-1}, or also the stock return, Y_{t-1}. The functional form of g is fixed for the simple parametric relationship, but remains fully flexible for the non- and semiparametric counterpart. The error terms ξ_t are mean zero variables given the past.

4.2.1 The Measure of Validation

Since we use non- and semiparametric techniques, we need an adequate measure for the predictive power. Classical in-sample measures like R^2 or adjusted R^2 cannot be used because various problems occur. For example, the classical R^2 favors always the most complex model or is also inconsistent, if the estimator is inconsistent, as shown by Valkanov (2003). Furthermore, the usual penalization for complexity via a degree-of-freedom adjustment gets meaningless in nonparametrics because it is still unclear what degrees-of-freedom are in this setting. Moreover, in prediction we are not interested in how well a model explains the variation inside the considered sample but, in contrast, would like to know how well it works out-of-sample. For this reasons, we use the *validated* R^2 of Nielsen and Sperlich (2003) which has some nice features and is defined as

$$R_V^2 = 1 - \frac{\sum_t \{Y_t - \hat{g}_{-t}\}^2}{\sum_t \{Y_t - \bar{Y}_{-t}\}^2}. \tag{4.2}$$

[24]Basically, we address the regression problem of estimating the conditional mean function $g(x) = E(Y|X = x)$ using n i.i.d. pairs (X_i, Y_i) observed from a smooth joint density and its multivariate generalization.

Note that in (4.2) cross-validated values $\hat{g}_{-t}$ and $\bar{Y}_{-t}$ are used, i. e. the (parametric or nonparametric) function g and the historical mean $\bar{Y}$ are predicted at t without the information contained in this point in time, and hence, the R_V^2 is an out-of-sample measure. The *validated* R^2 is independent of the amount of parameters (in the simple parametric case of g) and measures how well a given model and estimation principal predicts compared to the cross-validated historical mean. This means for positive R_V^2 values, that the predictor-based regression model (4.1) beats the historical average excess stock return.

Moreover, cross-validation not only punishes overfitting, i. e. pretending a functional relationship which does not really exist, but also allows us to find the optimal (prediction) bandwidth for the non- and semiparametric estimators[25]. This means that we use the *validated* R_V^2 for both, model selection and optimal bandwidth choice[26].

Note further that in standard out-of-sample tests, which estimate the model up to some year and test on the next years data, the underlying amount of data changes in size for different years. But the standard variance-bias trade-off is extremely dependent on the underlying amount of data. Due to cross-validation, our approach with the R_V^2 has almost the exact correct underlying size of data so that the variance-bias trade-off of our validation is therefore expected to be more accurate than current methods. Moreover, for a stationary process it should not matter, if we skip only the information in point t or all following points in time. The only difference would be that the remaining size of data is to small for the application of non- or semiparametric methods.

4.2.2 A Bootstrap Test

To show that our method works and does not give better results as the cross-validated historical mean just by chance, we propose a simple bootstrap test. In this, we test the parametric null that the true model is the cross-validated historical mean against a non/semiparametric alternative, i. e. that the true model is our proposed fully nonparametric (4.5) or semiparametric model with (4.8). In detail, we estimate the model under the null and under the alternative, and calculate the R_V^2 as well as

$$\tau = \frac{1}{T} \sum_t \left(\hat{g}_{-t} - \bar{Y}_{-t} \right)^2 . \tag{4.3}$$

The intention is now to simulate the distribution of R_V^2 and τ under the null. Since we do not know the distribution of the underlying random variables, the excess stock

[25]See, for example, Györfi et al. (1990).
[26]In which, of course, the bandwidth choice is a part of the model selection process.

returns[27], we cannot directly sample from them and thus apply the wild bootstrap. For this, we construct B bootstrap samples $\{Y_1^b, \dots, Y_{\mathcal{T}}^b\}$ using the residuals under the null

$$\hat{\varepsilon}_t^0 = Y_t - \bar{Y}_{-t}$$

and independent and identically distributed random variables with mean zero and variance one, for example, $u_t^b \sim N(0,1)$, such that

$$Y_t^b = Y_t + \hat{\varepsilon}_t^0 \cdot u_t^b.$$

In each bootstrap iteration b, we calculate now the cross-validated mean $\bar{Y}_{-t}^b$ of the Y_t^b, $t = 1, \dots, \mathcal{T}$, as well as the estimates of the alternative model $\hat{g}_{-t}^b$, and, finally, $R_V^{2,b}$ and τ^b like in (4.2) and (4.3) with this new estimates. To decide, if we reject or not, we use critical values from corresponding quantiles of the empirical distribution function of the B bootstrap analogues $R_V^{2,b}$ or τ^b, for example, from

$$F^*(u) = \frac{1}{B} \sum_b \mathbb{I}_{\{\tau^b \leq u\}}.$$

This is a well-known testing procedure, which has proved to be consistent in numerous tests, and has therefore been applied, of cause with certain modifications, to many non- or semiparametric testing problems.

4.2.3 The Simple Predictive Regression

For the sake of illustration, we develop our strategy step by step and start with the simple model. In empirical finance, often the linear predictive regression model

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t \tag{4.4}$$

is used to evidence predictability of excess stock returns. We are fully aware of the in the introduction mentioned problems with this model, nevertheless, we use it in this basic form, not only as starting point of our empirical study but also as a straightforward possibility to generate a simple prior.

For the American data, Table 4.1 shows both, the usual adjusted and the *validated* R^2. More or less the same values appear, whereas the adjusted R^2 is always greater than the *validated* R^2. But already Fama and French (1988a) note that the usual in-sample R^2 tend to overstate explanatory power due to possible bias. More important, both measures evidence the earnings yield as the variable with the most explanatory power, i. e. we start our analysis with a *validated* R^2 of 8.0 and will concentrate on the behavior of models which include this covariate.

[27] It is a stylized fact that stock returns are not normally distributed. Using the wild bootstrap, we avoid this poor approximation.

Table 4.1: *Predictive power of the simple linear model* (4.4).

	S	d	e	r	L	inf	b
R_V^2	-1.0	1.0	8.0	2.7	-1.1	-1.4	-0.4
R_{adj}^2	0.2	1.7	8.8	3.6	-0.6	-0.4	-0.1

Lagged explanatory variables: S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

Our findings directly confirm to the results of Lamont (1998), who mentions the additional power of the earnings-price ratio for the prediction of excess stock returns in his study using postwar U.S. data. Interestingly, the often used dividend-price ratio gives only poor results.

4.2.4 The Nonparametric Model

Following the growing evidence of nonlinear behavior in asset returns documented in the literature, we examine the relationship of excess stock returns and the financial variables of the last section using a flexible, because model-free, nonparametric estimator. The model

$$Y_t = g(X_{t-1}) + \xi_t \tag{4.5}$$

is estimated with a local-linear kernel smoother using the quartic kernel and the optimal bandwidth chosen by cross-validation, i. e. by maximizing the R_V^2 as described in Section 4.2.1. Note again, that no functional form is assumed. One should further keep in mind that the nonparametric method can estimate the linear function without any bias, since we apply a local-linear smoother. Thus, the simple linear model is automatically embedded in our approach[28]. Table 4.2 shows the results, the *validated* R^2 and the estimated p-values of the bootstrap test. Remember that we test the parametric null hypothesis, i. e. the true model is the cross-validated historical mean, against the nonparametric alternative, i. e. model (4.5) holds. The estimated p-value gives the probability that under the null a R_V^2 value can be found which is greater or equal to the observed one.[29] Using the usual significance levels, we find only the earnings variable with a p-value of 0.005 to be able to forecast stock returns better than the historical mean. We further find an almost factor 1.5 increase

[28]This also holds for all of the non- and semiparametric models proposed in the rest of this work.

[29]We focus here on the R_V^2 and its estimated p-values, since no essential differences occur between the decisions made for R_V^2 and τ. Nevertheless, we show the τ statistics and its estimated p-values in the corresponding tables, since the basic distinction of both is the fact that τ basically measures only the variation between the estimates of two procedures, while the R_V^2 compares the fit of them.

Table 4.2: *Predictive power of the one-dimensional nonparametric model* (4.5) *and corresponding estimated p-values of the bootstrap test.*

	S	d	e	r	L	inf	b
R^2_V	-1.2	0.9	11.8	2.5	-0.8	-1.6	-0.7
p-value	0.596	0.193	0.005	0.079	0.571	0.759	0.573
τ	0.029	0.078	0.384	0.139	0.008	0.013	0.026
p-value	0.543	0.253	0.047	0.062	0.643	0.645	0.482

Lagged explanatory variables: S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

in the *validated* R^2 from 8.0 to 11.8, compared to the simple regression. Note also, that at a 10% level the risk-free rate has small predictive power with a R^2_V value of 2.5, which is smaller than the one obtained with the linear model.

Figure 4.1 shows for both variables the estimated linear and nonlinear functions. While for risk-free an almost identical linear relationship is found, for earnings by price, nonlinearities appear. Economic theory predicts that the short-term interest rate has a negative impact on stock returns. Figure 4.1 confirms this relationship, since it shows an almost linear declining stock return for an increasing risk-free rate. An increase in the interest rate could raise financial costs, followed by a reduce of future corporate profitability and stock prices. Also the findings for earnings by price agree with the theory. A growing earnings-price ratio makes firms more interesting for investors, and thus stock returns should also increase, as can bee seen in the left part of Figure 4.1.

Motivated by this results that both, earnings and risk-free, explain to some extent stock returns, we broaden in the next subsection our model to the multivariate case.

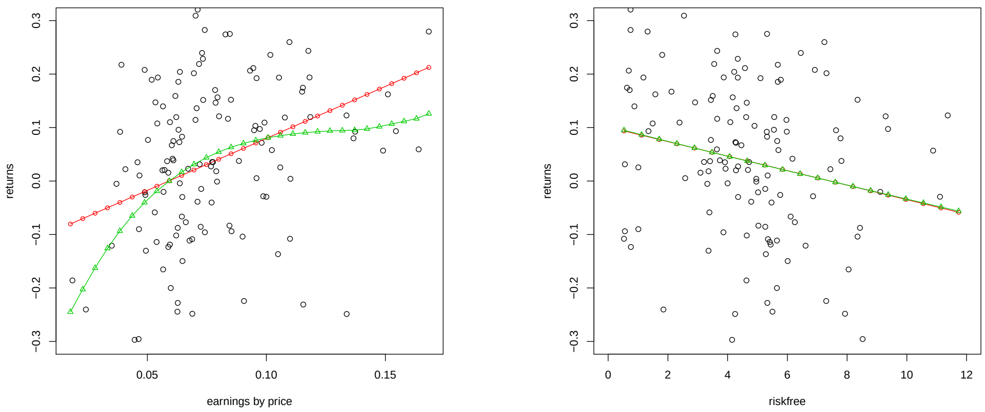
4.2.5 The Multivariate Parametric Model

The natural extension of model (4.4) is

$$Y_t = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X}_{t-1} + \varepsilon_t, \tag{4.6}$$

where $\boldsymbol{X}_{t-1}$ can be a vector of different explanatory variables, higher order terms, interactions of certain variables, or a combination of them. But again, we concentrate on the simple case, i. e. we use only two different regressor variables in (4.6) for creating a simple prior. Table 4.3 shows the results, the *validated* and the adjusted

Figure 4.1: *Left: stock returns and earnings by price, Right: stock returns and risk-free; both estimated with linear model* (4.4) *(circles) and nonlinear model* (4.5) *(triangles)*



R^2, for the regression of lagged earnings by price together with another variable on stock returns. We find again that the size of both measures is comparable. Moreover, the additional variables inflation, bond yield, and risk-free rate further improve the prediction, compared to the simple model (4.4) with earnings by price as unique explanatory variable, due to R_V^2 values greater than 8.0. In particular, even the one-dimensional nonparametric model (4.5) with earnings by price as covariate is outperformed by the multivariate linear model (4.6) using earnings by price and the risk-free rate as regressors. Here we find a R_V^2 of 12.2 instead of 11.8 for the former one.

Table 4.3: *Predictive power of the two-dimensional linear model* (4.6).

	e,S	e,d	e,r	e,L	e,inf	e,b
R_V^2	6.8	6.9	12.2	7.3	9.2	8.8
R_{adj}^2	8.5	8.7	13.9	8.7	10.7	10.0

Lagged explanatory variables: e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

4.3 Nonparametric Prediction Guided by Prior Knowledge

4.3.1 The Fully Nonparametric Model

To allow the use of more than one explanatory variable in a flexible nonparametric way, we consider the conditional mean equation

$$Y_t = g(\boldsymbol{X}_{t-1}) + \xi_t, \tag{4.7}$$

where the vector $\boldsymbol{X}_{t-1}$ includes now different regressor variables[30]. Table 4.4 gives the results, the *validated* R^2 and the estimated p-value of the proper bootstrap test, using again earnings by price together with another explanatory variable. Here, we find evidence that the appropriate functional form is nonlinear. For all these models we reject at the usual significance levels the null hypothesis that the true model would be the simple historical mean. Moreover, we find again for all models improved stock return predictions compared to the multivariate linear model (4.6) because all R_V^2 values are significantly higher. The best model at the moment is the fully two-dimensional one using earnings by price and the risk-free rate, resulting in a R_V^2 value of 13.7, what is a remarkable increase in predictive power of 12% compared to the parametric counterpart.

Here, we only apply two-dimensional models because more complex[31] nonparametric models would not end in better results. Typically, such settings are faced with essential difficulties, like the *curse of dimensionality*, boundary or bandwidth problems. We will see in the following how it is possible to circumvent or at least to reduce them in the combination of strategies that are usually applied individually.

4.3.2 Improved Smoothing through Prior Knowledge

In this subsection, we include prior information in our analysis. This could be, for example, a regression model coming from empirical data analysis or statistical modeling, or likewise a good economic model provided by the clever economist. We restrict ourselves to the former because already the use of such simple pilot estimates helps to improve the prediction of stock returns as we will demonstrate in the following.

[30]Different authors, for example, McMillan (2001), include here only the significant exogenous variables identified in the linear model (4.6). Although insignificant variables in a linear framework could be significant in nonlinear models, they follow Granger and Teräsvirta (1993) and argue that if the true data generating process is nonlinear then applying a linear model would result in an overfit of the data and in more significant parameters than required by the correct nonlinear specification.

[31]In the sense of more than two dimensions.

Table 4.4: *Predictive power of the fully two-dimensional nonparametric model (4.7) and corresponding estimated p-values of the bootstrap test.*

	e, S	e, d	e, r	e, L	e, inf	e, b
R_V^2	8.5	12.6	13.7	11.0	11.0	11.3
p-value	0.003	0.003	0.000	0.004	0.000	0.000
τ	0.307	0.503	0.485	0.383	0.452	0.430
p-value	0.045	0.018	0.005	0.021	0.011	0.001

Lagged explanatory variables: e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

The basic idea—see, for example, the well written paper of Glad (1998)—is the combination of the parametric pilot from model (4.4) or (4.6) and the nonparametric smoother from Subsections 4.2.4 or 4.3.1 in a semiparametric fashion, where the latter nonparametric estimator is multiplicatively guided by the former parametric and builds on the simple identity[32]

$$g(x) = g_\theta(x) \cdot \frac{g(x)}{g_\theta(x)}. \tag{4.8}$$

The essential fact is that if the prior captures some of the characteristics of the shape of $g(x)$, the second factor in (4.8) becomes less variable than the original $g(x)$ itself. Thus a nonparametric estimator of the *correction factor* $\frac{g(x)}{g_\theta(x)}$ gives better results with less bias.

Note again, that the global pilot could be generated by any parametric technique including simple linear methods, by more complex approaches like nonparametric regression[33] or regression splines with few knots, but also by well-founded economic theory. However, very often even a simple and rough parametric guide is enough to improve the estimate.

From (4.8) it is obvious that local problems for the above guided approach can occur if the prior itself crosses the x-axis one or more times. Two possible solutions are usually described in the literature. First, a suitable truncation is proposed, i. e. clipping the absolute value of the correcting factor, for example, below $1/10$ and above 10 makes the estimator more robust. Second, one could shift all response data

[32]Remember that we address the regression problem of estimating the conditional mean function $g(x) = E(Y|X = x)$, utilizing its standard solution, the fit of some parametric model $g_\theta(x)$, with the parameter θ, to the data. Of course, the prior should be estimated consistently.

[33]For the multiplicative bias correction in nonparametric regression, see Linton and Nielsen (1994).

Y_i a distance c in such a way that the new prior $g_\theta(x) + c$ is strictly greater than zero and does not anymore intersect the x-axis:

$$g(x) + c = (g_\theta(x) + c) \cdot \frac{g(x) + c}{g_\theta(x) + c}. \tag{4.9}$$

Note that the estimator becomes for increasing size of c more and more equal to the usual local polynomial which is invariant to such shifts, so that large values of c resolve the intersection problem, but diminish the effect of the guide.

Of course, parameter estimation variability also affects the result, but Glad (1998) shows that there is actually no loss in precision caused by the prior. Even for clear misleading guides she reports the tendency of ignoring the incorrect information and to end up with results similar to that one produced by the fully nonparametric estimator. Also in small samples the guided estimator has strong bias reducing properties. In her experiments, all not too unreasonable guides significantly reduce the bias for all sample sizes and level of noise.

Mainly in the multivariate version, this approach can improve prediction. The reason for it lies in the fact that traditional nonparametric estimators, like the in Section 4.3.1 presented one, have a rather slow rate of convergence in higher dimensions[34]. Also for a guided multivariate kernel estimator the possibility for bias reduction is essential if the parametric guide captures important features[35] of $g(x)$. Thus, the idea of guided nonparametric regression turns out to be even more helpful in such a setting.

It is also possible to interpret equations (4.8) or (4.9) as an optimal transformation of the nonparametric estimation problem. The subsequent nonparametric smoother of the transformed variables, i. e. of the *correction factor*, is characterized by less bias. For simple transformation techniques that improve nonparametric regression, see, for example, Park et al. (1997).

Table 4.5 shows the results[36], i. e. the *validated* R_V^2, of models based on (4.9) which use earnings by price together with another explanatory variable. The same variables are used to generate the simple linear prior with model (4.6) and to estimate the *correction factor*. We find that for earnings by price together with dividend by price as well as long term interest rate our strategy helps to improve the prediction power.

[34] It is a well known fact that the rate of convergence decreases dramatically for higher dimensions.

[35] Note that in the conditional asymptotic bias of the multivariate local-linear estimator the hessian of the true function appears. But for a "quasi linear" *correction factor* produced by a very good prior, the second derivatives should be very small and thus also the bias.

[36] We do not show the results of the bootstrap test for the following models guided by a prior because we will see that those models result with further improved R_V^2 than the fully nonparametric models (we have already seen in the applied bootstrap tests that the fully nonparametric models are significantly better than the simple historical mean).

Compared to the fully two-dimensional model (4.7) with the same variables, for the former we find a still not satisfying increase of the *validated* R^2 of 7%, and a notable one of 16% for the latter. For the other variables, our quality measure for the prediction decreases slightly. The reason for this lies in a poor prior or in the fact that the fully two-dimensional smoother already estimates the unknown relationship between stock returns and the used explanatory variables adequately.

Table 4.5: *Predictive power for model* (4.9).

	e, S	e, d	e, r	e, L	e, inf	e, b
R_V^2	6.6	13.5	12.1	12.8	9.5	8.0

In both steps, the prior and estimation of the correction factor, used lagged explanatory variables: e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

4.3.3 Prior Knowledge for Dimension Reduction

As discussed in the previous subsections, fully nonparametric models suffer in several aspects, with increasing number of dimensions, from the *curse of dimensionality*, and are faced with bandwidth or boundary problems. Since this type of estimator is based on the idea of local weighted averaging, the observations are sparsely distributed in higher dimensions causing unsatisfactory performance. To circumvent this, it is often proposed to import more structure in the estimation process, like additivity (cf. Stone (1985)) or semiparametric modelling. But, these are not the only possible solutions. Here, our in Section 4.3.2 proposed approach can also help to import more structure and reduce dimensionality in a multiplicative way. For example, instead of using for both, prior and nonparametric smoother of the *correction factor*, a two dimensional model, we reduce both to one-dimensional problems, but with different explanatory variables. For this, we first generalize (4.9) and concentrate on the analog identity

$$g(x_1) + c = (g_\theta(x_2) + c) \cdot \frac{g(x_1) + c}{g_\theta(x_2) + c}. \tag{4.10}$$

Please keep in mind that this is a separable model of x_1 and x_2. The results of this approach can be found in Table 4.6. Here, we use the simple linear parametric model (4.4) with different variables for the prior step. After that we estimate the *correction factor* with the one-dimensional nonparametric model (4.5) and earnings by price as covariate. Four of the six in Table 4.6 presented models improve stock return prediction, as we can observe an increased R_V^2 compared to the fully two-dimensional models from Subsection 4.3.1. For example, a simple linear prior with the risk-free rate and nonparametric smoother with earnings by price gives a *validated* R^2 of 15.8, a remarkable increase of 15% compared to our best model so far, the fully two-dimensional one with exact the same variables.

Table 4.6: *Predictive power for dimension reduction using identity* (4.10).

	S	d	r	L	inf	b
e	8.8	7.6	15.8	10.7	11.4	11.8

The prior is generated by a one-dimensional linear regression (4.4) *and uses as lagged explanatory variables S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor is estimated as in model* (4.5) *using only e earnings by price.*

The estimated functions for both models, the fully two-dimensional one (4.7) and the model guided by prior[37] with (4.10), as well as for the simple parametric counterpart are shown in Figure 4.2. Note that we fix one variable at a certain level and plot the relationship of stock returns with the remaining variable. On the left hand side of Figure 4.2, we fix the risk-free rate at values of 1.0, 6.0, and 12.0. For example, we see that the estimated function, which is guided by the prior, always forecasts negative stock returns for very high earnings by price. In contrast, the parametric and fully nonparametric fit show positive increasing stock returns for earnings by price from a value of 0.11. On the right hand side of Figure 4.2, we fix earnings by price at 0.03, 0.05, and 0.13. All displayed estimates are more or less linear and find at all levels of earnings by price a linear relationship between stock returns and the risk-free rate. Again, the negative impact of the risk-free rate on stock returns can bee seen. Only for a small earnings-price ratio, the estimator guided by the prior results in an almost constant line, what means that for small earnings by price the risk-free rate has no, or only a small, impact on stock returns.

Note further, that the approach with the prior results in a better fit in the boundary region compared to the fully nonparametric one[38], and thus in more reliable results. The reason for this lies again in the different number of dimensions used for the nonparametric part of the estimators.

[37] As just described, we use the simple linear prior (4.4) with the risk-free rate as regressor.

[38] Boundary effects are quite common in nonparametrics as well as different approaches to circumvent them.

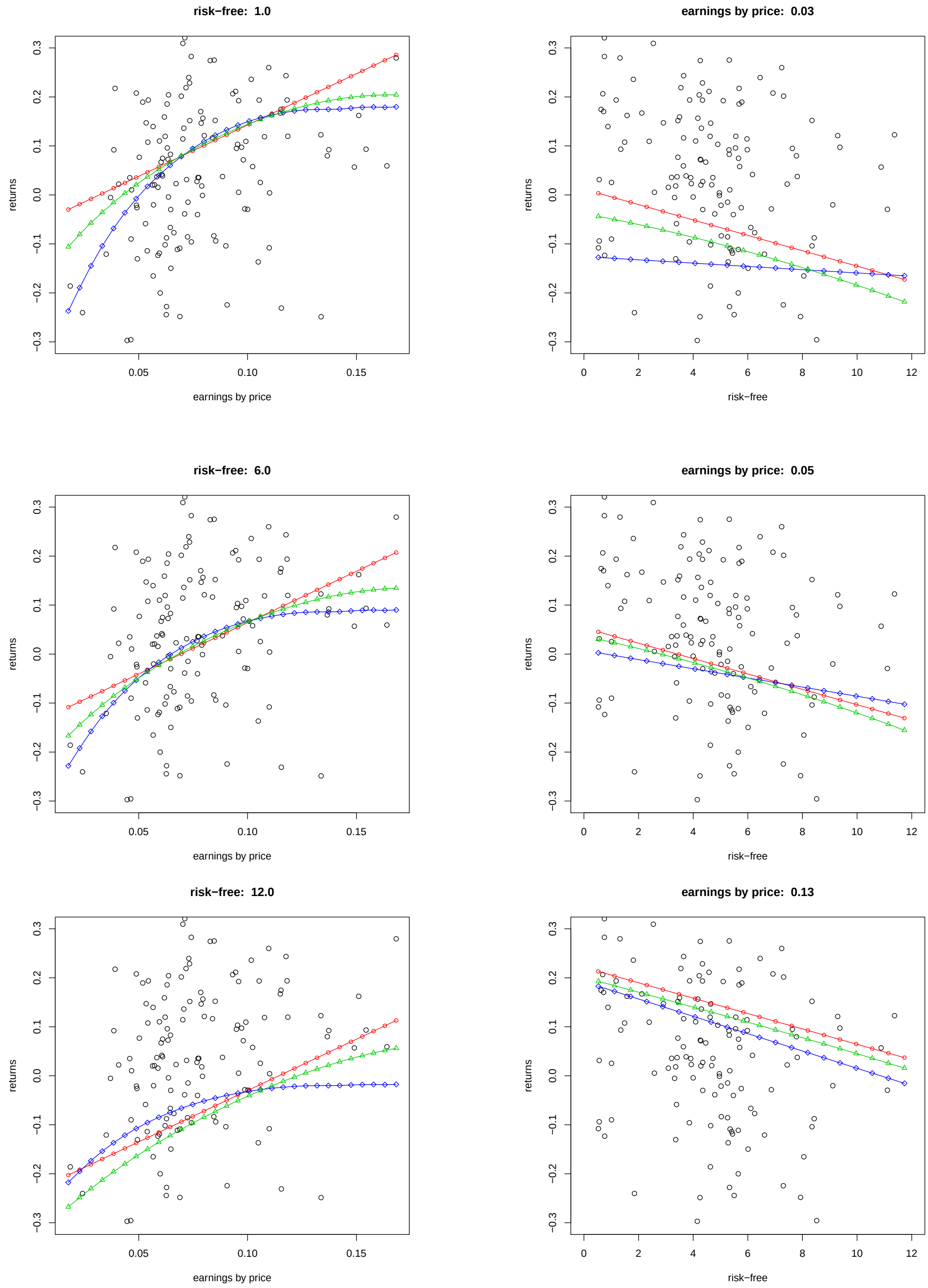


Figure 4.2: *Left: stock returns and earnings by price at different levels of risk-free, Right: stock returns and risk-free at different levels of earnings by price; both estimated with simple linear model* (4.6) *(circles), fully nonparametric model* (4.7) *(triangles), and the model guided by prior* (4.10) *(diamonds). The simple linear model* (4.4) *with the risk-free rate as regressor is used to generate the prior.*

4.3.4 Extensions to Higher Dimensional Models

The above approach can easily be extended in several ways. Here, we consider higher dimensions for x_1 and x_2 in (4.10) with possible overlapping covariates. For example, we could also use a two-dimensional linear prior in (4.10) and still estimate the *correction factor* with a one-dimensional nonparametric model. This results again in an improvement because we find a *validated* R^2 of 16.1 for the model that uses earnings by price and the risk-free rate for the simple linear prior and only earnings in the nonparametric step, as can be seen in Table 4.7. This is again a notable increase in predictive power of 18% compared to the best fully nonparametric model.

Table 4.7: *Predictive power for dimension reduction using identity* (4.10).

	e, S	e, d	e, r	e, L	e, inf	e, b
e	7.5	8.5	16.1	9.9	10.2	10.1

The prior is generated by a two-dimensional linear regression (4.6) *and uses as lagged explanatory variables* e *earnings by price together with* S *stock return,* d *dividend by price,* r *risk-free rate,* L *long-term interest rate,* inf *inflation, and* b *bond yield. The correction factor is estimated as in model* (4.5) *using only* e *earnings by price.*

The other way around is possible too. We use the simple one-dimensional parametric prior (4.4) together with a fully two-dimensional nonparametric smoother. In the application of this method, we find the results presented in Table 4.8. For example, using in the simple linear prior step the risk-free rate and in the nonparametric smoother earnings by price and the long-term interest rate, we find an R_V^2 of 18.5, an improvement of impressive 35% compared to the nonparametric model without prior, or an increase of 131% compared to the simple predictive regression, the starting point of our analysis. Also a simple linear prior with the long-term interest rate, together with earnings by price and again long-term interest rate in the nonparametric step, improves the prediction power by remarkable 29% compared to the fully nonparametric version of the model. This results are in accordance with economic theory since the most important part of the stock return is related to the change in interest rates and earnings.

In the above examples, we have seen that the simple extension to identity (4.10) combines transformation, bias and dimension reduction techniques in a new way and in a *single* approach, in contrast to the usual proposed separable or additive structures. Thus, boundary and bandwidth problems are easily alleviated and the

Table 4.8: *Predictive power for dimension reduction using identity* (4.10).

	S	d	e	r	L	inf	b
e, L	9.9	13.1	14.2	18.5	13.3	13.3	11.4

The prior is generated by a one-dimensional linear regression (4.4) *and uses as lagged explanatory variables S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor is estimated as in model* (4.7) *using e earnings by price and L long-term interest rate as covariates.*

curse of dimensionality circumvented.

4.4 Further Remarks and Conclusions

4.4.1 Wider Results

Up to now, we concentrated in this chapter on models which involved the variable earnings by price. Of course, we used other explanatory variables too. The results of such models can be found on the analogy to previous representations in Table 4.10–4.14 in the appendix. There, we also give a short overview of the used data. Table 4.9 presents summary statistics of the available variables. Note that we calculate the inflation variable as the percentage change of the consumer price index and the bond variable as the difference of the ten-year government bond.

As Table 4.10 and 4.11 indicate, it is hard to find a model that can better predict than the simple historical mean. But it is not surprising that, once we find such a model, the risk-free rate is an important part of it. For example, we find for the fully nonparametric model, risk-free rate together with dividend by price ($R_V^2 = 3.0$) or long-term interest rate ($R_V^2 = 8.5$), *validated* R^2 values that are significantly different from zero. However, these models do not have the predictive power found before for the model that uses earnings by price and risk-free ($R_V^2 = 13.7$).

In Table 4.12–4.14, we include the already shown results (for earnings by price) for reasons of clarity and comparability. We find that earnings by price consistently gives the best results[39], together with the interest rates. Moreover, we see that more complex models do not automatically imply better results. For example, if we use the simple linear prior (4.4) with the risk-free rate and estimate the *correction factor* along (4.10) with model (4.5) and earnings by price as covariate (see third line in

[39] In the sense of the largest R_V^2 value.

Table 4.12), we obtain a *validated* R^2 of 15.8. On the other side, if we also include the risk-free rate when we estimate the correction factor, i. e. with the more complex model (4.7), we get only a R_V^2 of 10.7 (see third line in Table 4.14). Furthermore, we stress again that the choice of the prior is crucial. This can bee seen, for example, in line three of Table 4.13, where we estimate the *correction factor* with model (4.5) and earnings by price as covariate. The use of the simple prior (4.6) with earnings by price and dividend by price gives a R_V^2 of 8.5, while we nearly double ($R_V^2 = 16.1$) the result if we take the same prior but the risk-free rate instead of dividend by price.

4.4.2 Summary and Outlook

The objective of this chapter is to show that the prediction of excess stock returns can essentially be improved by the approach of flexible non- and semiparametric techniques. We start with a fully nonparametric model and estimate this with a standard local-linear kernel regression, whereas we maximize the *validated* R^2 for the choice of the best model and the bandwidth. We further propose a simple wild-bootstrap test which allows us to decide whether we can accept the parametric null hypothesis, that the historical mean is the right model, or whether we prefer the non- or semiparametric alternative. After we have seen the usefulness of the nonparametric approach, we introduce a possibility to include prior knowledge in the estimation procedure. This can be, for example, a good economic model or likewise a simple parametric regression. We indicate, that even the inclusion of the latter in a semiparametric fashion, more precisely, in a multiplicative way, can enormously improve the prediction of stock returns. To illustrate the potential of our method, we apply it to annual American stock market data, which are provided by Robert Shiller and used for several other articles. Our results confirm to economic theory, namely that the most important part of stock returns is related to the change in interest rates and earnings.

To deliver a statistically insight into our method, we mention that, mainly in higher dimensions, a nonparametric approach would suffer from the *curse of dimensionality*, bandwidth or boundary problems. A possible adjustment for this problem is the imposition of more structure. Our method contributes to this strategy due to its new and innovative idea—a model directly guided by economic theory. We achieve by a simple transformation the combination of bias and dimension reduction, i. e. more structure to circumvent the *curse of dimensionality*. This means in our case that a reliable prior captures some of the characteristics of the shape of the estimating function, and thus a multiplicative correction can cause a bias and dimension reduction in the remaining nonparametric estimation process of the *correction factor*. Thus, we present here a method which greatly improves nonparametric regression in combination with a simple parametric technique.

An other possibility to impose more structure in the prediction process of excess stock returns could be the use of same years covariates. Usually, economic theory says that the price of a stock is driven by fundamentals and investors should focus on forward earnings and profitability. Thus, information on same years, instead of last years, earnings or interest rates can improve prediction. The problem which obviously occurs is that this information is unknown and must also be predicted in some way[40]. Furthermore, one should also take into account structural breaks or calendar effects. As already mentioned, also longer horizons are easily included in the analysis. Here, a possible improvement could be an error-correction method, like the one described in Bansal and Kiku (2011).

[40] Cf. the article of Scholz, Sperlich, and Nielsen (2011), where a two-step procedure for the inclusion of the same years bond yield is proposed, which is related to the change in interest rates.

4.5 Appendix: Tables of Additional Results

Table 4.9: *US market data (1872-2009).*

	Max	Min	Mean	Sd
S&P Stock Price Index	1479.22	3.25	165.08	345.39
Dividend Accruing to Index	28.39	0.18	3.96	6.27
Earnings Accruing to Index	81.51	0.16	8.69	15.54
Stock Returns	0.44	-0.62	0.04	0.18
Dividend by Price	0.09	0.01	0.05	0.01
Earnings by Price	0.17	0.02	0.08	0.03
Short-term Interest Rate	17.63	0.53	4.77	2.77
Long-term Interest Rate	14.59	1.95	4.67	2.27
Inflation	0.21	-0.16	0.02	0.06
Bond	2.03	-4.13	-0.02	0.77

Table 4.10: *Predictive power of the two-dimensional linear model* (4.6).

	S,d	S,r	S,L	S,inf	S,b	d,r	d,L	d,inf
R^2_V	0.7	1.1	-2.2	-2.3	-1.6	3.5	-0.1	-0.4
R^2_{adj}	2.5	3.4	-0.4	-0.1	-0.1	5.0	1.0	1.3

	d,b	r,L	r,inf	r,b	L,inf	L,b	inf,b
R^2_V	0.8	7.7	1.2	1.5	-2.5	-1.5	-1.9
R^2_{adj}	1.8	8.6	2.9	2.9	-1.1	-0.8	-0.7

Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

Table 4.11: *Predictive power of the fully two-dimensional nonparametric model* (4.7) *and corresponding estimated p-values of the bootstrap test.*

	S, d	S, r	S, L	S, inf	S, b	d, r	d, L	d, inf
R_V^2	0.3	0.5	-2.3	-2.4	-2.1	3.0	-0.3	-0.7
p-value	0.192	0.206	0.686	0.625	0.556	0.043	0.226	0.312
τ	0.123	0.165	0.039	0.050	0.052	0.202	0.079	0.086
p-value	0.240	0.185	0.589	0.574	0.499	0.069	0.233	0.369

	d, b	r, L	r, inf	r, b	L, inf	L, b	inf, b
R_V^2	0.3	8.5	0.7	1.9	-2.5	-1.6	-2.4
p-value	0.131	0.002	0.161	0.071	0.811	0.625	0.718
τ	0.123	0.319	0.140	0.144	0.014	0.026	0.031
p-value	0.082	0.013	0.186	0.101	0.818	0.659	0.665

Lagged explanatory variables: S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, b bond yield.

Table 4.12: *Predictive power for dimension reduction using identity* (4.10).

	S	d	e	r	L	inf	b
S	-4.8	-1.0	5.2	-0.5	-3.8	-3.9	-3.2
d	-1.0	-2.1	5.6	1.9	-1.8	-2.0	-0.6
e	8.8	7.6	9.3	15.8	10.7	11.4	11.8
r	-0.6	1.8	10.8	-0.8	-1.2	-1.0	-0.1
L	-3.5	-1.7	5.9	2.3	-3.7	-3.8	-2.6
inf	-3.9	-2.0	8.8	-0.5	-4.3	-4.9	-3.6
b	-3.4	-1.0	7.6	0.7	-3.3	-3.8	-3.3

The prior (columns) is generated by a one-dimensional linear regression (4.4) *and the correction factor (rows) is estimated as in model* (4.5). *Both use as lagged explanatory variables S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield.*

Table 4.13: *Predictive power for dimension reduction using identity* (4.10).

	e, S	e, d	e, r	e, L	e, inf	e, b
S	3.5	3.9	9.0	4.4	6.9	5.8
d	4.2	4.1	10.7	5.3	8.0	6.5
e	7.5	8.5	16.1	9.9	10.2	10.1
r	9.0	10.5	9.1	7.5	10.3	10.1
L	4.5	5.4	11.2	4.3	6.7	6.4
inf	7.9	8.8	11.0	7.4	7.8	8.3
b	6.2	6.9	11.4	6.6	7.9	6.0

The prior (columns) is generated by a two-dimensional linear regression (4.6) *and uses as lagged explanatory variables e earnings by price together with S stock return, d dividend by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor (rows) is estimated as in model* (4.5) *using only one of the explanatory variables.*

Table 4.14: *Predictive power for dimension reduction using identity* (4.10).

	S	d	e	r	L	inf	b
e, S	5.3	2.9	5.8	12.2	7.1	7.9	8.7
e, d	9.6	11.0	10.2	17.1	12.0	12.6	12.5
e, r	9.9	6.8	11.0	10.7	10.1	12.9	12.6
e, L	9.9	13.1	14.2	18.5	13.3	13.3	11.4
e, inf	7.9	5.7	8.9	13.0	9.5	8.7	10.2
e, b	8.3	5.6	8.5	15.9	9.9	10.6	9.1

The prior (columns) is generated by a one-dimensional linear regression (4.4) *and uses as lagged explanatory variables S stock return, d dividend by price, e earnings by price, r risk-free rate, L long-term interest rate, inf inflation, and b bond yield. The correction factor (rows) is estimated as in model* (4.7) *using e earnings by price together with another covariate.*

Chapter 5

Conclusions and Outlook

This thesis deals with the inclusion of economic prior knowledge in the statistical modelling process. Widely established non- and semiparametric approaches, namely the local-polynomial smoother and the wild bootstrap, are adequately adapted to particular economic problems. A major objective of research in this thesis is to show that, by suitable modification, these popular techniques can perform well in different economic fields. The reason for this lies in the fact that our proposed methods address the *curse of dimensionality and complexity.* In contrast to additivity or separability, which are often recommended in the statistical literature, this thesis makes use of the semiparametric nature of economic problems to reduce dimensionality. It directly takes advantage of the structure that is inherent in the economic process that generates the data. Prior knowledge not only improves the plausibility of the models but also the interpretability of their results. It can also be used to address some other problems that widely occur by the use of fully nonparametric approaches. For example, the estimation accuracy on boundaries can be improved, and the bias can be reduced by applying a semiparametric approach. This final chapter reviews the main ideas and results of the thesis. It also suggests possible extensions and directions in which further research can be carried out.

Chapter 2 considers a semiparametric approach to investigate consumer demand. Since the starting point is a model of indirect utility, rationality restrictions are easily imposed. Economic theory yields a model for expenditure shares that is given by a fraction whose numerator is partially linear, but whose denominator comprises the derivative of the numerator. The key point is the achieved dimension reduction in the fully nonparametric part of the model which has only one dimension. The model thus contributes to the methods that address the *curse of dimensionality.* In Chapter 2 a new asymmetric version of the wild bootstrap is introduced for inference. The bootstrap residuals are generated in a special way to take into account that expenditure shares cannot lie outside of the interval $[0, 1]$. A simulation study shows the potential of the method because it is possible to capture shapes, for example with a flat plateau or a bump, that are usually difficult to estimate. It further approves that the bootstrap procedure achieves an acceptable level of accuracy. The application of the method to Canadian price and expenditure data suggests that some expenditure shares exhibit remarkable degrees of nonlinearity. Some interesting challenges are still open for further research. Although the modest Monte Carlo study verified that the proposed techniques work, the mathematical justification for the approach and the properties of the estimator are of interest, for example, for the direct calculation of confidence bands. Since the denominator in the fraction of the model is the same for all budget share equations, it could also be possible to use this fact for the development of a slightly different, but even more efficient and robust algorithm. Moreover, the difficult *global regularity* problem may be addressed. The restriction of the estimated budget shares to be everywhere in the interval $[0, 1]$ should be directly considered in the estimation algorithm. Finally, in the empirical part it would be worth investigating the social welfare by the calculation, for example, of the cost of living impacts as a consequence of price changes.

Chapters 3 and 4 investigate the intensively debated question of whether equity returns (or premiums) can be predicted by empirical models. While many articles in the financial and actuarial literature favor the historical mean, or other simple parametric methods, both chapters focus on more sophisticated techniques. These approaches not only outperform the simple historical mean, they also result in a greatly improved prediction quality compared to fully nonparametric versions of the models. This could be achieved by the imposition of more structure in the estimation process which is an appropriate way to circumvent the *curse of dimensionality.*

Economic theory and statistical arguments motivate in Chapter 3 the inclusion of the same years bond yield in the fully nonparametric prediction approach for excess stock returns. Since the current bond yield is unknown, it is predicted in a prior step also with a fully nonparametric method. Chapter 3 thus presents the theoretical justification for the use of *constructed* variables in the nonparametric regression when the time series data are *dependent.* The proposed method is implemented on Danish stock and bond market data and shows, as already mentioned, a remarkable improvement of the prediction quality, as measured by the *validated* R^2. The results confirm the economic motivation for the inclusion of the same years bond yield. It captures the most important part of the stock return, namely the part related to the change in interest rates. In addition, the first step of the method (the construction of the current bond yield) can be seen as an optimal transformation of the predictors. This can significantly improve the prediction of stock returns because the nonparametric smoother of the transformed variables is characterized by less bias.

The starting point of Chapter 4 is the fully nonparametric model that allows a flexible form of the conditional mean function. A straightforward bootstrap test confirms that non- and semiparametric methods yield better forecasts than do parametric models, and shows therefore the usefulness of more sophisticated techniques. In Chapter 4 a new approach is introduced to include prior knowledge in the forecasting procedure of excess stock returns. Economic theory directly guides the modelling process. The available prior information is included in a semiparametric manner, where the nonparametric smoother is multiplicatively guided by the prior. The application of economic theory or the examination of standard parametric models leads to the necessary prior, whose choice is essential. In consequence of this approach a dimension and bias reduction is achieved, both to impose more structure to circumvent the *curse of dimensionality*. It can be shown that certain boundary and bandwidth difficulties, which could occur in fully nonparametric approaches, are thereby overcome in a *single* idea. To illustrate the potential of the proposed method, it is implemented on American stock market data. The results of the empirical part show a notable improvement of the prediction quality which is again measured by the *validated* R^2. As in Chapter 3, the findings confirm to economic theory. The most important part of the stock return is related to the change in interest rates and earnings.

Chapters 3 and 4 propose important and innovative techniques for the improvement of forecasts of excess stock returns. The aim of further research could be to investigate other possibilities that impose more structure in the estimation process. Economic theory usually indicates that the price of a stock is driven by fundamentals and that investors should focus on forward earnings and profitability. Thus, information on same years covariates, instead of last years data, as earnings or interest rates could further improve the prediction. Chapter 3 indicates a possible way how current regressors may be generated. The proposed approach could also be applied, for example, to adequate American, UK, or German data. Another worthwhile question is the investigation of longer horizons. Both chapters concentrate on the one-year period but are readily adapted to this problem. In doing so, attention should be paid in the cross-validation as described in detail in Chapter 3. A further improvement may be here the use of an error-correction method in spirit of the article of Bansal and Kiku (2011). Furthermore, structural breaks and calendar effects should be taken into account when predicting stock returns. It remains also for further research to apply the described techniques in related economic fields. For example, the Sharpe ratio is often calculated to measure the risk of an investment. Thus, the proposed methods could not only be extended to volatility but also combined with volatility to a forecast of risk.

Bibliography

ABHYANKAR, A., COPELAND, L. S., and WONG, W. (1997). Uncovering Nonlinear Structure in Real-Time Stock Market Indices. *Journal of Business and Economic Statistics* **15**, 1–14.

AMIHUD, Y., and HURVICH, C. M. (2004). Predictive Regression: A Reduced-Bias Estimation Method. *Journal of Financial and Quantitative Analysis* **39**, 813–841.

ANG, A., and BEKAERT, G. (2007). Stock Return Predictability: Is It There? *Review of Financial Studies* **20**, 651–707.

ASNESS, C. (2003). Fight the Fed Model: The Relationship Between Future Returns and Stock and Bond Market Yields. *Journal of Portfolio Management* **30**, 11–24.

BANKS, J., BLUNDELL, R., and LEWBEL, A. (1997). Quadratic Engel curves and consumer demand. *Review of Economics and Statistics* **79**, 527–539.

BANSAL, R., and KIKU, D. (2011). Cointegration and Long-Run Asset Allocation. *Journal of Business and Economic Statistics* **29**, 161–173.

BLUNDELL, R., CHEN, X., and KRISTENSEN, D. (2003). Semi-nonparametric IV estimation of shape-invariant Engel curves. Centre for Microdata Methods and Practice, working paper CWP15/03.

BLUNDELL, R., DUNCAN, A., and PENDAKUR, K. (1998). Semiparametric estimation of consumer demand. *Journal of Applied Econometrics* **13**, 435–461.

BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2nd ed. Springer, New York.

BOSSAERTS, P., and HILLION, P. (1999). Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *The Review of Financial Studies* **12**, 405–428.

BOUDOUKH, J. M. RICHARDSON, M. P. and WHITELAW, R. F.(2008). The Myth of Long-Horizon Predictability. *The Review of Financial Studies* **21**, 1577–1605.

Butler, A. W., Grullon, G., and Weston, J. P. (2005). Can Managers Forecast Aggregate Market Returns? *Journal of Finance* **60**, 963–986.

Campbell, J. Y. (1987). Stock Returns and the Term Structure. *Journal of Financial Economics* **18**, 373–399.

Campbell, J. Y. (1991). A Variance Decomposition for Stock Returns. *Economic Journal* **101**, 157–179.

Campbell, J. Y. and Cochrane, J. H. (1999). By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* **107**, 205–251.

Campbell, S. D., and Diebold, F. X. (2009). Stock Returns and Expected Business Conditions: Half a Century of Direct Evidence. *Journal of Business and Economic Statistics* **27**, 266–278.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Campbell, J. Y., and Shiller, R. J. (1987). Cointegration and Tests of Present Value Models. *Journal of Political Economy* **95**, 1067–1087.

Campbell, J. Y., and Shiller, R. J. (1988a). Stock Prices, Earnings, and Expected Dividends. *Journal of Finance* **43**, 661–676.

Campbell, J. Y., and Shiller, R. J. (1988b). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies* **1**, 195–228.

Campbell, J. Y., and Shiller, R. J. (1998). Valuation Ratios and the Long-Run Stock Market Outlook. *Journal of Portfolio Management* **1**, 195–227.

Campbell, J. Y., and Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies* **21**, 1455–1508.

Campbell, J. Y., and Viceira, L. M. (1999). Consumption and Portfolio Decisions when Expected Returns are Time-Varying. *Quarterly Journal of Economics* **144**, 433–495.

Campbell, J. Y., and Yogo, M. (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics* **81**, 27–60.

Cavanagh, C. L., Elliot, G., and Stock, J. H. (1995). Inference in Models With Nearly Integrated Regressors. *Econometric Theory* **11**, 1131–1147.

Chen, Q., and Hong, Y. (2009). Predictability of Equity Returns over Different Time Horizons: A Nonparametric Approach. Cornell University/Department of Economics. Working Paper.

Chiquoine, B., and Hjalmarsson, E. (2008). Jackknifing Stock Return Predictions. International Finance Discussion Papers. Number 932.

Chu, C. K., and Marron, J. S. (1991). Comparison of Two Bandwidth Selectors With Dependent Errors. *The Annals of Statistics* **19**, 1906–1918.

Clark, T. E. (2004). Can Out-of-Sample Forecast Comparisons Help Prevent Overfitting? *Journal of Forecasting* **23**, 115–139.

Cochrane, J. H. (1992). Explaining the Variance of Price-Dividend Ratios. *The Review of Financial Studies* **5**, 243–280.

Cochrane, J. H. (1999). New Facts in Finance. *Economic Perspectives* **23**, 36–58.

Deaton, A., and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review* **70**, 312–326.

Deschamps, P.J. (1988). A note on the maximum likehood estimation of allocation systems. *Computational Statistics & Data Analysis* **6**, 109–112.

Doukhan, P. (1994). *Mixing.* Springer, New-York.

Drobetz, W., and Hoechle, D. (2003). Parametric and Nonparametric Estimation of Conditional Return Expectations. WWZ/Department of Finance, Working Paper No. 9/03.

Engsted, T., and Tanggaard, C. (2001). The Danish Stock and Bond Markets: Comovement, Return Predictability and Variance Decomposition. *Journal of Empirical Finance* **8**, 243–271.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work . *Journal of Finance* **25**, 383–417.

Fama, E. F. (1990). Stock Returns, Expected Returns, and Real Activity. *Journal of Finance* **45**, 1089–1108.

Fama, E. F., and French, K. R. (1988). Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* **22**, 3–25.

Fama, E. F., and French, K. R. (1988). Permanent and Temporary Components of Stock Prices. *Journal of Political Economy* **96**, 246–273.

FAMA, E. F., and FRENCH, K. R. (1989). Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics* **25**, 23–49.

FAMA, E. F., and SCHWERT, G. W. (1977). Asset Returns and Inflation. *Journal of Financial Economics* **5**, 115–146.

FAN, J., and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

FAN, J., and YAO, Q. (2003). *Nonlinear Time Series.* Springer, New-York.

FAN, J., and ZHANG, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics* **27**, 1491–1518.

FERRATY, F., NÚÑEZ-ANTÓN, V., and VIEU, P. (2001). *Regresión No Paramétrica: Desde la Dimensión Uno Hasta la Dimensión Infinita.* Servicio Editorial da la Universidad del Pais Vasco, Bilbao.

FERSON, W. E., SARKISSIAN, S., and SIMIN, T. T. (2003). Spurious Regression in Financial Economics? *Journal of Finance* **58**, 1393–1413.

FRENCH, K. R., SCHWERT, G. W., and STAMBAUGH, R. F. (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics* **19**, 3–29.

FROOT, K., and OBSTFELD, M. (1991). Intrinsic Bubbles: The Case of Stock Prices. *American Economic Review* **81**, 1189–1217.

GLAD, I. K. (1998). Parametrically Guided Non-Parametric Regression. *Scandinavian Journal of Statistics* **25**, 649–668.

GOETZMANN, W. N., and JORION, P. (1993). Testing the Predictive Power of Dividend Yields. *Journal of Finance* **48**, 663–679.

GOETZMANN, W. N., and JORION, P. (1995). A Longer Look at Dividend Yields. *Journal of Business* **68**, 483–508.

GOETZMANN, W. N., and SANTA-CLARA, P. (2003). Idiosyncratic Risk Matters!. *Journal of Finance* **58**, 975–1007.

GOYAL, A., and WELCH, I. (2003). The Myth of Predictability: Does the Dividend Yield Forecast the Equity Premium? *Management Science* **49**, 639–654.

GOYAL, A., and WELCH, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* **21**, 1455–1508.

GRANGER, C. W. J., and TERÄSVIRTA, T. (1993). *Modelling nonlinear economic relationships.* Oxford University Press, Oxford.

GUO, H., and SAVICKAS, R. (2006). Idiosyncratic Volatility, Stock Market Volatility, and Expected Stock Returns. *Journal of Business and Economic Statistics* **24**, 43–56.

GYÖFRI, L., HÄRDLE, W., SARDA, P., and VIEU, PH. (1990). *Nonparametric Curve Estimation from Time Series*(Lecture Notes in Statistics). Springer, Heidelberg.

HAAG, B. R., HODERLEIN, S., and PENDAKUR, K. (2009). Testing and imposing Slutsky symmetry in nonparametric demand systems. *Journal of Econometrics* **153**, 33–50.

HÄRDLE, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Boston.

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

HÄRDLE, W., MÜLLER, M., SPERLICH, S., and WERWATZ A. (2004). *Nonparametric and semiparametric models.* Springer Verlag, Berlin.

HARVEY, C. R.(2001). The Specification of Conditional Expectations. *Journal of Empirical Finance* **8**, 573–638.

HASTIE, T. J., and TIBSHIRANI, R. J. (1984). Generalized aditive models. Department of Statistics, Stanford University, Technical Report 98.

HASTIE,T. J., and TIBSHIRANI, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

HODRICK, R. J.(1992). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. *The Review of Financial Studies* **5**, 357–386.

INOUE, A., and KILIAN, L. (2004). In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews* **23**, 371–402.

JANSSON, M., and MOREIRA, M. J. (2006). Optimal Inference in Regression Models With Nearly Integrated Regressors. *Econometrica* **74**, 681–714.

JORGENSEN, D., LAU, L., and STOKER, T. (1980). Welfare comparison and exact aggregation. *American Economic Review* **70**, 268–272.

KEIM, D. B., and STAMBAUGH, R. F. (1986). Prediction of Stock Returns in the Stock and Bond Markets. *Journal of Financial Economics* **17**, 357–390.

KILIAN, L. (1999). Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?. *Journal of Applied Econometrics* **14**, 491–510.

KIM, T. Y., PARK, B. U., MOON, M. S., and KIM, K. (2009). Using bimodal kernel for inference in nonparametric regression with correlated errors. *Journal of Multivariate Analysis* **100**, 1487–1497.

KOTHARI, S., and SHANKEN, J. (1992). Stock Return Variation and Expected Dividens: A Time-Series and Cross-Sectional Analysis. *Journal of Financial Economics* **31**, 177–210.

KOTHARI, S., and SHANKEN, J. (1997). Book-to-market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis. *Journal of Financial Economics* **44**, 169–203.

LAMONT, O. (1998). Earnings and Expected Returns. *Journal of Finance* **53**, 1563–1587.

LETTAU, M., and LETTAU, M. (2005). *Handbook of Financial Econometrics* Measuring and Modeling Variation in the Risk-Return Tradeoff.

LETTAU, M., and NIEUWERBURGH, S. V. (2008). Reconciling the Return Predictability Evidence. *The Review of Financial Studies* **21**, 1607–1652.

LEWBEL, A. (2000). Consumer demand systems and household equivalence scales, *Handbook of Applied Econometrics*, Volume II: Microeconomics. Blackwell Publishers Ltd, Oxford.

LEWBEL, A., and PENDAKUR, K. (2009). Tricks with Hicks: The EASI demand system. *American Economic Review* **99**, 827–863.

LEWELLEN, J. (2004). Predicting Returns With Financial Ratios. *Journal of Financial Economics* **74**, 209–235.

LINTON, O., and NIELSEN, J. P. (1994). A multiplicative bias reduction method for nonpametric regression. *Statistics and Probability Letters* **19**, 181–187.

LO, A. W., and MACKINLAY, A. C. (1990). Data-Snooping Biases in Tests of Financial Asset Pricing Models. *Review of Financial Studies* **3**, 431–467.

LUND, J., and ENGSTED, T. (1996). GMM and present value tests of the C_CAPM: Evidence from Danish, German, Swedish, and UK stock markets. *Journal of International Money and Finance* **15**, 497–521.

Mark, N. C. (1995). Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability. *American Economic Review* **85**, 201–218.

Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic Theory.* Oxford University Press, New York.

Masry, E., and Fan, J. (1997). Local Polynomial Estimation of Regression Functions for Mixing Processes. *Scandinavian Journal of Statistics* **24**, 165–179.

Mazzocchi, M. (2006). Time patterns in UK demand for alcohol and tobacco: an application of the EM algorithm. *Computational Statistics & Data Analysis* **50**, 2191–2205.

McMillan, D. G. (2001). Nonlinear Predictability of Stock Market Returns: Evidence from Nonparametric and Treshold Models. *International Review of Economics and Finance* **10**, 353–368.

Moral-Arce, I., Rodríguez-Póo, J. M., and Sperlich, S. (2007). Feasible semiparametric estimation of censored expenditure equations. Preprint, University of Göttingen, ZfS-2007-13.

Millimet, D., and Rusty T. (2008). Estimating demand systems and measuring consumer preferences, *Journal of Econometrics* **147**, 384–395.

Nelson, C., and Kim, M. (1993). Predictable Stock Returns: Reality or Statistical Illusion? *Journal of Finance* **48**, 641–661.

Nielsen, J. P., and Linton, O. (1998). An optimization interpretation of integration and back-fitting estimators for separable nonparametric models. *Journal od Royal Statistical Society* B **60**, 217–222.

Nielsen, J. P., and Sperlich, S. (2003). Prediction of Stock Returns: A new way to look at it. *Astin Bulletin* **33**, 399–417.

Orbe, S., Ferreira, E., and Rodríguez-Póo, J. M. (2003). An algorithm to estimate time-varying parameter SURE models under different types of restriction. *Computational Statistics & Data Analysis* **42**, 363–383.

Park, B. U., Kim, W. C., Ruppert, D., Jones, M. C., Signorini, D. F., and Kohn, R. (1997). Simple Transformation Techniques for Improved Nonparametric Regression. *Scandinavian Journal of Statistics* **24**, 145–163.

Pendakur, K. (2002). Taking prices seriously in the measurement of inequality. *Journal of Public Economics* **86**, 47–69.

Pendakur, K., Scholz, M., and Sperlich, S. (2010). Semiparametric indirect utility and consumer demand. *Computational Statistics and Data Analysis* **54**, 2763–2775.

Pendakur, K., and Sperlich, S. (2010). Semiparametric estimation of consumer demand systems in real expenditure. *Journal of Applied Econometrics* **25**, 420–457.

Perez-Quiros, G., and Timmermann, A. (2000). Firm Size and Cyclical Variations in Stock Returns. *Journal of Finance* **55**, 1229–1262.

Polk, C., Thompson, S., and Vuolteenaho, T. (2006). Cross-Sectional Forecasts of the Equity Premium. *Journal of Financial Economics* **81**, 101–141.

Pollack, R., and Wales, T. (1991). *Demand system specification and estimation.* Oxford University Press, New York.

Ponti, J., and Schall, L. D. (1998). Book-to-Market Ratios as Predictors of Market Returns. *Journal of Financial Economics* **49**, 141–160.

Poterba, J., and Summers, L. (1988). Mean Reversion in Stock Returns: Evidence and Applications. *Journal of Financial Economics* **22**, 27–60.

Qi, M. (1999). Nonlinear Predictability of Stock Returns Using Financial and Economic Variables. *Journal of Business and Economic Statistics* **17**, 419–429.

Racine, J. (2001). On the Nonlinear Predictability of Stock Returns Using Financial and Economic Variables *Journal of Business and Economic Statistics* **19**, 380–382.

Rapach, D. E., and Wohar, M. E. (2006). In-Sample vs. Out-of-Sample Tests of Stock Return Predictabilty in the Context of Data Mining. *Journal of Empirical Finance* **13**, 231–247.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies* **23**, 821–862.

Rey, D. (2004). Stock Market Predictability: Is it There? A Critical Review. WWZ/Department of Finance, Working Paper No. 12/03.

Rio, E. (2000). Théorie asymptotique pour des processus aléatoires faiblement dépendants. In *SMAI, Mathématiques et Applications* **31**. Springer, Berlin.

Roy, R. (1947). La Distribution du Revenu Entre Les Divers Biens, *Econometrica* **15**, 205-225.

Rozeff, M. (1984). Dividend Yields Are Equity Risk Premiums. *Journal of Portfolio Management* **11**, 68–75.

Sarmiento, C. (2005). A varying coefficient approach to global flexibility in demand analysis: a semiparametric approximation. *American Journal of Agricultural Economics* **87**, 38–47.

Scholz, M., Sperlich, S., and Nielsen, J. P. (2011). Nonparametric Prediction of Stock Returns With Generated Bond Yields. *Working Paper*.

Scholz, M., Nielsen, J. P., and Sperlich, S. (2011). Nonparametric Prediction of Stock Returns Guided by Prior Knowledge. *Working Paper*.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shiller, R. J. (1989). *Market Volatility*. MIT-Press, Cambridge.

Shiller, R. J., and Beltratti, A. E. (1992). Stock Prices and Bond Yields. *Journal of Monetary Economics* **30**, 25–46.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Sperlich, S. (2009). A note on nonparametric estimation with predicted variables. *The Econometrics Journal* **12**, 382–395.

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

Stambaugh, R. (1999). Predictive Regressions. *Journal of Financial Economics* **54**, 375–421.

Stone, C.J. (1985). Additive Regression and Other Nonparametric Models. *Annals of Statistics* **13**, 689–705.

Torous, W., and Valkanov, R. (2000). Boundaries of Predictability: Noisy Predictive Regressions. Working Paper, UCLA.

Torous, W., Valkanov, R., and Yan, S. (2004). On Predicting Stock Returns with Nearly Integrated Explanatory Variables. *Journal of Business* **77**, 937–966.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics* **68**, 201–232.

Varian, H. R. (1978). *Microeconomic Analysis.* Norton, New York.

West, K. (1988). Dividend Innovations and Stock Price Volatility. *Econometrica* **56**, 37–61.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica* **68**, 1097–1126.

Wolf, M. (2000). Stock Returns And Dividend Yields Revisited: A New Way To Look At An Old Problem. *Journal of Business and Economic Statistics* **18**, 18–30.