

CONTRIBUTIONS TO FUNCTIONAL DATA
ANALYSIS WITH APPLICATIONS TO
MODELING TIME SERIES AND PANEL DATA

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2013

vorgelegt von

Dipl.-Vw. Dominik Liebl

aus

Passau

Referent: Prof. Dr. Karl Mosler (Universität zu Köln)

Korreferent: Prof. Dr. Alois Kneip (Universität Bonn)

Tag der Promotion: 21. Aug. 2013

Contents

Contents	i
List of Figures	v
Introduction	1
1 Modeling and Forecasting Electricity Spot Prices: A Functional Data Perspective	7
1.1 Introduction	8
1.2 Electricity data	12
1.3 Functional factor model	16
1.4 Estimation procedure	18
1.4.1 Estimation of the price-demand functions X_t	19
1.4.2 Estimation of the basis system $\{f_1, \dots, f_T\}$	20
1.4.3 Random domains $\mathcal{D}(X_t) = [a_t, b_t]$	22
1.4.4 The estimator $\{\hat{f}_1, \dots, \hat{f}_K\}$	23
1.4.5 A note on convergence	24
1.5 Application	26
1.5.1 Interpretation of the factors and exemplary analysis of the scores	29
1.5.2 Validation of the model assumptions	30
1.6 Forecasting	32
1.6.1 Forecasting with the FFM	32
1.6.2 Competing forecast models	35
1.6.3 Evaluation of forecast performances	39

1.7	Conclusion	43
2	A fundamental model for electricity spot prices using functional data analysis	45
2.1	Introduction	45
2.2	Data	51
2.3	Model & estimators	52
2.3.1	Unifying regression model	56
2.3.2	The estimator	59
2.3.3	Random design of the prediction points	60
2.3.4	Boundary estimators	62
2.4	Technical assumptions & main results	64
2.4.1	Optimal bandwidth selection	69
2.4.2	Estimation of the eigencomponents	70
2.4.3	Estimation of the principal component scores	71
2.5	Application	74
2.6	Conclusion	81
2.7	Appendix: Proofs	82
2.7.1	Proof of Proposition 2.4.1	82
2.7.2	Proof of Theorem 2.4.1	82
2.7.3	Proof of Theorem 2.4.2	90
3	The R-package phtt: Panel Data Analysis with Heterogeneous Time Trends	95
3.1	Introduction	95
3.2	Panel models for heterogeneity in time trends	100
3.2.1	Computational details	103
3.2.2	Application	105
3.3	Panel criteria for selecting the number of factors	108
3.3.1	Application	111
3.4	Panel models with stationary common factors	115
3.4.1	Model with known number of factors d	115
3.4.2	Model with unknown number of factors d	116

Contents	iii
3.4.3 Application	119
3.5 Models with additive and interactive unobserved effects	122
3.5.1 Specification tests	125
3.5.1.1 Testing the sufficiency of classical additive effects	126
3.5.1.2 Testing the existence of common factors	128
3.6 Interpretation	130
3.7 Summary	133
Conclusions	135
Danksagung	139
References	141

List of Figures

1.1	Time series of electricity market data	10
1.2	Univariate time series of fitted price-demand functions	13
1.3	Empirical covariance function and VARIMAX rotated basis functions	28
1.4	Time series of principal component scores	28
1.5	P-values of Granger-causality tests	30
1.6	Hourly electricity spot prices	33
1.7	Forecast of the price-demand function of January 25, 2008	34
1.8	Comparison of the spot prices and the 1 day ahead forecasts	36
1.9	Root mean squared errors	41
1.10	Mean and trimmed mean values of the interval scores $S_{\alpha}^{int}(h, \ell)$	42
2.1	Discretization points of the random price functions	47
2.2	Scatter plot of prediction points	49
2.3	Contour plots of the estimated mean functions	75
2.4	First and second eigenfunctions	77
2.5	Temperature dependent variance shares	78
2.6	Fitted price functions	80
3.1	Plots of the dependent variable and the regressor variables	100
3.2	Estimated factors and time-varying individual effects	108
3.3	Scree plot produced by the <code>plot()</code> -method	114
3.4	Estimated factors and time-varying individual effects	122
3.5	Estimated common factors, time-, and individual-effects	126
3.6	Visualization of the differences in the individual effects	132

Introduction

This thesis contributes to a specific branch of nonparametric statistics called “*functional data analysis*”. While classical statistics deals with the analysis of random scalars, vectors, and matrices, functional data analysis (FDA) refers to the statistical analysis of random functions. Examples of random functions include biological and biomechanical data, e.g., growth curves, which can be sampled (nearly) without noise at (in principle) arbitrary discretization points. This type of functional data was the motivating starting point for the development of suitable statistical methods [see Rao (1958) for an early reference].

The next generation of methods for functional data focuses on latent random functions from which we observe only finitely many, noisy discretization points—a situation typically encountered in economic contexts. Here, the functional nature of the data represents a qualitative assumption on the underlying data-generating process. Examples in the literature of this kind of functional data include the analysis of ebay auction prices, production indices, and implied volatility functions [see, e.g., Wang *et al.* (2008b), Ramsay & Ramsey (2002), and Benko *et al.* (2009)].

The potentially infinite dimension of functional data demands for methods of dimensionality reduction in advance of almost any further statistical analysis. In this regard, the functional version of principal components analysis (FPCA) became the de facto standard for decomposing random functions into useful basis components. The resulting decomposition of random functions into linear combinations of basis functions (called eigenfunctions) and univariate random variables (called scores), leads to the well-known Karhunen-Loève decomposition [see, e.g., Ash & Gardner (1975)]. The underlying idea is that of a function space with pairwise orthonormal basis functions. FPCA chooses basis functions, in an op-

timal way such that a low number of basis functions suffices to approximate the original random functions with high accuracy.

Nowadays many classical statistical models are generalized for the case of random functions and very often FPCA builds the core method that allows for this kind of generalization. Prominent examples are functional linear regression models, functional auto regressive models, and functional canonical correlation analysis [see Ramsay & Dalzell (1991), Bosq (2000) and He *et al.* (2003)]. However, there are also statistical topics and problems exclusively encountered in FDA, such as the statistical analysis of differential equations and the problem of registration in the case of misaligned functions [see, e.g., Ramsay (1996) and Kneip & Ramsay (2008)]. An excellent overview of FDA methods can be found in the monographs of Ramsay & Silverman (2005) and Ferraty & Vieu (2006).

FPCA is well studied for the classical case, where it is assumed that a sample of functions is observed precisely; however, this is a situation rarely fulfilled in practice. Recent work focuses on the more realistic case of latent functional data with (finitely many) noisy discretization points per function.

Generally, there are two different strategies for conducting FPCA in these more challenging cases: First, one can either pre-smooth each function and then estimate the covariance function. Second, one can estimate the covariance function directly from the noisy discretization points using nonparametric smoothing procedures. The former possibility is applied in Chapters 1 and 2, while the latter is used in Chapter 2.

The latter possibility has the further advantage that it also works for sparse functional data with only very few observations per function. But both strategies involve nonparametric smoothing procedures, and those procedures crucially depend on an appropriate choice of smoothing parameters. This problem is of particular interest and is carefully treated in all of the following chapters. Contrary to the usual case, optimization has to be done with respect to the common basis (or eigen-)functions and not with respect to the single random functions.

In this thesis we focus on the following three classical statistical models and their application in the context of functional data analysis: Chapter 1 deals with the functional factor model, in Chapter 2 we introduce multivariate nonparametric regression model as a tool for FPCA, and in Chapter 3 we discuss panel data

models that allow for functional factor structures in the error term.

In Chapter 1, we propose a new perspective on modeling and forecasting electricity spot prices. Our approach is motivated by the data-generating process of electricity spot prices, which is well described what is called the merit order model. The merit order model is a micro economic model based on the assumption that spot prices on electricity exchanges are determined by the marginal generation costs of the last power plant that is required to cover the demand. The resulting merit order curve reflects the increasing generation costs of the installed power plants. Correspondingly, we suggest interpreting hourly electricity spot prices as noisy discretization points of smooth price functions.

These price functions are modeled by a functional factor model (FFM) for which we discuss a two-step estimation procedure. The first step is a classical pre-smoothing step in order to estimate the single price functions from the noisy discretization points. The second step then aims for a robust estimation of a finite set of common basis functions from the pre-smoothed price functions. In doing this, we carefully consider the issue of finding an optimal smoothing parameter.

The presentation of our functional factor model concludes with an extensive forecast study which compares our FFM with alternative time series models that have been successfully applied in the literature on electricity spot prices. The forecast study clearly confirms the superior power of our functional factor model and the use of price functions as underlying structures of electricity spot prices in general.

A slightly modified version of Chapter 1 is forthcoming as a single-authored article in *The Annals of Applied Statistics*; see Liebl (2013).

Chapter 2 further discusses the problem of modeling electricity spot prices. On the one hand, we extend the concept of price function introduced in Chapter 1 by two additional covariables. On the other hand, we focus on a generally deeper theoretical consideration of the involved multivariate nonparametric regression model, which is used as a tool for FPCA.

We extend existing theoretical results with respect to FPCA for sparse functional data by considering the asymptotic bias and variance of the multivariate local linear estimator of the mean and the covariance functions. Here, we carefully consider the effects of between-correlations, which are caused by the time series

context, and the effects of within-correlations, which are caused by the functional nature of the data.

In order to demonstrate the usefulness of our model we analyze the effects of Germany's nuclear moratorium on March 14, 2011. This event describes a natural experiment, since in the course of Germany's nuclear moratorium on March 14, 2011, eight nuclear power plants were phased out [Nestle (2012)]. The data set analyzed in Chapter 2 covers exactly one year before and one year after Germany's nuclear power phase-out. We apply our model separately to these two time spans in order to contrast the different market situations.

Chapter 2 is based on a joint research project together with Prof. Dr. Alois Kneip (University Bonn). All paperwork as well as theoretical work is done by the author of this thesis, Prof. Dr. Alois Kneip checks plausibility of the theoretical work.

In Chapter 3 we pick up the successful application of FDA within the literature on panel data models. Recent panel data models allow us to control for complex unobserved heterogeneity effects by the incorporation of latent factor models. This new kind of panel data models extends the classical concept of individual random (scalar) effects to random processes or random functions [see, e.g., Bai *et al.* (2009), Bai (2009), and Kneip *et al.* (2012a)].

Even though this class of panel models is of high relevance for practical problems such as stochastic frontier analysis, they are still rarely applied in the empirical literature. Our implementation of these methods in the statistical software package of Bada & Liebl (2013b) provides a first step towards facilitating their application. Beyond this, we solve some open estimation problems, extend the suggested models for classical fixed effects, and introduce a new Hausman-type specification test.

As the estimation procedure of Kneip *et al.* (2012a) involves nonparametric smoothing methods, the choice of a reliable procedure to find an optimal smoothing parameter is most important for implementing the estimation procedure in a statistical software package. We consider this problem and suggest using Generalized Cross Validation (GCV) in order to determine an upper bound for the optimal smoothing parameter. However, it is impossible to apply the classical GCV formulas as proposed, e.g., in Craven & Wahba (1978) since we do not

know the parameters β and $v_i(t)$. Our computational algorithm for determining the GCV smoothing parameter is based on the parameter cascading method suggested in Cao & Ramsay (2010).

This GCV smoothing parameter builds an upper bound for the optimal smoothing parameter, since it does not account for the qualitative assumption of an underlying factor structure. The final optimal smoothing parameter lies somewhere between the GCV smoothing parameter and zero. Knowledge of this interval allows for a reasonable implementation of the computationally costly cross validation criterion.

A slightly modified version of Chapter 3 is accepted as a co-authored article for the *Journal of Statistical Software*; see Bada & Liebl (2013a). The main sections of this article are Section 3.2, which describe the implementation of the estimation procedures proposed by Kneip *et al.* (2012a), and Section 3.4, which describe the implementation of the estimation procedures proposed in Bai (2009). The author of this thesis is responsible for Section 3.2, while the co-author of the article, Oualid Bada (Univeristy Bonn), is responsible for Section 3.4. Sections 3.3 and 3.5 discuss further estimation problems, model extensions, as well as specification tests. Generally speaking, those parts, which refer to the panel model of Kneip *et al.* (2012a) are written by the author of this thesis, those parts, which refer to the panel model of Bai (2009), are written by the co-author. The remaining parts of the latter sections, which refer to both models, as well as the introductory and the concluding Sections 3.1 and 3.6 are joint works of both authors.

The outline of this thesis is as follows: In Chapter 1 we present our perspective on modeling and forecasting electricity spot prices. In Sections 2.3 and 1.2 we motivate the conceptual idea and discuss peculiarities of electricity market data. In Sections 1.3 and 1.4 we introduce the functional factor model and the two step estimation procedure. Chapter 1 is completed by an application of the model in Section 1.5 and a forecast study in Section 1.6. The results are discussed in Section 1.7.

In Chapter 2 we introduce multivariate nonparametric regression as a tool for FPCA. In Sections 2.1 and 2.2 we discuss the general ideas and the data. The nonparametric regression models as well as the multivariate local linear estimator are discussed in Section 2.3. In Section 2.4 we present our main theoretical results.

In Section 2.5 we apply the procedures to real data. Chapter 2 completed by the conclusion in Section 2.6.

In Chapter 3 we introduce our statistical software package Bada & Liebl (2013b). Section 3.1 discusses the implemented panel data models at a general level. In Sections 3.2 and 3.4 the models are discussed in more detail. In Section 3.3 we present several criteria to determine the factor dimension. In Section 3.5 we augment the panel data models for classical fixed effects and propose a Hausman-type specification test. In Section 3.6 we outline an example of the application of the procedures. Chapter 3 is completed by a short summary in Section 3.7.

Chapter 1

Modeling and Forecasting Electricity Spot Prices: A Functional Data Perspective

Classical time series models have serious difficulties in modeling and forecasting the enormous fluctuations of electricity spot prices. Markov regime switch models belong to the most often used models in the electricity literature. These models try to capture the fluctuations of electricity spot prices by using different regimes, each with its own mean and covariance structure. Usually one regime is dedicated to moderate prices and another is dedicated to high prices. However, these models show poor performance and there is no theoretical justification for this kind of classification. The merit order model however, the most important micro-economic pricing model for electricity spot prices, suggests a continuum of mean levels with a functional dependence on electricity demand.

We propose a new statistical perspective on modeling and forecasting electricity spot prices that accounts for the merit order model. In a first step, the functional relation between electricity spot prices and electricity demand is modeled by daily price-demand functions. In a second step, we parameterize the series of daily price-demand functions using a functional factor model. The power of this new perspective is demonstrated by a forecast study that compares our functional factor model with two established classical time series models as well as

two alternative functional data models.

1.1 Introduction

Time series of hourly electricity spot prices have peculiar properties. They differ substantially from time series of equities and other commodities because electricity still cannot be stored efficiently and, therefore, electricity demand has an untempered effect on the electricity spot price [Knittel & Roberts (2005)].

The development of models for electricity spot prices was triggered by the liberalization of electricity markets in the early 1990s. Hourly electricity spot prices are usually considered to be multivariate (24-dimensional) time series since for each day t the 24 intra-day spot prices are settled simultaneously the day before [Huisman *et al.* (2007)].

However, classical time series models adopted for electricity spot prices such as autoregressive, jump diffusion, or Markov regime switch models reduce the multivariate time series to univariate time series either by taking daily averages of the 24 hourly spot prices [Weron *et al.* (2004), Kosater & Mosler (2006), and Koopman *et al.* (2007)] or by considering each hour h separately [Karakatsani & Bunn (2008)]. These unnatural aggregations and separations of the data necessarily come with great losses in information.

Our model, a functional factor model (FFM), is not a mere adaption of a classical time series model but is motivated by the data-generating process of electricity spot prices itself. Pricing in power markets is explained by the merit order model. This model assumes that the spot prices at electricity exchanges are based on the marginal generation costs of the last power plant that is required to cover the demand. The resulting so-called merit order curve reflects the increasing generation costs of the installed power plants. Often, nuclear and lignite plants cover the minimal demand for electricity. Higher demand is mostly served by hard coal and gas fired power plants.

Due to its importance the merit order model is referred to as a fundamental market model [Burger *et al.* (2008), Chapter 4]. Essentially, the consideration of this fundamental model yields to the superior forecast performance of our FFM in comparison to state of the art time series models and alternative functional

data models.

It is important to emphasize that the merit order model is not a static model. The merit order curve rather depends on the variations of the daily prices for raw materials, the prices of CO₂ certificates, the weather, plant outages, and maintenance schedules of power plants.

The merit order curve is most important for the explanation of electricity spot prices in the literature on energy economics and justifies our view on the set of hourly electricity spot prices $\{y_{t1}, \dots, y_{t24}\}$ of day t . We do not interpret them as 24-dimensional vectors but rather as noisy discretization points of a smooth price-demand function X_t , which can be formalized as follows:

$$y_{th} = X_t(u_{th}) + \varepsilon_{th},$$

where u_{th} denotes electricity demand at hour h of day t and ε_{th} is assumed to be a white noise process.

The price-demand function $X_t(u)$ can be seen as the empirical counterpart of the merit order curve estimated non-parametrically from the $N = 24$ hourly price-demand data pairs $(y_{t1}, u_{t1}), \dots, (y_{tN}, u_{tN})$. Five exemplary estimated price-demand functions $\hat{X}_t(u)$ are shown in the lower panel of Figure 1.1. Figure 1.2 visualizes the temporal evolution of the time series of price-demand functions by showing the univariate time series $\hat{X}_1(u), \dots, \hat{X}_T(u)$ for a fixed value of electricity-demand $u = 58,000$ MW for the whole observed time span of $T = 717$ work days (Mo.-Fr.) from January 1, 2006 to September 30, 2008.

In order to capture the dynamic component of the price-demand functions we assume them to be generated by a functional factor model defined as

$$X_t(u) = \sum_{k=1}^K \beta_{tk} f_k(u),$$

where the factors or basis functions f_k are time constant and the corresponding scores β_{tk} are allowed to be non-stationary time series.

We do not specify a constant mean function in our FFM, since we allow the time series of price-demand functions ($X_t(u)$) to be non-stationary. Consequently, the classical interpretation of the factors f_k as perturbations of the mean does

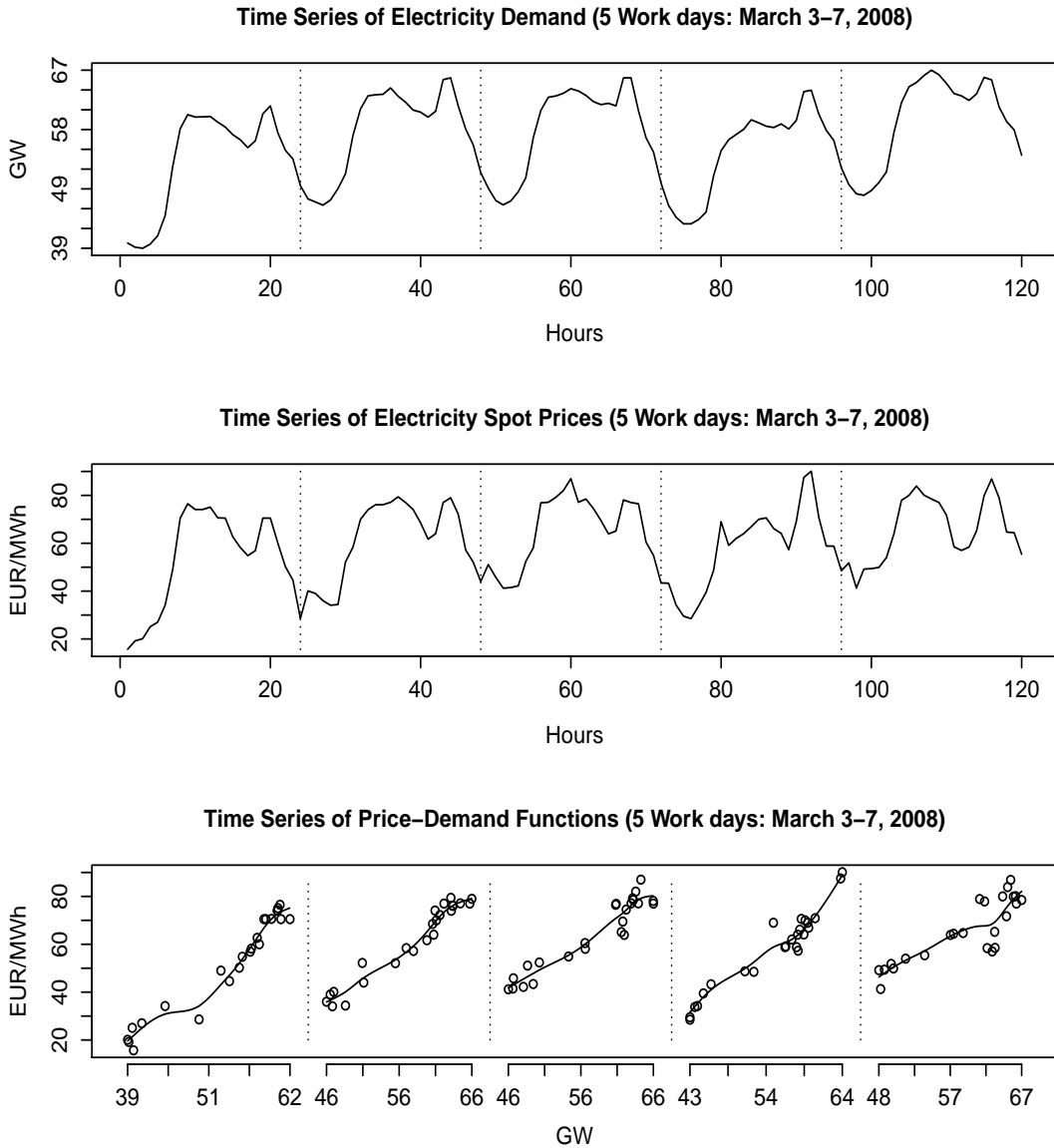


Figure 1.1: UPPER PANEL: Time series of electricity demand (u_{th}), measured in GW (1 GW = 1000 MW). MIDDLE PANEL: Electricity spot prices (y_{th}). LOWER PANEL: Price-demand functions (\hat{X}_t) with noisy discretization points $(y_{t1}, u_{t1}), \dots, (y_{tN}, u_{tN})$.

not apply—as common in the literature on dynamic (functional) factor models; see, e.g., Hays *et al.* (2012).

Note that the five price-demand functions in the lower panel of Figure 1.1 are observed on different domains. This distinguishes our functional data set from classical functional data sets, where all functions are observed on a common domain. We refer to this feature as *random domains* and its consideration in Sections 1.4.3 and 1.4.5 is a central part of our estimation procedure.

We use a two-step estimation procedure. The first step is to estimate the daily price-demand functions \hat{X}_t by cubic spline smoothing for all days $t \in \{1, \dots, T\}$. The second step is to determine a $K < \infty$ dimensional common functional basis system $\{f_1, \dots, f_K\}$ for the estimated price-demand functions $\hat{X}_1, \dots, \hat{X}_T$. Given this system of basis functions we model the estimated daily price-demand functions by a functional factor model—using the basis functions as common factors. The fitted discrete hourly electricity spot prices \hat{y}_{th} are then obtained through the evaluation of the modeled price-demand functions at the corresponding hourly values of demand for electricity; formally written as: $\hat{y}_{th} = \hat{X}_t(u_{th})$.

Functional data analysis (FDA) can share our perspective on electricity spot prices. A broad overview of many different FDA methods can be found in the monographs of Ramsay & Silverman (2005) and Ferraty & Vieu (2006). Particularly, Chapter 8 in Ramsay & Silverman (2005) and the non-parametric methods for computing the empirical covariance function as proposed in Staniswalis & Lee (1998), Yao *et al.* (2005), Hall *et al.* (2006), and Li & Hsing (2010) are important methodological references for this paper.

The application of models from the functional data literature to electricity market data is not new. For example, there is a vast literature on modeling and forecasting electricity *demand*; see, e.g., Ferraty & Vieu (2006) and Antoch *et al.* (2010). However, modeling and forecasting electricity spot prices is much more difficult than modeling and forecasting electricity demand. The semi-functional partial linear model (SFPL) of Vilar *et al.* (2012) is one of the very rare cases in which FDA methods are used to forecast electricity spot prices.

Two very recent examples of other functional factor models are given by the functional factor analysis in Liu *et al.* (2012) and the functional dynamic factor model (FDFM) in Hays *et al.* (2012). Liu *et al.* (2012) propose a new rotation

scheme for the functional basis components. Hays *et al.* (2012) model a time series of yield curves and estimate their model by the EM algorithm. In contrast to the FDFM of Hays *et al.* (2012) we do not have to make a priori assumptions on the stochastic properties of the time series of scores in order to estimate our model components. Furthermore, we are able to model and forecast functional time series observed on random domains.

Very close to the FDFM of Hays *et al.* (2012) is the Dynamic Semiparametric Factor Model (DSFM) of Park *et al.* (2009). As our functional factor model the DSFM does not need a priori assumptions on the time series of scores. This and the fact that the DSFM was already successfully applied to electricity prices [Borak & Weron (2008) and Härdle & Trück (2010)] makes the DSFM a perfect competitor for our FFM.

The main difference between the FDFM of Hays *et al.* (2012) and the DSFM of Park *et al.* (2009) in comparison to our FFM is that the FFM can deal with functional times series observed on random domains. Furthermore, Park *et al.* (2009) use an iterating optimization algorithm to estimate the basis functions of the DSFM, whereas we standardize the elements of the time series (X_t) so that we can robustly estimate the basis functions by functional principal component analysis. Our estimation procedure is much simpler to implement and faster with respect to computational time than the Newton-Raphson algorithm suggested in Park *et al.* (2009).

The next section is devoted to the introduction of our data set and to a critical consideration of the stylized facts of electricity spot prices usually claimed in the electricity literature. In Section 1.3 we present our functional factor model and in Section 1.4 its estimation. An application of the model to real data is presented in Section 1.5. Finally, the performance of the functional factor model is demonstrated by an extensive forecast study in Section 1.6.

1.2 Electricity data

We demonstrate our functional factor model by modeling and forecasting electricity spot prices of the German power market traded at the European Energy Exchange (EEX) in Leipzig. The German power market is the biggest power

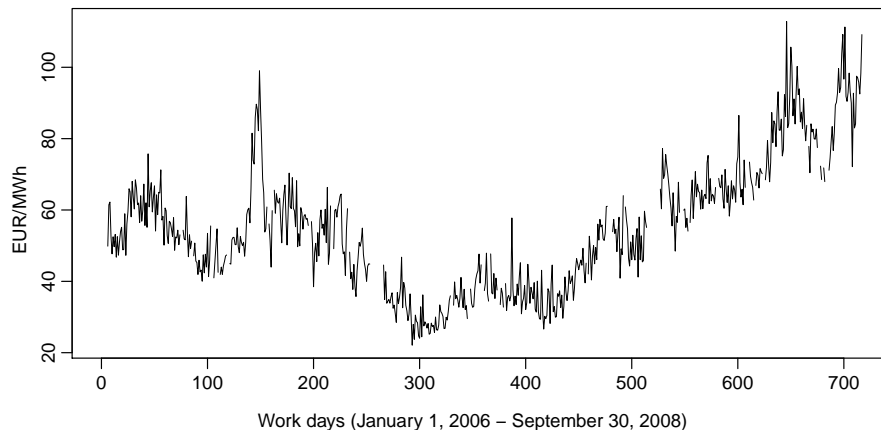


Figure 1.2: Univariate time series of fitted price-demand functions $\hat{X}_1(u), \dots, \hat{X}_T(u)$ evaluated at $u = 58000$ MW. Gaps correspond to holidays.

market in Europe in terms of consumption. The wholesale market is fragmented into an Over The Counter (OTC) market and the EEX. While the OTC market has a continuous trade, the EEX has a single uniform price auction with a gate closure for the day ahead market at 12 p.m. the day before physical delivery. Although three-fourths of the trading volume is settled via bilateral OTC contracts, the EEX spot price is of fundamental importance as benchmark and reference point for other markets, such as OTC or forward markets [Ockenfels *et al.* (2008), Chapter 1].

The data for this analysis stem from three different publicly available sources. The hourly spot prices of the German electricity market are provided by the European Energy Exchange (www.eex.com), hourly values of Germany's gross electricity demand are provided by the European Network of Transmission System Operators for Electricity (www.entsoe.eu), and German wind power infeed data are provided by the EEX Transparency Platform (www.transparency.eex.com).

In the German electricity market, as in most of the electricity markets in the world, renewable energy sources are usually provided with purchase guarantees. Therefore, not the hourly values of gross electricity demand are relevant for the pricing at the EEX but rather the hourly values of gross demand minus the hourly electricity infeeds from renewable energy sources. We consider only wind

power infeed data since the influences of other renewable energy sources such as photovoltaic and biomass on electricity spot prices are still negligible for the German electricity market (and their explicit consideration essentially would lead to the same results).

The data consists of pairs (y_{th}, u_{th}) with y_{th} denoting the electricity spot price and u_{th} the electricity demand of hour $h \in \{1, \dots, 24\}$ at day t . We define electricity demand u_{th} as the gross electricity demand of hour h and day t minus the wind power infeed of electricity at the corresponding hour h and day t .

The data set analyzed in this article covers $T = 717$ work days (Mo.-Fr.) within the time horizon from January 1, 2006 to September 30, 2008. For the sake of clarity, only working days are considered in our analysis since for weekends there are different compositions of the power plant portfolio. The same reasoning applies to holidays and so-called Brückentage, which are extra days off that bridge single working days between a bank holiday and the weekend. Therefore we set all holidays and Brückentage to NA-values.

As a referee noted, the time span of our data set is peculiar. Starting around January, 2007 a price bubble for raw commodities such as coal and gas was formed, which induced a strong increase in the electricity spot prices. Interestingly, the increase in the electricity spot prices is hardly visible in the original time series as shown in Figure 1.6. But it catches the eye in the plot of Figure 1.2, which shows the time series of price-demand functions $(\hat{X}_t(u)|u)$ evaluated for a certain value of electricity demand $u = 58,000$ MW. The reason is that at this relatively high value of electricity demand usually coal and gas fired power plants cover the demand.

Very few (only 0.5%) of the data pairs (y_{th}, u_{th}) with prices $y_{th} > 200$ EUR/MWh have to be treated as outliers since they cannot be explained by the merit order model. Even in exceptional situations the marginal costs of electricity production do not exceed the value of 200 EUR/MWh. Prices above this threshold are referred to as price spikes and have to be explained using an additional scarcity premium [Burger *et al.* (2008), Chapter 4]. The analysis of price spikes is a research topic on its own [Christensen *et al.* (2009)] and is not within the scope of this paper.

We exclude the outliers for the estimation of our model and denote the amount

of data pairs of day t used for estimation by $N_t \leq N = 24$. Nevertheless, we use the whole data set, including the outliers, in order to assess the forecast performance of our model in Section 1.6.

Review: Stylized facts of electricity data Our functional perspective on electricity spot prices allows us to review critically the so-called “stylized facts” of hourly electricity spot prices (y_{th}). Usually, time series of electricity spot prices are assumed (i) to have deterministic daily, weekly and yearly seasonal patterns, (ii) to show price dependent volatilities, and (iii) to be stationary (after controlling for the seasonal patterns); see Huisman & De Jong (2003), Knittel & Roberts (2005), Kosater & Mosler (2006), Huisman *et al.* (2007), and many others.

At first glance these stylized facts seem to be reasonable; see the middle panel in Figure 1.1. However, the first two stylized facts, (i) and (ii), are misleading since both have their origins in the time series of electricity demand: the characteristics of electricity demand are rather carried over to the time series of electricity spot prices.

This can be explained by a micro-economic point of view, again using the merit order model. The merit order curve induces a monotone increasing supply function for electricity, which implies higher electricity spot prices for higher values of electricity demand, where electricity demand can be considered as inelastic. Given this micro-economic point of view, we can regard the daily supply functions for electricity as diffusers in the transmission from electricity demand u_{th} to the electricity spot price y_{th} .

Additional diffusion comes from the variations of the daily supply functions caused by the varying input-costs of, e.g., coal and gas. Compare to this the time series of electricity demand with the time series of electricity spot prices shown in the upper and middle panels of Figure 1.1 respectively. The seasonal patterns of electricity spot prices are just a diffused version of the smoother seasonal patterns of electricity demand.

Price dependent volatility (ii) can be explained by the slope of the merit order curve, which is increasing with electricity demand. Changes in electricity demand have greater price effects for greater values of electricity demand and therefore

cause greater volatilities than is the case for lower values of electricity demand.

Stationarity (iii) has to be considered critically, too. Recently, Bosco *et al.* (2010) were able to show empirically that electricity spot prices at the EEX have a unit root. The authors point out that the stationarity assumption might be wrong in markets that are influenced by price-enhancing sources such as prices for coal and gas since time series of coal and gas prices are commonly found to be non-stationary. Our functional factor model allows for non-stationarity in the time series of price-demand functions (X_t) and, in fact, tests indicate that the estimated series of price-demand functions is non-stationary; see Section 1.5.2.

This short review of electricity spot prices demonstrates that electricity data are complex with dynamics induced by the variations of the merit order curve (mainly caused by varying input-costs) and separate additional dynamics induced by electricity demand. To the best of our knowledge, our functional factor model is the first model that allows for a separate consideration of these two stochastic sources. The variations dedicated to the dynamics of the merit order curve are captured by the price-demand functions and modeled by our functional factor model. The problem of modeling and forecasting electricity demand is “outsourced” and the statistician can choose powerful specialized models for time series of gross electricity demand [Antoch *et al.* (2010)] and time series of wind power [Lau & McSharry (2010)]. This separation corresponds to the real data generating process.

1.3 Functional factor model

As mentioned above, electricity spot prices y_{t1}, \dots, y_{t24} are actually one-day-ahead future prices since they are settled simultaneously at day $t - 1$. This implies that there is some degree of uncertainty about the next day world in the electricity spot price y_{th} , which we model non-parametrically as

$$y_{th} = X_t(u_{th}) + \varepsilon_{th}. \quad (1.1)$$

The error terms ε_{th} are assumed to be iid white noise errors with finite variance $V(\varepsilon_{th}) = \sigma_\varepsilon^2$ and each function X_t is assumed to be continuous and square

integrable.

For each function X_t the values of electricity demand u_{th} are only observed within random sub-domains $\mathcal{D}(X_t) = [a_t, b_t]$, where $[a_t, b_t] \subseteq [A, B] \subset \mathbb{R}$. The unobserved univariate time series (a_t) and (b_t) are assumed to be time series processes with $A \leq a_t < b_t \leq B$ and marginal pdf's of a_t and b_t given by $f_a(z_a) > 0$ and $f_b(z_b) > 0$ for all $z_a, z_b \in [A, B]$ and $t \in \{1, \dots, T\}$.

The price-demand functions are relatively homogeneous. All of them look very similar to the five randomly chosen price-demand functions shown in the lower panel of Figure 1.1. The underlying reason for this homogeneity is that, on the one hand, the merit order curve induces rather simple monotone increasing price-demand functions. On the other hand, the general portfolio of power plants, which is reflected by the merit order curve, is changing very slowly and can be considered as constant over the period of our analysis. We formalize this homogeneity of the price-demand functions by the assumption that the time series of price-demand functions (X_t) is generated by a functional factor model with time constant basis functions.

Given this assumption, every price-demand function X_t can be modeled by the same set of $K < \infty$ (unobserved) basis functions $f_1, \dots, f_k, \dots, f_K$ with $f_k \in L^2[A, B]$, which span the K -dimensional functional space $\mathcal{H}_K \subset L^2[A, B]$ such that we can write

$$X_t(u) = \sum_{k=1}^K \beta_{tk} f_k(u) \quad \text{for all } u \in [a_t, b_t], \quad (1.2)$$

where the common basis functions f_k as well as the scores β_{tk} are unobserved and have to be determined from the data. We use the usual orthonormal identification restrictions for the basis functions, which require that $\int_A^B f_k^2(u) du = 1$ and $\int_A^B f_k(u) f_l(u) du = 0$ for all $k < l \in \{1, \dots, K\}$.

The K real time series $(\beta_{t1}), \dots, (\beta_{tK})$ are defined as

$$\begin{pmatrix} \beta_{t1} \\ \vdots \\ \beta_{tK} \end{pmatrix} = \begin{pmatrix} \int_{a_t}^{b_t} f_1^2 & \cdots & \int_{a_t}^{b_t} f_1 f_K \\ \vdots & \ddots & \vdots \\ \int_{a_t}^{b_t} f_1 f_2 & \cdots & \int_{a_t}^{b_t} f_K^2 \end{pmatrix}^{-1} \begin{pmatrix} \int_{a_t}^{b_t} f_1 X_t \\ \vdots \\ \int_{a_t}^{b_t} f_K X_t \end{pmatrix} \quad (1.3)$$

and are allowed to be arbitrary non-stationary processes. Note that for $a_t = A$ and $b_t = B$ the definition of the scores β_{tk} corresponds to the classical definition, given by $\beta_{tk} = \int_A^B X_t(u) f_k(u) du$.

In the following section we propose an estimation algorithm for the functional factor model.

1.4 Estimation procedure

As outlined in Sections 2.3 and 1.2 we do not observe the series (X_t) directly but have to estimate each price-demand function X_t from the corresponding data pairs $(y_{t1}, u_{t1}), \dots, (y_{tN_t}, u_{tN_t})$. After this initial estimation step, which is discussed in Section 1.4.1, we show in Section 1.4.2 how to determine an orthonormal K -dimensional basis system $\{f_1, \dots, f_K\}$ for the classical functional data case when all price-demand functions X_1, \dots, X_T are observed on the deterministic domain $\mathcal{D}(X_t) = [A, B]$. In Section 1.4.3 we generalize the determination of the orthonormal K -dimensional basis system $\{f_1, \dots, f_K\}$ to our case, where the price-demand functions X_t are observed only on random domains $\mathcal{D}(X_t) = [a_t, b_t]$. Finally, we define our estimator $\{\hat{f}_1, \dots, \hat{f}_K\}$ in Section 1.4.4.

As usually for (functional) factor models, the set of factors $\{f_1, \dots, f_K\}$ in Eq. (1.2) is only determined up to orthonormal rotations. Furthermore, the determination of an orthonormal K -dimensional basis system $\{\hat{f}_1, \dots, \hat{f}_K\}$ for a given series (\hat{X}_t) is, in the first instance, a mere algebraic problem. But it is also a statistical estimation problem in the sense that $\hat{\mathcal{H}}_K$, with $\hat{\mathcal{H}}_K = \text{span}(\hat{f}_1, \dots, \hat{f}_K)$, is a consistent estimator of the theoretical counterpart \mathcal{H}_K . The crucial assumption is that X_t comes from the FFM (1.2). Consistency of the estimation follows from the consistency of the single non-parametric estimators $\hat{X}_t(u)$, which converge in probability against $X_t(u)$ as $N_t \rightarrow \infty$ for all $u \in [a_t, b_t]$ and all $t \in \{1, \dots, T\}$ [Benedetti (1977)]. Below in Section 1.4.5 we consider this issue in more detail.

1.4.1 Estimation of the price-demand functions X_t

The estimation of the functions X_t from the data pairs $(y_{t1}, u_{t1}), \dots, (y_{tN_t}, u_{tN_t})$ is done by minimizing

$$d(\mathcal{X}|t) = \sum_{h=1}^{N_t} (y_{th} - \mathcal{X}(u_{th}))^2 + b \int_{a_t}^{b_t} (D^2 \mathcal{X}(u))^2 du \quad (1.4)$$

over all twice continuously differentiable functions \mathcal{X} , where $D^2 \mathcal{X}$ denotes the 2nd derivative of \mathcal{X} and $b > 0$ is a preselected smoothing parameter. Spline theory assures that any solution \hat{X}_t of the minimization problem (1.4) can be expanded by a natural spline basis [De Boor (2001a)]. Therefore we can use the expansion $\mathcal{X}(u) = \mathbf{c}' \boldsymbol{\phi}(t)$, where $\boldsymbol{\phi}$ is the $(N_t + 2)$ -vector of natural spline basis functions of degree 3 and \mathbf{c} is the $(N_t + 2)$ -vector of coefficients over which Eq. (1.4) is minimized. This procedure is usually denoted as cubic spline smoothing and the interested reader is referred to the monographs of De Boor (2001a) and Ramsay & Silverman (2005).

An important issue that remains to be discussed is the selection of the smoothing parameter b . Usually, the optimal smoothing parameter b^{opt} is chosen by (generalized) cross-validation such that the trade-off between bias and variance of the estimate \hat{X}_t is optimized asymptotically with respect to the mean integrated squared error (MISE) criterion. However, our aim is not an optimal single estimate \hat{X}_t but rather an optimal estimation of the basis system $\{f_1, \dots, f_K\}$ for which we can use the information of all price-demand functions X_1, \dots, X_T .

Consequently, we do not have to optimize the MISEs of the single estimators \hat{X}_t but those of their weighted averages $\hat{f}_1, \dots, \hat{f}_K$. In this case an undersmoothing parameter $\underline{b}_K < b^{opt}$ has to be chosen. This was discussed for the first time in Benko *et al.* (2009). The underlying reason is that the estimators $\hat{f}_1, \dots, \hat{f}_K$ essentially are weighted averages over all $\hat{X}_1, \dots, \hat{X}_T$. Averaging reduces the overall variance and therefore opens the possibility for a further reduction in the MISEs of the estimators $\hat{f}_1, \dots, \hat{f}_K$ by a further reduction of the bias-component through choosing $\underline{b}_K < b^{opt}$ in the minimization of Eq. (1.4). Benko *et al.* (2009) propose to approximate an optimal undersmoothing parameter \underline{b}_K by minimizing the

following cross-validation criterion:

$$CV(b_K) = \sum_{t=1}^T \sum_{i=1}^{N_t} \left\{ y_{th} - \sum_{k=1}^K \hat{\gamma}_{tk} \hat{f}_{k,-t}(u_{th}) \right\}, \quad (1.5)$$

over $0 \leq b_K \leq \infty$, where $\hat{\gamma}_{tk}$ are the OLS estimators of $\hat{\beta}_{tk}$ and $\hat{f}_{k,-t}$ denote the estimators of f_{tk} based on the data pairs (y_{sh}, u_{sh}) with $s \in \{1, \dots, t-1, t+1, \dots, T\}$. We denote the estimators of X_t based on an undersmoothing parameter \underline{b}_K by $\tilde{X}_1, \dots, \tilde{X}_T$ and those based on b^{opt} by $\hat{X}_1, \dots, \hat{X}_T$.

As can be seen in Eq. (1.5) an optimal undersmoothing parameter \underline{b}_K depends on the dimension K . The problem of choosing K can be seen as a model selection problem, which generally can be solved using information criteria. For our application in Section 1.5 we use the simple cumulative variance criterion as well as the AIC type criterion proposed in Yao *et al.* (2005).

1.4.2 Estimation of the basis system $\{f_1, \dots, f_T\}$

Our estimation procedure uses the property that any orthonormal basis system $\{f_1, \dots, f_K\}$ of the series (X_t) has to fulfill the minimization problem

$$\sum_{t=1}^T \|X_t - \sum_{k=1}^K \beta_{tk} f_k\|_2^2 = \min_{B_K} \sum_{t=1}^T \min_{\gamma_{t1}, \dots, \gamma_{tK} \in \mathbb{R}} \|X_t - \sum_{k=1}^K \gamma_{tk} g_k\|_2^2 \quad (1.6)$$

over all possible K -dimensional orthonormal basis systems $B_K = \{g_1, \dots, g_K\}$, where $g_1, \dots, g_K \in L^2[A, B]$ and $\|\cdot\|_2$ denotes the functional L2 norm $\|x\|_2 = \sqrt{\int_A^B x^2(u) du}$ for any $x \in L^2[A, B]$. This property is a direct consequence of the FFM (1.2).

The minimization problem (1.6) can be used to define an estimator for a basis system $\{f_1, \dots, f_K\}$. We would only have to replace the unobserved functions X_t with their undersmoothed estimators \tilde{X}_t and try to find a basis system that minimizes the right hand side (rhs) of Eq. (1.6) using functional principal component analysis (FPCA).

Before we present the analytic solution we adjust the minimization problem (1.6). This adjustment yields a robustification, which is needed since we allow the K time series $(\beta_{t1}), \dots, (\beta_{tK})$ to be non-stationary processes. The non-

stationarity of the scores $(\beta_{t1}), \dots, (\beta_{tK})$ implies that different functions X_t and X_s can be of very different orders of magnitude, i.e., $\|X_t\|_2 \ll \|X_s\|_2$. In such cases, the squared L2-norm on the rhs of Eq. (1.6) sets an overproportional weight on functions with great orders of magnitude, and a functional principal component estimator based on Eq. (1.6) would be distorted toward those functions X_s that have great orders of magnitude $\|X_s\|_2$.

If we were only interested in the determination of some set of basis functions $\{f_1, \dots, f_K\}$ that spans the same space \mathcal{H}_K as the set of functions $\{X_1, \dots, X_T\}$, we would not have to care about functions X_s with great orders of magnitude $\|X_s\|$. However, if we are interested in the interpretation of the basis functions f_k we want them to be representative for all functions X_1, \dots, X_T .

A general solution to this problem is to replace the price-demand functions X_t in (1.6) with their standardized counterparts $X_t^* = X_t/\|X_t\|$, which have equal orders of magnitude $\|X_t^*\| = 1$ for all $t \in \{1, \dots, T\}$. Using this replacement yields the following new minimization problem:

$$\sum_{t=1}^T \|X_t^* - \sum_{k=1}^K \beta_{tk}^* f_k^*\|_2^2 = \min_{B_K} \sum_{t=1}^T \min_{\gamma_{t1}, \dots, \gamma_{tK} \in \mathbb{R}} \|X_t^* - \sum_{k=1}^K \gamma_{tk} g_k\|_2^2 \quad (1.7)$$

Solving Eq. (1.7) by FPCA generally will yield different basis functions f_k^* than solving Eq. (1.6). However, both minimization problems (1.6) and (1.7) are equivalent in the sense that both sets of basis functions $\{f_1^*, \dots, f_K^*\}$ and $\{f_1, \dots, f_K\}$ are equivalent up to orthonormal rotations and therefore span the same space \mathcal{H}_K . The standardization yields to a simple base change, which can be seen by the fact that the original price-demand functions X_t can be written in terms of the basis functions f_k^* as $X_t = \sum_{k=1}^K (\|X_t\| \cdot \beta_{tk}^*) f_k^*$.

The standardization of all price-demand functions X_t in the minimization problem (1.7) allows us to establish a non-distorted estimator $\{\hat{f}_1, \dots, \hat{f}_K\}$ that represents all price-demand functions equally well. This approach is similar to robust estimation procedures proposed by Locantore *et al.* (1999) and Gervini (2008) but differs conceptually insofar as we do not consider any functional observation X_t as an outlier.

We construct our estimator $\{\hat{f}_1, \dots, \hat{f}_K\}$ from the analytic solution of the minimization problem (1.7). The solutions of the inner minimization problem

with respect to the scores γ_{tk} are given by least squares theory, and we can write

$$\sum_{t=1}^T \|X_t^* - \sum_{k=1}^K \beta_{tk}^* f_k^*\|_2^2 = \min_{B_K} \sum_{t=1}^T \|X_t^* - P_K^g X_t^*\|_2^2, \quad (1.8)$$

where P_K^g , defined as $P_K^g X_t^* = \sum_{k=1}^K \left(\int_A^B X_t^*(v) g_k(v) dv \right) g_k$, is a linear projection operator that projects the standardized price-demand functions X_t^* into the subspace of $L^2[A, B]$ spanned by the orthonormal basis system $B_K = \{g_1, \dots, g_K\}$.

It is well known that a solution of the minimization problem (1.8) with respect to all K -dimensional orthonormal basis systems B_K can be determined by FPCA. This so-called ‘‘best basis property’’ of the empirical eigenfunctions $e_{T,1}, \dots, e_{T,K}$ is of central importance for this paper; see Section 8.2.3 in Ramsay & Silverman (2005), among others. Note that the eigenvalues $\lambda_{T,k}$ with $k \in \{1, \dots, K\}$ may be of multiplicity $L > 1$; in this case $e_{T,k} \in E_k$ with $E_k = \text{span}(e_{T,k,1}, \dots, e_{T,k,L})$.

A solution of (1.8) is given by the set of eigenfunctions $\{e_{T,1}, \dots, e_{T,K}\}$ that belong to the first K greatest eigenvalues $\lambda_{T,1} > \lambda_{T,2} > \dots > \lambda_{T,K} > 0$ of the empirical covariance operator Γ_T defined as

$$(\Gamma_T x)(u) = \int_A^B \gamma_T(u, v) x(v) dv \quad \text{for all } x \in L^2[A, B], \quad (1.9)$$

where the empirical covariance function $\gamma_T(u, v)$ is defined as local linear surface smoother in Eq. (1.10). We use this non-parametric version of $\gamma_T(u, v)$, since it can be applied to the classical case of deterministic domains $\mathcal{D}_t(X_t) = [A, B]$ as well as to the case of random domains $\mathcal{D}_t(X_t) = [a_t, b_t]$ discussed in the following Section 1.4.3. Contrary to this, the classical textbook definition of $\gamma_T(u, v)$ cannot be applied to the case of random domains¹.

1.4.3 Random domains $\mathcal{D}(X_t) = [a_t, b_t]$

From a computational perspective, functional data observed on random domains cause problems similar to sparsely observed functional data. For the latter case there is already a broad stream of literature based on the papers of Staniswalis & Lee (1998), Yao *et al.* (2005), Hall *et al.* (2006), and Li & Hsing (2010).

¹The classical definition is given by $\gamma_T(u, v) = T^{-1} \sum_{t=1}^T X_t(u) X_t(v)$.

We follow Yao *et al.* (2005) and compute the covariance function γ_T by local linear surface smoothing. Here, $\gamma_T(u, v) = \beta_{T,0}$ and $\beta_{T,0}$ is determined by minimizing

$$\sum_{t=1}^T \sum_{i,j=1}^{N_t} \kappa_2 \left(\frac{(u_{ti} - u)}{b_\gamma}, \frac{(u_{tj} - v)}{b_\gamma} \right) \{X_t^*(u_{ti})X_t^*(u_{tj}) - f(\beta_T, (u, v), (u_{ti}, u_{tj}))\}^2 \quad (1.10)$$

over $\beta_T = (\beta_{T,0}, \beta_{T,11}, \beta_{T,12})' \in \mathbb{R}^3$, where $f(\beta_T, (u, v), (u_{ti}, u_{tj})) = \beta_{T,0} + \beta_{T,11}(u - u_{ti}) + \beta_{T,12}(v - u_{tj})$, u_{ti} are the observed values of electricity demand, b_γ is the smoothing parameter that can be determined, for instance, by (generalized) cross-validation, and $\kappa_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a two-dimensional kernel function such as the multiplicative kernel $\kappa_2(x_1, x_2) = \kappa(x_1) \kappa(x_2)$ with κ being a standard univariate kernel such as the Epanechnikov kernel. See Yao *et al.* (2005) and Fan & Gijbels (1996) for further details.

1.4.4 The estimator $\{\hat{f}_1, \dots, \hat{f}_K\}$

Given the analytic solution of the minimization problem (1.8) we can now define the estimator $\{\hat{f}_1, \dots, \hat{f}_K\}$. The only thing that we have to do is to replace the standardized price-demand functions X_t^* in Eq. (1.10) by their undersmoothed and standardized estimators \tilde{X}_t^* , where $\tilde{X}_t^* = \tilde{X}_t / \|\hat{X}_t\|_2$.

Note that we scale the undersmoothed estimator \tilde{X}_t with the L2 norm of the estimator \hat{X}_t , which is optimally smoothed with respect to the *single* observation X_t . Undersmoothing of the price-demand functions is always important if the target quantity, such as the covariance function $\gamma_T(u, v)$, consist of an average over all functions. The approximation of the norm $\|X_t\|$ does not involve averages over all functions, such that we are better off to use the norm of the classically smoothed curves $\|\hat{X}_t\|$.

Let us denote the estimator of the empirical covariance operator Γ_T by $\hat{\Gamma}_T$, defined as

$$(\hat{\Gamma}_T x)(u) = \int_A^B \hat{\gamma}_T(u, v)x(v)dv \quad \text{for all } x \in L^2[A, B], \quad (1.11)$$

where $\hat{\gamma}_T(u, v)$ is determined by minimizing Eq. (1.10) after replacing $X_t^* = X_t / \|X_t\|$ by $\tilde{X}_t^* = \tilde{X}_t / \|\hat{X}_t\|_2$. Accordingly, we denote the first K ordered eigen-

values and the corresponding eigenfunctions of $\hat{\Gamma}_T$ by $\hat{\lambda}_{T,1} > \dots > \hat{\lambda}_{T,K}$ and $\hat{e}_{T,1}, \dots, \hat{e}_{T,K}$.

The estimator $\{\hat{f}_1, \dots, \hat{f}_K\}$ is then defined as any orthonormal rotation of the orthonormal basis system $\{\hat{e}_{T,1}, \dots, \hat{e}_{T,K}\}$ determined by Eq. (1.8). The trivial case would be to use the empirical eigenfunctions $\hat{e}_{T,1}, \dots, \hat{e}_{T,K}$ directly as basis functions such that $\hat{f}_k = \hat{e}_{T,k}$ for all $k \in \{1, \dots, K\}$. It is generally left to the statistician to choose an appropriate orthonormal rotation scheme, which facilitates the interpretation. In our application we use the well-known VARIMAX-rotation.

Following our assumptions on the data generating process in Eq. (1.2) we use the basis system $\{\hat{f}_1, \dots, \hat{f}_K\}$ in order to re-estimate the functions X_1, \dots, X_T by

$$\hat{X}_t^f = \sum_{k=1}^K \hat{\beta}_{tk} \hat{f}_k, \quad (1.12)$$

where the parameters $\hat{\beta}_{tk}$ are defined according to Eq. (1.3). This is a crucial step of the estimation procedure. Given that our model assumption in Eq. (1.2) is true, the original single cubic smoothing splines estimates \hat{X}_t will be much less efficient estimators of the price-demand functions X_t than the estimators \hat{X}_t^f since the latter use the information of the whole data set.

1.4.5 A note on convergence

Assume that we are able to observe the (unobservable) set of functions $\{X_1, \dots, X_T\}$ as defined in Eq. (1.2) but with deterministic domains $\mathcal{D}(X_t) = [A, B]$. In this case the K empirical eigenfunctions $e_{T,1}, \dots, e_{T,K}$ can be determined from the empirical covariance operator Γ_T as defined in Eq. (1.9) based on the classical definition of the empirical covariance function $\gamma_T(u, v) = K^{-1} \sum_{k=1}^K X_{t_k}(u)X_{t_k}(v)$. Actually, only a subset of at least K linear independent functions, say X_{t_1}, \dots, X_{t_K} , would suffice to determine the K empirical eigenfunctions $e_{T,1}, \dots, e_{T,K}$.

In this case, the determination of the basis system $\{e_{T,1}, \dots, e_{T,K}\}$ is a mere algebraic problem. Furthermore, the space \mathcal{H}_K spanned by the basis system $\{e_{T,1}, \dots, e_{T,K}\}$ does not depend on the data. By the definition in Eq. (1.2), two sets of functions $\{X_1, \dots, X_T\}$ and $\{X_1, \dots, X_{T'}\}$ span the same space \mathcal{H}_K for

all $T' \geq T \geq K$, such that also the corresponding basis systems $\{e_{T,1}, \dots, e_{T,K}\}$ and $\{e_{T',1}, \dots, e_{T',K}\}$ span the same space \mathcal{H}_K .

Though, we do not observe the functions X_t but the noisy discretization points $y_{th} = X_t(u_{th}) + \varepsilon_{th}$. Starting with the first scenario of deterministic domains $\mathcal{D}(X_t) = [A, B]$, the determination of the estimated basis system $\{\hat{e}_{T,1}, \dots, \hat{e}_{T,K}\}$ can be done from the estimated empirical covariance operator $\hat{\Gamma}_T$ defined in Eq. (1.11). Again, this on its own is a mere algebraic problem but it yields to our consistency argument.

The estimated eigenfunction \hat{e}_{Tk} we can write as a continuous function of $\hat{X}_1, \dots, \hat{X}_T$, say $\hat{e}_{Tk} = g_k(\hat{X}_1, \dots, \hat{X}_T)$. By the continuous mapping theorem $\hat{e}_{Tk} = g_k(\hat{X}_1, \dots, \hat{X}_T)$ converges to $e_{Tk} = g_k(X_1, \dots, X_T)$ as $\hat{X}_t(u) \rightarrow X_t(u)$, e.g., in probability as $N_t \rightarrow \infty$ for all $u \in [A, B]$ and $t \in \{1, \dots, T\}$ with $T \geq K$. We additionally have to assume that all involved smoothing parameters go against zero appropriately fast such that $N\underline{b}_K \rightarrow \infty$, $Nb^{opt} \rightarrow \infty$, and $NTb_\gamma \rightarrow \infty$ [Benedetti (1977)]. Eigenfunctions e_k are determined up to sign changes and it is assumed that the correct signs are chosen.

In this sense we can state that $\hat{\mathcal{H}}_K = \text{span}(\hat{e}_{T,1}, \dots, \hat{e}_{T,K})$ converges to $\mathcal{H}_K = \text{span}(e_{T,1}, \dots, e_{T,K})$, which is all we can achieve for (functional) factor models, since the single factors f_k remain unidentifiable.

Finally, it only remains to consider the scenario of random domains $\mathcal{D}(X_t) = [a_t, b_t]$. Also in this case any two points $(u, v) \in [A, B]$ have to be covered by at least K price-demand functions, which is fulfilled asymptotically. By our assumptions the time series (a_t) and (b_t) are processes with $A \leq a_t < b_t \leq B$ and the marginal pdf's of a_t and b_t are given by $f_a(z_a) > 0$ and $f_b(z_b) > 0$ for all $z_a, z_b \in [A, B]$ and $t \in \{1, \dots, T\}$. This yields that

$$\Pr(a_t \in [A, A + \varepsilon]) > 0 \quad \text{and} \quad \Pr(b_t \in [B - \varepsilon, B]) > 0 \quad \text{for any } \varepsilon > 0,$$

such that for $T \rightarrow \infty$ with probability one there are sub-series (a_s) and (b_s) of (a_t) and (b_t) for which the boundary points A and B are accumulation points. From this it follows that as $T \rightarrow \infty$ we can find always more than K functions X_t that cover the points $u, v \in [A, B]$.

To conclude, consistency of our estimation procedure relies on our model

assumptions in Eq.'s (1.1) and (1.2) and is driven by the consistency of the first step estimators of the price-demand functions $X_t(u)$.

1.5 Application

In this section we apply our estimation procedure of the FFM described in Section 1.4 to the data set described in Section 1.2. In Section 1.5.1 we show how to interpret the factors and demonstrate an exemplary analysis of the scores and in Section 1.5.2 we validate the crucial model assumptions.

A drawback of the cross-validation criterion in Eq. (1.5) is that it depends on the unknown dimension K . Therefore, we, first, determine optimal undersmoothing parameters \underline{b}_K for several values of K and, second, choose the dimension K , which minimizes the AIC of Yao *et al.* (2005).

The AIC type criterion indicates an optimal dimension of $K = 2$ (AIC values in Table 1.1 are shown as differences from the lowest AIC value). These first two basis functions are able to explain 99.95% of the variance. The minimization of the cross-validation criterion (1.5) for $K = 2$ yields an undersmoothing parameter of \underline{b}_2 that is only two tenth of the usual cross-validation smoothing parameter b^{opt} ; see Table 1.1.

K	\underline{b}_K/b^{opt}	AIC	Cum. Var.
1	0.1	596.4	92.62%
2	0.2	0	99.95%
3	0.3	52.9	99.97%

Table 1.1: Undersmoothing parameters \underline{b}_K (shown as fractions of the usual cross-validation smoothing parameter b^{opt}), AIC values (shown as differences from the lowest AIC value), and cumulative variances for the dimensions $K \in \{1, 2, 3\}$.

Based on the undersmoothed and scaled estimators $\tilde{X}_t^* = \tilde{X}_t/||\hat{X}_t||$ we compute the estimator $\hat{\gamma}_T$ of the empirical covariance function γ_T by local linear surface smoothing, as explained in Section 1.4.4. The result is shown in the left panel of Figure 1.3. The plot of the estimator $\hat{\gamma}_T$ shows clearly that the sample variance of the standardized price-demand functions \tilde{X}_t^* increases monotonically with electricity demand.

The estimators of the first two empirical eigenfunctions $\hat{e}_{T,1}$ and $\hat{e}_{T,2}$ are determined from the eigendecomposition of a discretized version of the estimated empirical covariance function $\hat{\gamma}_T$ using an equidistant grid of $n \times n$ discretization points of the plane $[A, B]^2$. The estimation of the smooth eigenfunctions by discretizing the smooth covariance function is common in the FDA literature; see, e.g., Rice & Silverman (1991).

In order to find an appropriate number n of discretization points there is the following trade off, which has to be considered: On the one hand n must be small enough that the algorithm to compute the eigendecomposition runs stable. On the other hand n must be great enough that the $n \times n$ -matrix of discretization points forms a good approximation to the covariance function. The choice of $n = 50$ appears to be appropriate for our application. As a robustness check we also tried values of n ranging from 20 to 70, which yield nearly identical results.

We rotate the basis system of the estimated eigenfunctions $\{\hat{e}_{T,1}, \hat{e}_{T,2}\}$ by the VARIMAX-criterion in order to get interpretable basis functions \hat{f}_1 and \hat{f}_2 . The two rotated basis functions \hat{f}_1 and \hat{f}_2 explain 58.63% and 41.32% of the total sample variance of the price-demand functions \hat{X}_t .

It is convenient to choose an appropriate scaling of the graphs of the basis functions \hat{f}_1 and \hat{f}_2 in order to plot them with a reasonable order of magnitude. We scale the graphs by their corresponding average scores $\bar{\hat{\beta}}_i = T^{-1} \sum_{t=1}^T \hat{\beta}_{ti}$ for $i \in \{1, 2\}$. In the right panel of Figure 1.3 the graph of $\hat{f}_1 \bar{\hat{\beta}}_1$ is plotted as a solid line, whereas the graph of $\hat{f}_2 \bar{\hat{\beta}}_2$ is plotted as a dashed line.

Given the basis system $\{\hat{f}_1, \hat{f}_2\}$ we re-estimate the functions X_1, \dots, X_T by Eq. (1.12) such that

$$\hat{X}_t^f = \hat{\beta}_{t1} \hat{f}_1 + \hat{\beta}_{t2} \hat{f}_2.$$

To simplify the notation we write $\hat{X}_t = \hat{X}_t^f$ from now on. The coefficients $\hat{\beta}_{t1}$ and $\hat{\beta}_{t2}$ are determined by OLS regressions of \hat{X}_t simultaneously on \hat{f}_1 and \hat{f}_2 after discretizing the functions at the N_t observed values of electricity demand u_{t1}, \dots, u_{tN_t} . The time series of the scores are shown in Figure 1.4.

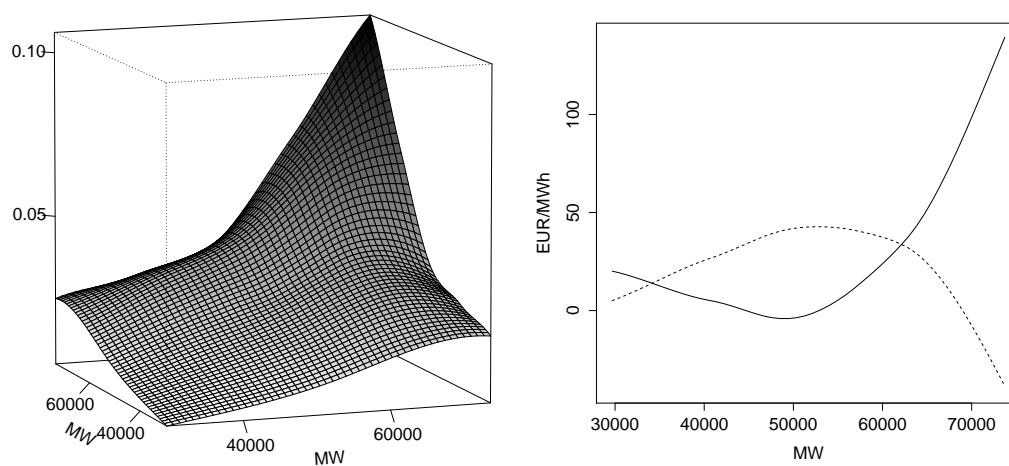


Figure 1.3: LEFT PANEL: Empirical covariance function $\hat{\gamma}_T$. RIGHT PANEL: VARIMAX rotated basis functions \hat{f}_1 (solid line) and \hat{f}_2 (dashed line), scaled by the average scores $\hat{\beta}_{.1}$ and $\hat{\beta}_{.2}$, respectively.

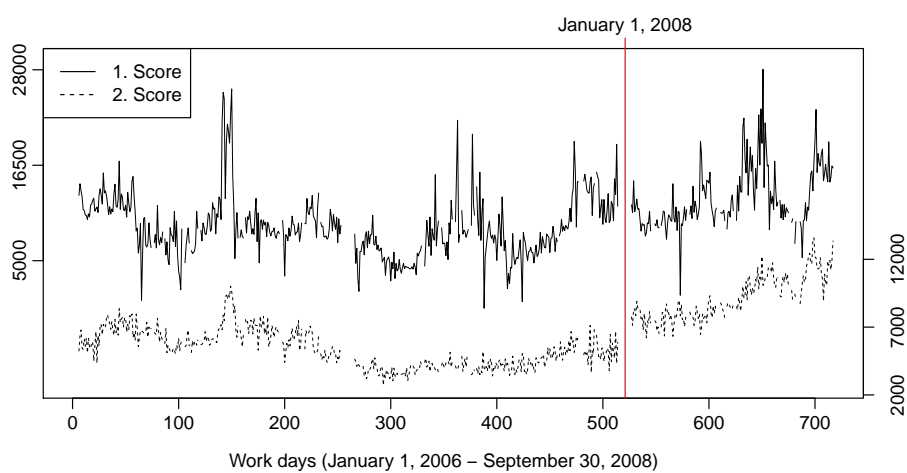


Figure 1.4: Time series of the first scores ($\hat{\beta}_{t1}$) (solid line) and second scores ($\hat{\beta}_{t2}$) (dashed line). The vertical red line separates the initial learning sample from the initial forecasting sample. Gaps in the time series correspond to holidays.

1.5.1 Interpretation of the factors and exemplary analysis of the scores

Remember that we do not use a mean function in our FFM. Consequently, the classical interpretation of the factors \hat{f}_k as perturbations of the mean does not apply. A reasonable interpretation of the estimated factors \hat{f}_1 and \hat{f}_2 can be derived from the classical micro-economic point of view on electricity spot prices; see also the discussion in Section 1.2.

This point of view allows us to interpret the price-demand functions $\hat{X}_t(u) = \hat{\beta}_{t1}\hat{f}_1(u) + \hat{\beta}_{t2}\hat{f}_2(u)$ as daily *empirical merit order curves* or *empirical supply functions*, where the shape of the curves \hat{X}_t is determined by the factors \hat{f}_1 and \hat{f}_2 and the scores $\hat{\beta}_{t1}$ and $\hat{\beta}_{t2}$. For example, steep empirical supply functions have high score ratios $\hat{\beta}_{t1}/\hat{\beta}_{t2}$ and vice versa. Since steep supply functions are associated with high prices we could interpret the first factor \hat{f}_1 as the high-price component and the second factor \hat{f}_2 as the moderate-price component. In general, any interpretation of the factors has to be done with caution since they are only identified up to orthonormal rotations.

Particularly, the scores $\hat{\beta}_{t1}$ and $\hat{\beta}_{t2}$ are useful for a further analysis of the dynamics of the empirical supply functions. For example, researchers or risk analysts, who wish to predict days with high electricity prices, could try to predict days with steep empirical supply functions \hat{X}_t .

Days with steep supply functions represent market situations with capacity constraints, i.e., situations in which power plants with high generation costs are needed to supply the demanded amount of electricity. There are several causes for capacity constraints, such as extreme temperatures or power plant outages.

In fact, the time varying steepness of the empirical supply functions (quantified as time series of score ratios $(\hat{\beta}_{t1}/\hat{\beta}_{t2})$) is Granger-caused by the time series of extreme temperatures (defined as absolute temperature deviations from the mean temperature), where the temperature data is available from the German Weather Service (www.dwd.de). Figure 1.5 shows the p-values of the corresponding Granger-causality tests (Granger, 1969).

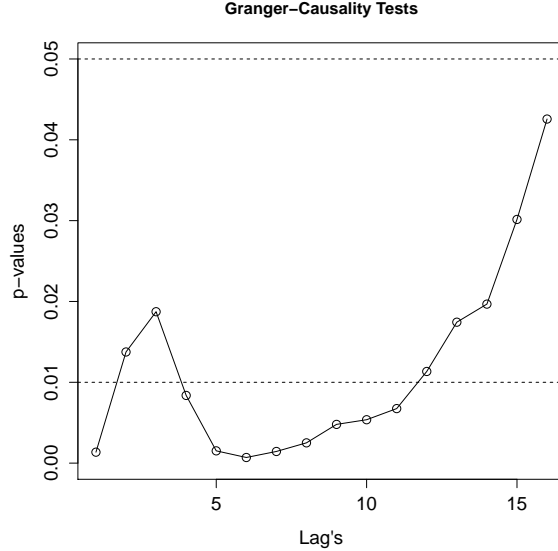


Figure 1.5: P-values of Granger-causality tests of whether the time varying steepness of the price-demand functions (quantified as time series of score ratios $(\hat{\beta}_{t1}/\hat{\beta}_{t2})$) is Granger-caused by past values of the time series of extreme temperatures.

1.5.2 Validation of the model assumptions

The overall in-sample data fit of the estimated spot prices $\hat{y}_{th} = \hat{X}_t(u_{th})$, measured by the R^2 -parameter, is given by $R^2 = 0.92$ and indicates a good model fit. Nevertheless, our implicit stability assumption in (1.2) that $X_t \in \mathcal{H}_K$ for all $t \in \{1, \dots, T\}$, i.e., that all functions X_t are elements of the same space \mathcal{H}_K , may be seen as critical.

In our context it is impossible to validate the stability assumption by statistical tests such as in Benko *et al.* (2009) since we do not assume that the time series of the scores (β_{t1}) and (β_{t2}) are stationary. However, we can compare different basis systems estimated from subsets of the data with each other. The stability assumption can be seen as supported if all of these subset-basis-functions span the same space $\hat{\mathcal{H}}_K$.

We define half-yearly data-subsets 6-1, 6-2, 7-1, 7-2, and one nine-month data-subset 8-1 by choosing according subsets of the index set $\{1, \dots, T\}$ and investigate the R^2 -parameters from subset regressions—such as, e.g., $\hat{e}_1^{(6-1)}(u)$

simultaneously on $\hat{e}_1^{(6-2)}(u)$ and $\hat{e}_2^{(6-2)}(u)$. This assesses whether the eigenfunction $\hat{e}_1^{(6-1)}$ can be seen as an element of the space spanned by the basis system $\{\hat{e}_1^{(6-2)}, \hat{e}_2^{(6-2)}\}$.

The results are given in Table 1.2 and clearly support our assumption that $X_t \in \mathcal{H}_K$ for all $t \in \{1, \dots, T\}$. The R^2 -values with respect to the first eigenfunctions $\hat{e}_1^{(6-1)}, \hat{e}_1^{(6-2)}, \dots, \hat{e}_1^{(8-1)}$ and $\hat{e}_{T,1}$ are all greater than or equal to 0.99. Also, the R^2 -values with respect to the second eigenfunctions $\hat{e}_2^{(6-1)}, \hat{e}_2^{(6-2)}, \dots, \hat{e}_2^{(8-1)}$, and $\hat{e}_{T,2}$ indicate no clear violation of our model assumption.

The R^2 -values with respect to the second eigenfunctions are systematically smaller than those with respect to the first eigenfunctions, since the first order bias term of an estimated eigenfunction is inversely related to the pairwise distances of its eigenvalue to all other eigenvalues; see Benko *et al.* (2009), Theorem 2 (iii). By construction, these distances are greatest for the first eigenvalue.

	$\hat{e}_1^{(6-1)}, \hat{e}_2^{(6-1)}$	$\hat{e}_1^{(6-2)}, \hat{e}_2^{(6-2)}$	$\hat{e}_1^{(7-1)}, \hat{e}_2^{(7-1)}$	$\hat{e}_1^{(7-2)}, \hat{e}_2^{(7-2)}$	$\hat{e}_1^{(8-1)}, \hat{e}_2^{(8-1)}$
$\{\hat{e}_1^{(6-1)}, \hat{e}_2^{(6-1)}\}$	— —	0.99 0.95	1.00 0.96	1.00 0.89	1.00 0.99
$\{\hat{e}_1^{(6-2)}, \hat{e}_2^{(6-2)}\}$	0.99 0.95	— —	0.99 0.83	1.00 0.98	1.00 0.95
$\{\hat{e}_1^{(7-1)}, \hat{e}_2^{(7-1)}\}$	1.00 0.96	0.99 0.83	— —	1.00 0.78	1.00 0.95
$\{\hat{e}_1^{(7-2)}, \hat{e}_2^{(7-2)}\}$	1.00 0.89	1.00 0.98	1.00 0.78	— —	1.00 0.89
$\{\hat{e}_1^{(8-1)}, \hat{e}_2^{(8-1)}\}$	1.00 0.99	1.00 0.95	1.00 0.95	1.00 0.89	— —
$\{\hat{e}_{T,1}, \hat{e}_{T,2}\}$	1.00 0.99	1.00 0.96	1.00 0.95	1.00 0.90	1.00 0.99

Table 1.2: Descriptive validation of the assumption that $X_t \in \mathcal{H}_K$ for all $t \in \{1, \dots, T\}$. The list elements are R^2 -values, which stem from subset regressions of, e.g., the eigenfunction $\hat{e}_1^{(6-1)}$ on the eigenfunctions $\{\hat{e}_1^{(6-2)}, \hat{e}_2^{(6-2)}\}$ in the upper left case with $R^2 = 0.99$.

Finally, we test for (non-)stationarity of the time series of the scores $(\hat{\beta}_{t1})$ and $(\hat{\beta}_{t2})$ using the usual testing procedures such as the KPSS-tests for stationarity and ADF-tests for non-stationarity (with a 5%-significance level for all tests). The results allow us to assume that the time series of the scores $(\hat{\beta}_{t1})$ and $(\hat{\beta}_{t2})$ are non-stationary. Detailed reports are not shown for reasons of space.

This section demonstrates a very good and stable in-sample fit of our FFM. Of course, this cannot guarantee a good out-of-sample performance.

1.6 Forecasting

For our forecasting study we divide the data set into a learning sample of days $t \in \{1, \dots, T_L\}$ and a forecasting sample of days $t \in \{T_L + 1, \dots, T\}$, where the initial $T_L + 1$ corresponds to January 1, 2008 and T to September 30, 2008. The learning sample is used to estimate the parameters and the forecasting sample is used to assess the forecast performance. We enlarge the learning sample after each ℓ days ahead forecast by one day. For $\ell \in \{1, \dots, 20\}$ days ahead forecasts this leads to $T - T_L - (\ell - 1) = 717 - 521 - (\ell - 1) = 197 - \ell$ work days that can be used to assess the forecast performance of our model.

Figure 1.6 shows the whole data set of $717 \cdot 24 = 17,208$ hourly electricity spot prices along with the indicated time spans of the initial learning and forecasting sample. Gaps in the time series correspond to holidays. At least from a visual perspective, the learning sample and the forecasting sample are of comparable complexity.

In the following Section 1.6.1 we discuss forecasting of electricity spot prices using the FFM. In Section 1.6.2 we formally introduce four competing forecasting models (two classical and two FDA models), and in Section 1.6.3 we compare their predictive performance.

As noted above for the FFM, the two competing FDA models also use learning data with electricity spot prices below 200 EUR/MWh only. In contrast, the forecasting sample retains all data, including those with spot prices above 200 EUR/MWh. Advanced outlier forecast procedures, which might yield better predictive performances, are beyond the scope of this paper.

1.6.1 Forecasting with the FFM

The computation of the ℓ days ahead forecast $\hat{y}_{T_L,h}(\ell) \in \mathbb{R}$ of the electricity spot price $y_{T_L+\ell,h}$ given the information set of the learning sample, say \mathcal{I}_{T_L} , involves the computation of the conditional expectation of a nonlinearly transformed random variable, namely $E[f_k(u_{T_L+\ell,h})|\mathcal{I}_{T_L}]$. We approximate the latter using the naive plug-in predictor $\hat{f}_k(\hat{u}_{T_L,h}(\ell))$, where $\hat{u}_{T_L,h}(\ell) = E[u_{T_L+\ell,h}|\mathcal{I}_{T_L}]$. This yields to

$$\hat{y}_{T_L,h}(\ell) = \hat{\beta}_{T_L,1}(\ell)\hat{f}_1(\hat{u}_{T_L,h}(\ell)) + \hat{\beta}_{T_L,2}(\ell)\hat{f}_2(\hat{u}_{T_L,h}(\ell)), \quad (1.13)$$

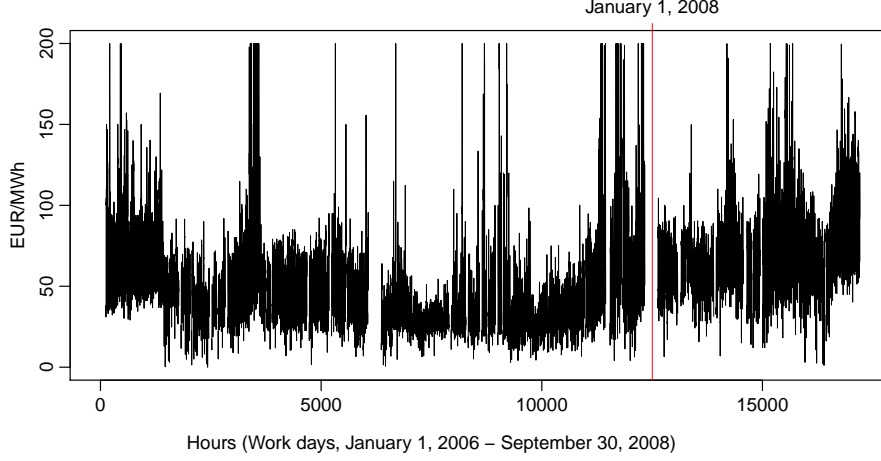


Figure 1.6: The whole data set of 17,208 hourly electricity spot prices. The vertical red line separates the initial learning sample from the initial forecasting sample. Gaps in the time series correspond to holidays.

where $\hat{\beta}_{T_L,1}(\ell)$, $\hat{\beta}_{T_L,2}(\ell)$, and $\hat{u}_{T_L,h}(\ell)$ are the ℓ days ahead forecasts of the scores $\hat{\beta}_{T_L+\ell,1}$, $\hat{\beta}_{T_L+\ell,2}$, and of the electricity demand value $u_{T_L+\ell,h}$.

The naive plug-in predictor $\hat{f}_k(\hat{u}_{T_L,h}(\ell))$ is a rather simple approximation of the conditional expectation $E[f_k(u_{T_L+\ell,h})|\mathcal{I}_{T_L}]$. Here, it performs very well because the basis functions are relatively smooth. In the case of more complex basis functions it might be necessary to improve the approximation using higher order Taylor expansions of \hat{f}_k around $\hat{u}_{T_L,h}(\ell)$.

We use the following univariate SARIMA(0, 1, 6) \times (0, 1, 1)₅-models to forecast the time series of the scores $(\hat{\beta}_{ti})$ with $i \in \{1, 2\}$:

$$(1 - B)(1 - B^5)\hat{\beta}_{ti} = \left(1 + \sum_{l=1}^6 \delta_{il} B^l\right) (1 + \delta_i^S B^5) \omega_{ti}, \quad (1.14)$$

where B is the back shift operator. In order to ensure that the SARIMA models (1.14) are not sample dependent, we select them from a set of reasonable alternative SARIMA models, where all of them are confirmed by the usual diagnostics on the residuals. Each of the confirmed models is applied to different subsets of the learning sample and the final model selection is done by the AIC. As usual, the ℓ days ahead forecasts $\hat{\beta}_{T_L,1}(\ell)$ and $\hat{\beta}_{T_L,2}(\ell)$ are given by the conditional ex-

pectations of $\hat{\beta}_{T_L+\ell,1}$ and $\hat{\beta}_{T_L+\ell,2}$ given the data from the learning sample; see, e.g., Brockwell & Davis (1991).

A first visual impression of the forecast performance is given in Figure 1.7, which compares the 24 hourly spot prices y_{th} with the 1 day ahead forecast of the price-demand function X_t . The 1 day ahead forecast of the price-demand function X_t is defined as

$$\hat{X}_{T_L}(\ell) = \hat{\beta}_{T_L,1}(\ell)\hat{f}_1 + \hat{\beta}_{T_L,2}(\ell)\hat{f}_2 \in L^2[A, B]. \quad (1.15)$$

Additionally, a 95% forecast interval is plotted as a gray shaded band. The forecast interval is computed on the basis of the 95% forecast intervals of the SARIMA forecasts $\hat{\beta}_{T_L,1}(\ell)$ and $\hat{\beta}_{T_L,2}(\ell)$ and has to be interpreted as a conditional forecast interval given the realizations \hat{f}_1 and \hat{f}_2 .

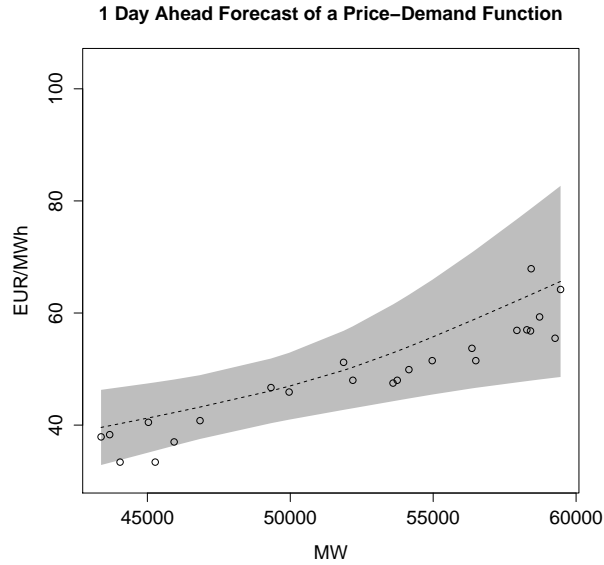


Figure 1.7: A comparison of the 24 hourly electricity spot prices y_{th} (circles) and the 1 day ahead forecast of the price-demand function X_t (dashed line) of January 25, 2008. The 95% forecast interval is plotted as a gray shaded band.

In order to be able to forecast the hourly electricity spot prices y_{th} , we also have to forecast the hourly values of electricity demand u_{th} ; see Eq. (1.13). Given our definition of electricity demand in Section 1.2, a ℓ days ahead forecast of

electricity demand $u_{T_L,h}(\ell)$ involves forecasting gross demand for electricity as well as wind power infeed data. The statistician has to choose appropriate models—one for gross electricity demand such as that proposed in Antoch *et al.* (2010) and another for wind power such as that proposed in Lau & McSharry (2010). For the sake of simplicity, we use the two reference cases of a “persistence” and an “ideal” forecast of electricity demand:

persistence The persistence (or “no-change”) forecast $\hat{u}_{T_L,h}^{persi}(\ell)$ is given by the last value of electricity demand that is still within the learning sample, i.e.,

$$\hat{u}_{T_L,h}^{persi}(\ell) = u_{T_L,h}.$$

ideal The ideal forecast is given by $u_{T_L+\ell,h}$ itself, i.e., $\hat{u}_{T_L,h}^{ideal}(\ell) = u_{T_L+\ell,h}$

This yields a range for possible electricity demand forecasts with bounds that can be easily interpreted.

A first visual comparison of the observed hourly electricity spot prices $y_{T_L+1,h}$ with their 1 day ahead forecasts $\hat{y}_{T_L,h}(1)$ is given in Figure 1.8. The left panel demonstrates the ideal forecast case and shows the spot prices y_{th} (circles) and their 1 day ahead forecasts $\hat{y}_{T_L,h}(1)$ (dotted line) based on the electricity demand forecasts $\hat{u}_{T_L,h}^{ideal}(1)$. The right panel demonstrates the persistence case and shows the spot prices y_{th} (circles) and their 1 day ahead forecasts $\hat{y}_{T_L,h}(1)$ (dotted line) based on the electricity demand forecasts $\hat{u}_{T_L,h}^{persi}(1)$. The 95% forecast intervals are plotted as gray shaded bands. The forecast interval shown in the right panel is much broader than that shown in the left panel. This is because the forecasted electricity spot prices based on the persistence electricity demand forecasts are too high, and higher electricity spot prices have broader 95% forecast intervals; see Figure 1.7.

1.6.2 Competing forecast models

In this section we introduce the four competing forecast models (two classical and two FDA models). The two classical models, referred to as AR and MR models, are archetypal representatives of the classical approaches in the literature on forecasting electricity spot prices; see, e.g., Kosater & Mosler (2006). The AR

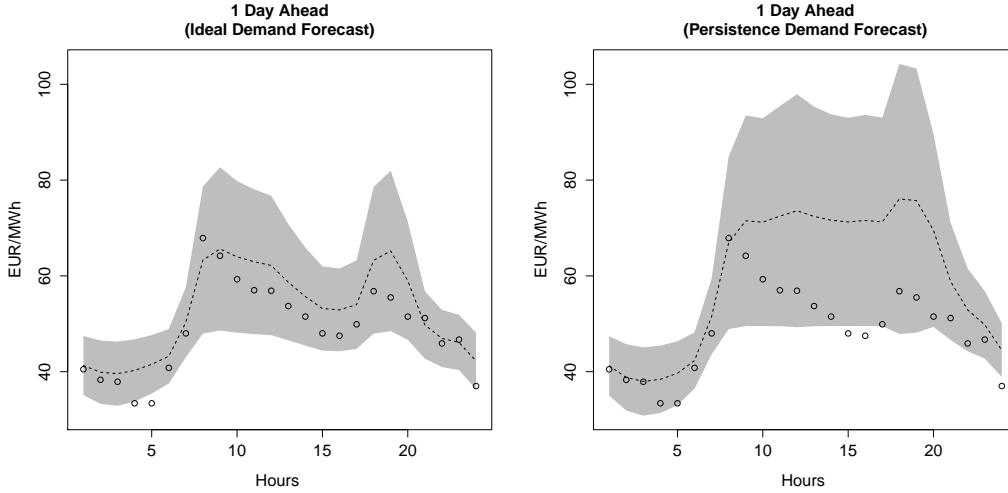


Figure 1.8: LEFT PANEL: Comparison of the spot prices y_{th} (circles) and the 1 day ahead forecasts $\hat{y}_{T_L,h}(\ell)$ (dashed line) of January 25, 2008 based on ideal demand forecasts. LEFT PANEL: Comparison of the spot prices y_{th} (circles) and the 1 day ahead forecasts $\hat{y}_{T_L,h}(\ell)$ (dashed line) of January 25, 2008 based on persistence demand forecasts. BOTH PANELS: The 95% forecast intervals are plotted as gray shaded bands.

model is an autoregressive model and the MR model is the Markov regime switch model for electricity spot prices proposed by Huisman & De Jong (2003).

The two FDA models are the above-discussed DSFM model of Park *et al.* (2009) and the semi-functional partial linear (SFPL) model of Vilar *et al.* (2012). Both of these FDA models have been successfully applied to forecast electricity spot prices [Härdle & Trück (2010) and Vilar *et al.* (2012)] and are expected to be more challenging competitors for our FFM than the two classical models.

Before the formal introduction of the four alternative forecast models, we need some unifying notation. The problem is that the two classical models, AR and MR, are designed to forecast only daily *aggregated* peakload and baseload spot prices defined as

$$y_t^P = \log\left(\frac{1}{12} \sum_{h=9}^{20} y_{th}\right) \quad \text{and} \quad y_t^B = \log\left(\frac{1}{24} \sum_{h=1}^{24} y_{th}\right).$$

In contrast to this, the three FDA models are designed to forecast the *hourly*

electricity spot prices y_{th} . Therefore, we define the forecasts of the peakload-aggregates $\hat{y}_{T_L}^P(\ell|\mathbf{Model})$ and baseload-aggregates $y_{T_L}^B(\ell|\mathbf{Model})$ of the FDA models as

$$\begin{aligned}\hat{y}_{T_L}^P(\ell|\mathbf{Model}) &= \log\left(\frac{1}{12}\sum_{h=9}^{20}\hat{y}_{T_L,h}(\ell|\mathbf{Model})\right) \quad \text{and} \\ \hat{y}_{T_L}^B(\ell|\mathbf{Model}) &= \log\left(\frac{1}{24}\sum_{h=1}^{24}\hat{y}_{T_L,h}(\ell|\mathbf{Model})\right),\end{aligned}$$

where $\hat{y}_{T_L,h}(\ell|\mathbf{Model})$ is the ℓ days ahead hourly electricity spot price forecast of the $\mathbf{Model} \in \{\text{FFM}, \text{DSFM}, \text{SFPL}\}$. By Jensen's inequality these definitions yield aggregated forecasts of the FDA models, which tend to be too high, i.e., $E[y_{T_L+\ell}^A|\mathcal{I}_{T_L}] \leq \hat{y}_{T_L}^A(\ell|\mathbf{Model})$ with $A \in \{P, B\}$. Therefore, the RMSEs of the FDA models shown in Figure 1.9 tend to be inflated and can be interpreted as being conservative.

In the following we formally introduce the four competing forecast models. Further details can be found in Kosater & Mosler (2006), Park *et al.* (2009), and Vilar *et al.* (2012).

AR The first benchmark model is the classical AR(1) model with an additive constant drift component and a time varying deterministic component. The AR model can be defined as

$$y_t^A = d^A + g_t^A + \alpha y_{t-1}^A + \omega_t^A, \quad \omega_t^A \sim \mathcal{N}(0, \sigma_{\omega^A}^2), \quad (1.16)$$

where $A \in \{P, B\}$ refers to the type of aggregation (peakload or baseload), d^A is the constant drift parameter, and g_t^A captures daily, weekly and yearly deterministic effects of the peakload and baseload prices respectively.

MR The second benchmark model is the Markov regime switch model proposed by Huisman & De Jong (2003). The MR model extends the AR model (1.16) and distinguishes between two different regimes $R_t^A \in \{M, S\}$, where M denotes the regime of moderate prices and S denotes the regime of price spikes. The MR

model can be defined as

$$\begin{aligned} y_{M,t}^A &= d^A + \alpha^A y_{M,t-1}^A + \omega_{M,t}^A \\ y_{S,t}^A &= \mu_S^A + \omega_{S,t}^A, \end{aligned} \quad (1.17)$$

where $A \in \{P, B\}$ refers to the type of aggregation (peakload or baseload), $\omega_{M,t}^A \sim \mathcal{N}(0, \sigma_{MA}^2)$, and $\omega_{S,t}^A \sim \mathcal{N}(0, \sigma_{SA}^2)$. The conditional probabilities of the transitions from one regime to another given the regime at $t-1$ are captured by the transition matrix

$$\begin{pmatrix} \text{P}\left(R_t^A = M | R_{t-1}^A = M\right) & \text{P}\left(R_t^A = M | R_{t-1}^A = S\right) \\ \text{P}\left(R_t^A = S | R_{t-1}^A = M\right) & \text{P}\left(R_t^A = S | R_{t-1}^A = S\right) \end{pmatrix} = \begin{pmatrix} q & 1-p \\ 1-p & p \end{pmatrix}$$

and have to be estimated, too.

DSFM The third model, the DSFM of Park *et al.* (2009), is a functional factor model, which is very similar to our FFM. Its application to electricity spot prices, as suggested by Härdle & Trück (2010), differs from our application, since it models the hourly spot prices y_{th} based on the classical time series point of view on electricity spot prices. That is, Härdle & Trück model and forecast non-parametric price-hour functions, say $\chi_t(h)$, and thereby fail to consider the merit order model. The DSFM can be written as

$$y_{th} = \chi_t(h) + \omega_{th}, \quad h \in \{1, \dots, 24\} \quad (1.18)$$

with $\chi_t \in L^2[1, 24]$ defined as

$$\chi_t(h) = f_0^{DSFM}(h) + \sum_{l=1}^L \beta_{tl}^{DSFM} f_l^{DSFM}(h),$$

where $f_0^{DSFM}(h)$ is a non-parametric mean function, $f_l^{DSFM}(h)$ are non-parametric functional factors, β_{tl}^{DSFM} are the univariate scores, and ω_{th} is a Gaussian white noise process.

Park *et al.* suggest selecting the number of factors L by the proportion of explained variation. We choose the factor dimension $\hat{L} = 2$, since this factor di-

mension yields the same proportion of explained variation as the factor dimension $\hat{K} = 2$ for our FFM.

Given the estimates of the time invariant model components, $\hat{f}_0^{DSFM}(h)$, $\hat{f}_l^{DSFM}(h)$, and \hat{L} , forecasting of the daily price-hour functions $\chi_t(h)$ can be done by forecasting the estimated univariate time series of scores. As for our FFM we use SARIMA models to forecast the univariate time series $(\hat{\beta}_{t1}^{DSFM})$ and $(\hat{\beta}_{t2}^{DSFM})$, where the model selection procedure for the SARIMA models is the same as for our FFM.

SFPL The fourth model, the SFPL model of Vilar *et al.* (2012), is a very recent functional data model, which was exclusively designed for forecasting electricity spot prices. The SFPL has the nice property of allowing us to include the values of electricity demand u_{th} as additional co-variables. Vilar *et al.* use dummy variables for work days and holidays as additional covariates, which we do not have to do, since we consider only work days. Note that, like the DSFM, the SFPL model uses price-hour functions $\chi_t(h)$ and therefore does not consider the merit order model. The definition of the SFPL is given in the following:

$$y_{t+\ell,h} = \alpha u_{th} + m(\chi_t(h)) + \omega_{th}, \quad (1.19)$$

where $m : L^2[1, 24] \rightarrow \mathbb{R}$ is a function that maps the price-hour function χ_t to a real value and ω_{th} is a Gaussian white noise process.

Forecasting electricity spot prices y_{th} with the SFPL model can be easily done using the R package `fda.usc` of Febrero-Bande & Oviedo de la Fuente (2012). However, in order to compute the forecasts $\hat{y}_{T_L,h}(\ell|\text{SFPL})$ of the electricity spot prices $y_{T_L+\ell,h}$ we also need forecasts of the electricity demand values $u_{T_L+\ell,h}$. We cope with this problem as suggested above for our FFM by using a persistence forecast and an ideal forecast.

1.6.3 Evaluation of forecast performances

The two plots of Figure 1.9 show the values of the RMSEs for the $\ell \in \{1, \dots, 20\}$ days ahead forecasts of the peakload prices (left panel) and the baseload prices (right panel). The two gray shaded regions in each plot show the possible RMSE

values of the FFM (solid line borders) and the SFPL model (dotted line borders). The lower bounds of the regions are based on the ideal electricity demand forecast $\hat{u}_{T_L,h}^{ideal}(\ell)$. The upper bounds are based on the persistence forecast of electricity demand $\hat{u}_{T_L,h}^{persi}(\ell)$.

The poor performance of the two classical time series models, AR and MR, in comparison to the three FDA models, FFM, DSFM, and SFPL, can be explained by the different approaches to model the aggregated peakload and baseload prices. The two classical models try to forecast the aggregated prices directly, whereas the three FDA models try to forecast the hourly electricity spot prices; aggregation is done afterward.

The superior performance of our FFM in comparison to the other two FDA models, DSFM and SFPL, can be explained by the FFM's explicit consideration of the merit order model. Both models, the DSFM and the SFPL, work with daily price-hour functions $\chi_t(h)$, which are based on a rather simple transfer of the classical time series point of view to a functional data point of view. By contrast, the FFM works with daily price-demand functions $X_t(u)$, which are based on the merit order model, the most important model for explaining electricity spot prices; see our discussion in Section 2.3. Finally, the DSFM generally performs better than the SFPL model. This might be explained by the fact that the SFPL model of Vilar *et al.* (2012) is an autoregressive model of order one. Vilar *et al.* do not discuss the possibility of extending the order-structure of their SFPL model.

The above study of the RMSE's only gives us insights into the forecast performances with respect to point forecasts. In order to complement the forecast comparisons, we also consider interval forecasts. In this regard, the interval score, proposed by Gneiting & Raftery (2007), is a very informative statistic. The interval score can be defined as

$$S_{\alpha}^{int}(h, \ell) = (\hat{b}_u - \hat{b}_l) + \frac{2}{\alpha}(\hat{b}_l - y_{T_L+\ell,h})\mathbb{I}\{y_{T_L+\ell,h} < \hat{b}_l\} + \frac{2}{\alpha}(y_{T_L+\ell,h} - \hat{b}_u)\mathbb{I}\{y_{T_L+\ell,h} > \hat{b}_u\},$$

where $\hat{b}_u = \hat{b}_{u,T_L,h}(\ell)$ and $\hat{b}_l = \hat{b}_{l,T_L,h}(\ell)$ are the lower and upper bounds of the

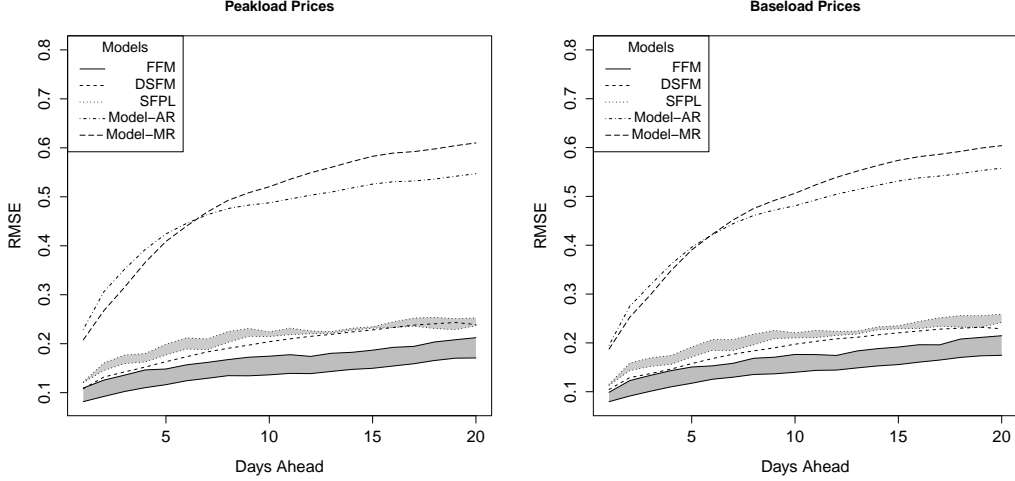


Figure 1.9: Root mean squared errors of the FFM (solid lines) and the alternative models, DSFM (short-dashed lines), SFPL (dotted lines), AR (dash-dotted lines), and MR (long-dashed lines) for peakload prices y_t^P (left panel) and baseload prices y_t^B (right panel). The gray shaded regions for the FFM and the SFPL model are lower bounded based on the ideal demand forecast, and upper bounded based on the persistence forecast.

$(1 - \alpha)\%$ forecast interval for the electricity spot price $y_{T_L+\ell,h}$. The interval score punishes a broad prediction interval ($\hat{b}_u - \hat{b}_l$) and adds an additional punishment if the actual observation $y_{T_L+\ell,h}$ is not within the prediction interval. In general, a lower interval score is a better one.

Unfortunately, we cannot compute the interval scores for all five models. For example, Vilar *et al.* (2012) do not propose any prediction intervals for the SFPL model. Furthermore, while it is easy to compute forecast intervals of the FFM and the DSFM for hourly sport prices, it is not trivial to compute them for the aggregated (peakload and baseload) prices.

Therefore, we focus on the hourly forecasts of electricity spot prices of the FFM and DSFM models. For both models, the 95% forecast intervals can be computed on the basis of the 95% forecast intervals of the SARIMA forecasts given the estimated factors.

Due to the enlargement of the learning sample after each ℓ days ahead forecast by one day and due to pooling all hours $h \in \{1, \dots, 24\}$ we have for each ℓ days ahead forecast $24 \cdot (197 - \ell)$ interval scores $S_\alpha^{int}(h, \ell)$ in order to compare the ℓ

days ahead forecast performances of our FFM and the DSFM. In Figure 1.10 we present the (trimmed) mean values of these pooled interval scores $S_\alpha^{int}(h, \ell)$ with $\alpha = 0.05$ for each $\ell \in \{1, \dots, 20\}$. The 5% trimmed mean values are used, since for both models there are some extreme values of the interval score (due to the outliers in the forecast sample), which distort the mean values.

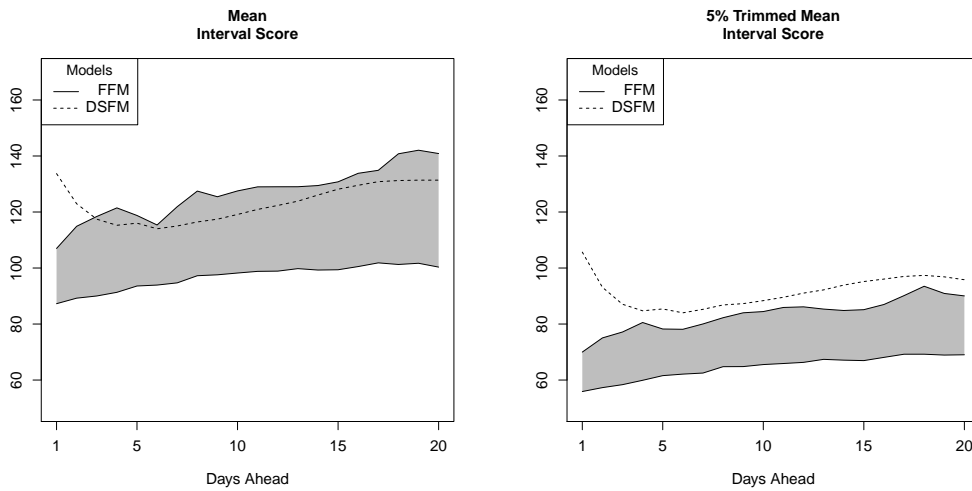


Figure 1.10: Mean and trimmed mean values of the interval scores $S_\alpha^{int}(h, \ell)$, pooled for all hours $h \in \{1, \dots, 24\}$. A low interval score stands for precise predictions with narrow prediction intervals. The dashed line corresponds to the interval scores of the DSFM of Park *et al.* (2009). The gray shaded regions for the FFM are lower bounded based on the ideal demand forecast, and upper bounded based on the persistence forecast.

Figure 1.10 clearly confirms the good forecast performance of the FFM. Besides some technical issues, the main conceptual difference between the DSFM and our FFM is that the DSFM works with daily price-*hour* functions $\chi_t(h)$, whereas our FFM works with daily price-*demand* functions $X_t(u)$, which are suggested by the merit order model. This demonstrates that the consideration of the merit order model yields better point forecasts as well as better interval forecasts.

1.7 Conclusion

In this paper we suggest interpreting hourly electricity spot prices as noisy discretization points of smooth price-demand functions. This functional data perspective on electricity spot prices is motivated as well as theoretically underpinned by the merit order model—the most important pricing model for electricity spot prices.

We propose a functional factor model in order to model and forecast the non-stationary time series of price-demand functions and discuss a two-step estimation procedure. In the first step we estimate the single price-demand functions from the noisy discretization points. In the second step we robustly estimate from these a finite set of common basis functions. The careful consideration of the merit order model yields a very parsimonious functional factor model with only two common basis functions, which together explain over 99% of the total sample variation of the price-demand functions.

Our approach allows us to separate the total variations of electricity spot prices into one part caused by the variations of the merit order curves (mainly variations of input-costs) and another part caused by the variations of electricity demand. The first part is modeled by our FFM and the second part can be modeled by specialized methods proposed in the literature. We decided to keep the model parsimonious; nevertheless, it is easily possible to include the input cost for resources (coal, gas, etc.) into our FFM. Researchers are invited to extend the FFM for these co-variables.

The presentation of our functional factor model is concluded by a real data application and a forecast study which compares our FFM with four alternative time series models that have been proposed in the electricity literature. The real data application demonstrates the use of the functional factor model and a possible interpretation of the unobserved common basis functions. The forecast study clearly confirms the power of our functional factor model and the use of price-demand functions as underlying structures of electricity spot prices in general.

Chapter 2

A fundamental model for electricity spot prices using functional data analysis

Contrary to usual financial market data, electricity spot prices can be predicted very well by a simple micro economic auction model, the so-called merit order model. This model assumes that the electricity spot prices come from price functions, called merit order curves.

Motivated by the merit order model, we suggest a functional data model for the analysis of electricity spot prices, but here we face a rare and interesting case of functional data. The noisy discretization points of the price functions are only observed within random subintervals of the electricity demand domain, which causes problems similar to the case of sparse functional data. We extend existing theoretical results by considering the asymptotic bias and variance of the local linear estimator of the mean and the covariance function for the case of time dependent functional data observed on random subintervals.

2.1 Introduction

The merit order model assumes that the spot prices at electricity exchanges are based on the marginal generation costs of the last power plant that is required

to cover the demand. The resulting (monotonically increasing) merit order curve reflects the increasing generation costs of a specific portfolio of power plants. Often, nuclear and lignite plants cover the minimal demand for electricity. Higher demand is mostly served by hard coal and gas fired power plants [Burger *et al.* (2008), Chapter 4, and Nestle (2012) Figure 1].

On the European Electricity Exchange (EEX), as at many other electricity exchanges, each of the 24 hourly electricity spot prices of a day t is determined in a separate auction, and all 24 auctions are settled simultaneously at 12 am at day $t - 1$ [Ockenfels *et al.* (2008), Chapters 2 & 4]. In contrast to the temporal dependencies between whole days, it is not necessary to consider temporal dependencies of the 24 spot prices within a day t due to the simultaneity of settlements [Huisman *et al.* (2007)]. The prices are determined by the intersection of the aggregated supply and demand functions, where in the case of perfect competition the supply function equals the merit order curve.

Daily mean merit order curves, or more generally, price functions \mathcal{P}_t can be estimated from the $n = 24$ data pairs $(P_{it}, \mathcal{D}_{it})$ of hourly electricity spot prices P_{it} and electricity demand values \mathcal{D}_{it} with $i \in \{1, \dots, n\}$ and $t \in \{1, \dots, T\}$; see the left panel of Figure 2.1.

Modeling and forecasting electricity spot prices using this functional data point of view turns out to be very useful and yields a superior forecast performance [Liebl (2013)]. This paper can be seen as an extension of Liebl (2013) inasmuch as we use the concept of price functions \mathcal{P}_t , but augment it by two additional covariables.

The first covariable is a dummy (or indicator) variable that allows us to condition on the peak hours (8am – 8pm) and off-peak hours (all non-peak hours). This leads to daily mean *peak* price functions and *off-peak* price functions, which account for the different market situations during the day- and night-time. The left panel of Figure 2.1 demonstrates the effect of conditioning on peak and off-peak hours for a randomly chosen day. This was not considered in Liebl (2013), since the estimation procedure used therein would suffer from this halving of the amount of discretization points per price function.

The introduction of this dummy variable corresponds to a trivial split of the dataset. Notationally, we do not differentiate between peak and off-peak price

functions, but simply refer to them as price functions \mathcal{P}_t . If not otherwise stated, all Figures are produced for the case of peak price hours.

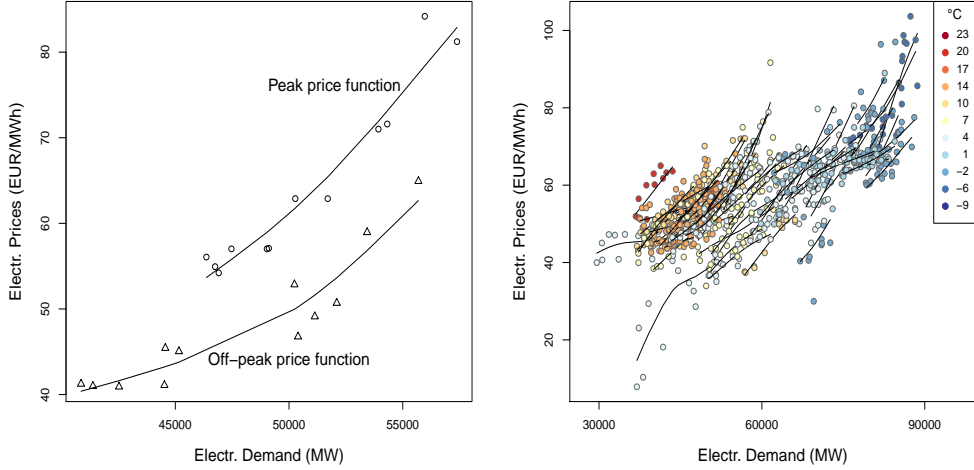


Figure 2.1: LEFT PANEL: Off-peak and peak price functions of the randomly chosen day October 12, 2010. RIGHT PANEL: Scatter plot of daily peak price functions (from March 14, 2010 to March 14, 2011); only every 3rd price function is plotted in order to maintain a clear plot.

The second covariable is the daily air temperature, which means that we regard electricity spot prices as noisy discretization points of daily *bivariate* random price functions \mathcal{P}_t with prediction points electricity demand \mathcal{D}_{it} and temperature \mathcal{T}_t , such that

$$P_{it} = \mathcal{P}_t(\mathcal{D}_{it}, \mathcal{T}_t) + \epsilon_{it}, \quad (2.1)$$

where ϵ_{it} is an iid mean zero error term having finite variance. The dependencies of electricity spot prices P_{it} over the time t are modeled using the assumption that the price functions \mathcal{P}_t come from a weakly stationary process of square integrable random functions. We do not use hourly air temperature values, since the merit order curve is determined on a daily basis.

Using the well-known Karhunen-Loève decomposition, the time series of bi-

variate random price functions can be written as

$$\begin{aligned}\mathcal{P}_t(\mathcal{D}_{it}, \mathcal{T}_t) &= \mu(\mathcal{D}_{it}, \mathcal{T}_t) + Z_t(\mathcal{D}_{it}, \mathcal{T}_t) \quad \text{with} \\ Z_t(\mathcal{D}_{it}, \mathcal{T}_t) &= \sum_{k=1}^{\infty} \beta_{tk}(\mathcal{T}_t) p_k(\mathcal{D}_{it}, \mathcal{T}_t),\end{aligned}\tag{2.2}$$

where μ is a common bivariate mean function, the function $p_k(\cdot, \mathcal{T}_t)$ is the k th conditional eigenfunction given the temperature value \mathcal{T}_t , which belongs to the k th (decreasingly ordered) conditional eigenvalue $\lambda_k(\mathcal{T}_t)$ of the autocovariance operator of the random functions \mathcal{P}_t . The random variables $\beta_{tk}(\mathcal{T}_t)$ are the conditional principal component scores, where our assumption in Eq. (2.1) of a weakly stationary process of random functions (\mathcal{P}_t) yields that the time series ($\beta_{tk}(\mathcal{T}_t)$) are weakly stationary stochastic processes for all $k \in \{1, 2, \dots\}$ and fixed $\mathcal{T}_t = \tau$ [see, e.g., Rice & Silverman (1991) for the introduction of the unconditional Karhunen-Loève decomposition].

The effect of electricity demand on electricity spot prices can be described by a simple, strictly monotonically increasing relationship; see the left panel of Figure 2.1. The effect of temperature on electricity spot prices is a bit more involved. On the one hand, high/low temperatures induce a left/right shifts of the price functions \mathcal{P}_t , caused by seasonal differences in the composition of the power plant portfolio; see the right panel of Figure 2.1.

On the other hand, there is also an indirect relationship through electricity demand, which itself also depends on temperature [Engle *et al.* (1986), Harvey & Koopman (1993)]. The latter dependency is visualized by the red line in Figure 2.2, which shows a local linear estimate for the smooth conditional mean function of electricity demand given the observed temperature values.

The scatter plot shown in Figure 2.2 visualizes two further very interesting peculiarities of our functional data set. First, the discretization points of single price functions are clustered. Second, due to physical constraints there are lower and upper bounds, say $a(\tau)$ and $b(\tau)$, for electricity demand, which depend on the temperature τ ; see the gray shaded region in Figure 2.2. The lower bound represents the amount of electricity needed to ensure the minimal maintenance level (e.g., for medical care, etc). The upper bound for electricity demand is given

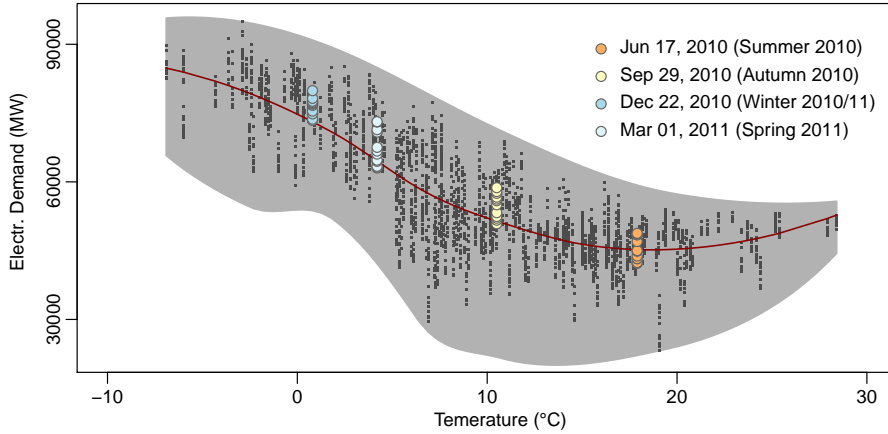


Figure 2.2: Scatter plot of the prediction points $(\mathcal{D}_{it}, \mathcal{T}_t)$ for the time span one year before the moratorium (from March 15, 2010 to March 14, 2011). The red line visualizes a local linear estimate for the smooth conditional mean function of electricity demand given the observed temperature values. The prediction points of four particular days are emphasized by color-filled circle points, where the days are chosen to have season-specific median temperatures; see Figure 2.1 for the color code.

by the maximum amount of electricity that can be provided by the power plant portfolio, since demand cannot exceed supply. In the following we refer to the set that is described by the interval $[a(\tau), b(\tau)]$ as

$$\mathcal{S} = \{(\delta, \tau) \in \mathbb{R}^2 \mid \delta \in [a(\tau), b(\tau)], \tau \in \mathcal{I}_{\mathcal{T}}\},$$

where $\mathcal{I}_{\mathcal{T}} \subset \mathbb{R}$ and $0 < a(\tau) < b(\tau) < \infty$. Here and in what follows, we denote realizations of \mathcal{D}_{it} as δ and realizations of \mathcal{T}_t as τ .

The model in Eq. (2.1) is a so-called fundamental model for electricity spot prices based on the fundamental (or exogenous) variables electricity demand and air temperature. Despite the fact that the model is theoretically well underpinned by the merit order model, it sets a counterpart to the literature on modeling electricity spot prices, which is dominated by the usage of classical time series models such as autoregressive, jump diffusion, or Markov regime switch models.

Classical time series models also try to capture the temperature effects and account for them by the incorporation of classical seasonal additive model compo-

nents. However, as discussed above, the merit order model suggests other, more involved relationships. Furthermore, the adoption of classical time series model for electricity spot prices is not straight forward. This is reflected by the fact that usually only daily aggregates of the hourly electricity spot prices are considered, such as daily averages [Weron *et al.* (2004), Kosater & Mosler (2006), and Koopman *et al.* (2007)] or each hour separately [Karakatsani & Bunn (2008)]. These unnatural aggregations and separations of the data necessarily come with great losses in information.

In order to estimate the model components in Eq. (2.2) we use functional principal component analysis (FPCA). FPCA is a core method for analyzing functional data and is well studied for the situation when the functions can be assumed to be observed precisely [Ramsay & Dalzell (1991), Besse & Ramsay (1986), and Hall & Hosseini-Nasab (2006)]. In economic contexts, however, we usually observe only finitely many, noisy discretization points per function. Generally, there are two different FPCA strategies in this case. One can either pre-smooth each function and then estimate the covariance function. Or one can estimate the covariance function directly by smoothing the pooled noisy discretization points. The latter possibility has the advantage that it works also for sparse functional data with only a few noisy discretization points per function randomly spread over the whole common domain.

FPCA for sparse functional data is considered for the first time in Yao *et al.* (2005). Based on a “large-T, small-n” asymptotic, the authors derive upper bounds describing uniform-in-probability convergence of the empirical mean function, covariance function, eigenvalues and eigenfunctions. More precise uniform-in-probability convergence rates for the eigenvalues and eigenfunctions are derived in the follow-up paper of Hall *et al.* (2006). Li & Hsing (2010) reconsider the problem and derive uniform, almost sure convergence rates. In the latter two papers, the authors take a unifying point of view and derive convergence rates for all possible scenarios between “large-T, small-n” and “large-T, large-n” asymptotics.

The case of sparse functional data is related to our situation, but there are three important differences. First, it is rather borderline to speak of sparse functional data, when considering $n = 12$ discretization points per price function.

Therefore, we adopt the unifying point of view as in Hall *et al.* (2006) and Li & Hsing (2010) and allow the number of discretization points $n = (T)$ to remain bounded as T diverges, as well as to diverge with T . Second, we consider a time series context, which has not been done for sparse functional data so far. Third, in contrast to the usual assumption, the discretization points of the price functions are not spread over the whole domain, but only observed within some random subintervals. This has implications for the estimation of the principal component scores, which we discuss in Section 2.4.3.

All of the above-cited theoretical work on sparse functional data considers local linear estimators [Fan & Gijbels (1996)]. We contribute to this literature by deriving explicit asymptotic bias and variance expressions of the multivariate local linear estimator, where we account for the temporal dependencies between the random price functions, as well as the dependencies between spot prices coming from the same price functions.

In order to demonstrate the usefulness of our model we analyze the effects of Germany's nuclear moratorium on March 14, 2011. This event describes a natural experiment, since in the course of it eight nuclear power plants were phased-out [Nestle (2012)]. The data set analyzed in this paper covers exactly one year before and one year after Germany's nuclear power phase-out. We apply our model separately to these two time spans in order to contrast the different market situations. This separation of the data set will not be made notationally explicit, but will be clear from the context.

2.2 Data

As already mentioned above, we demonstrate our model using electricity spot prices of the German power market traded at the European Energy Exchange (EEX) in Leipzig. The German power market is the biggest power market in Europe in terms of consumption. The EEX spot price is of fundamental importance as benchmark and reference point for other markets, such as over the counter and forward markets [Ockenfels *et al.* (2008), Chapter 1].

Nowadays, electricity markets usually provide purchase guarantees for renewable energy sources (RES). This applies to the German electricity market,

where our data comes from, as well as to almost all other electricity markets in the world. In this case it is not gross electricity demand, say \mathcal{G}_{it} , that is relevant for pricing, but rather gross electricity demand minus electricity infeeds from RES. Correspondingly, (residual) electricity demand \mathcal{D}_{it} is defined as $\mathcal{D}_{it} = \mathcal{G}_{it} - \text{RES.Infeed}_{it}$.

The data for our analysis stem from three different publicly available sources. The hourly spot prices of the German electricity market are provided by the European Energy Exchange (www.eex.com), hourly values of Germany's gross electricity demand are provided by the European Network of Transmission System Operators for Electricity (www.entsoe.eu), German wind and solar power infeed data are provided by the EEX Transparency Platform (www.transparency.eex.com), and German air temperature data are available from the German Weather Service (www.dwd.de).

Very few (only 0.2%) of the data pairs $(P_{it}, \mathcal{D}_{it})$ with prices $P_{it} > 200$ EUR/MWh have to be treated as outliers since they cannot be explained by the merit order model. Even in exceptional situations the marginal costs of electricity production do not exceed the value of 200 EUR/MWh. Prices above this threshold are referred to as price spikes and have to be explained using an additional scarcity premium [Burger *et al.* (2008), Chapter 4]. We simply set the few spot prices that exceed the value of 200 EUR/MWh equal to 200 EUR/MWh. The analysis of price spikes is a research topic on its own [Christensen *et al.* (2009)] and is not within the scope of this paper.

2.3 Model & estimators

We model the electricity spot price $P_{it} \in \mathbb{R}$ of hour $i \in \{1, \dots, n\}$ and day $t \in \{1, \dots, T\}$ as a noisy discretization point of a weakly stationary time series of square integrable, bivariate random functions with hourly electricity demand $\mathcal{D}_{it} \in [a(\mathcal{T}_t), b(\mathcal{T}_t)]$ and daily temperature $\mathcal{T}_t \in \mathcal{I}_{\mathcal{T}}$ as prediction points, where $[a(\mathcal{T}_t), b(\mathcal{T}_t)] \subset \mathbb{R}$ and $\mathcal{I}_{\mathcal{T}} \subset \mathbb{R}$.

Bringing together Eq.'s (2.1) and (2.2) we can write our statistical model for the electricity spot price as the following nonparametric multivariate regression

problem

$$P_{it} = \mu(\mathcal{D}_{it}, \mathcal{T}_t) + Z_t(\mathcal{D}_{it}, \mathcal{T}_t) + \epsilon_{it} \quad \text{with} \quad (2.3)$$

$$Z_t(\mathcal{D}_{it}, \mathcal{T}_t) = \sum_{k=1}^{\infty} \beta_{tk}(\mathcal{T}_t) p_k(\mathcal{D}_{it}, \mathcal{T}_t), \quad (2.4)$$

where ϵ_{it} is a uncorrelated, mean zero error term having finite variance and μ is a common bivariate mean function. The random functions $Z_t \in L_2[\mathcal{S}]$ are elements of the space of square integrable functions on the compact domain $\mathcal{S} \subset \mathbb{R}^2$ with $E(Z_t(\delta, \tau)) = 0$ for all $\delta \in [a(\tau), b(\tau)]$ and $\tau \in \mathcal{I}_{\mathcal{T}}$.

Model (2.3) is a nonparametric multivariate regression model with a nonstandard error term, which is composed of a random functional part Z_t and a random scalar part ϵ_{it} . In order to estimate the mean function μ , we have to take into account the stochastics of this composed error term. The scalar error term does not introduce any remarkable difficulties. However, the functional error term imposes within-function covariances, which are typical for functional error terms, as well as between-function covariances, which are caused by our time series context. See Section 2.3.1 below for a more formal discussion of the latter covariances.

Besides estimation of the mean function μ , it is of practical interest to estimate the unobserved smooth price functions $\mathcal{P}_t(\cdot, \mathcal{T}_t) \in L_2[a(\mathcal{T}_t), b(\mathcal{T}_t)]$ for the observed realizations of \mathcal{T}_t . This requires to approximate the unobserved realizations of random functions $Z_t(\cdot, \mathcal{T}_t)$ (remember Eq. (2.2)). The Karhunen-Loève decomposition in Eq. (2.4) justifies to approximate specific realizations of $Z_t(\cdot, \mathcal{T}_t)$ using the first, say K , principal component scores and eigenfunctions, i.e.,

$$Z_t(\cdot, \mathcal{T}_t) \approx \sum_{k=1}^K \beta_{tk}(\mathcal{T}_t) p_k(\cdot, \mathcal{T}_t),$$

where $\beta_{tk}(\mathcal{T}_t)$ is the k th conditional principal component score given the temperature value \mathcal{T}_t . Indeed the well known best-basis property of the eigenfunctions assures that a rather small number of, e.g., $K = 2$ or $K = 3$ leads to satisfactory approximations in practical problems. Here and in what follows we assume the eigenfunctions $p_k(\cdot, \mathcal{T}_t)$ to be ordered according to the decreasing sequence of nonzero eigenvalues, say $\lambda_k(\mathcal{T}_t)$ [see Section 8.2.3 in Ramsay & Silverman (2005)].

A prerequisite to estimate the first K conditional eigenfunctions and eigenvalues is the estimation of the conditional covariance function

$$\gamma((\delta, \delta'), \tau) = \text{Cov}(Z_t(\delta, \tau), Z_t(\delta', \tau)).$$

Note that we use only one temperature value τ , since we anyways observe only one realization of the prediction point \mathcal{T}_t per each time point t . The corresponding covariance operator Γ is then defined by

$$(\Gamma z)(\delta, \tau) = \int_{[a(\tau), b(\tau)]} \gamma((\delta, \delta'), \tau) z(\delta', \tau) d\delta' \quad (2.5)$$

for any function $z(\cdot, \tau) \in L_2[a(\tau), b(\tau)]$.

Once an estimate of the covariance operator is determined, the conditional eigenfunctions $p_k(\cdot, \tau)$, the corresponding eigenvalues $\lambda_k(\tau)$, as well as the conditional principal component scores

$$\beta_{tk}(\tau) = \int_{[a(\tau), b(\tau)]} Z_t(\delta, \tau) p_k(\delta, \tau) d\delta$$

can be (in principle) estimated using FPCA. However, the above discussed non-standard random design of the predictions points \mathcal{D}_{it} and \mathcal{T}_t imposes some further difficulties with respect to the estimation of the conditional principal component scores $\beta_{tk}(\tau)$. A possible solution is discussed in Section 2.4.3.

In the following we demonstrate that the estimation of the mean function μ and the estimation of the covariance function γ are both essentially the same nonparametric multivariate regression problems, which then allows us to develop our theory with respect to a common unifying regression model. The reason for this is simple: the problem of estimating the variance of a random variable, say X , is equivalent to the problem of estimating the mean of the transformed random variable $(X - E(X))^2$. Correspondingly, in order to estimate the covariance function γ , we simply have to transform the original data triples $(P_{it}, (\mathcal{D}_{it}, \mathcal{T}_t))$ into new data tuples

$$(C_{ijt}, (\mathcal{D}_{it}, \mathcal{D}_{jt}, \mathcal{T}_t)) \text{ with } t \in \{1, \dots, T\},$$

where $C_{ijt} \in \mathbb{R}$ are so-called ‘‘raw covariances’’ defined as

$$C_{ijt} = (P_{it} - \mu(\mathcal{D}_{it}, \mathcal{T}_t))(P_{jt} - \mu(\mathcal{D}_{jt}, \mathcal{T}_t)) \quad (2.6)$$

[see also Yao *et al.* (2005)].

By inserting Eq. (2.3) into Eq. (2.6), it follows that each raw covariance actually comes from the following nonparametric multivariate regression problem

$$C_{ijt} = \gamma((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t) + G_t((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t) + \varepsilon_{ijt} \quad (2.7)$$

again with a nonstandard error term composed of a functional part and a scalar part

$$\begin{aligned} G_t((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t) &= Z_t(\mathcal{D}_{it}, \mathcal{T}_t) Z_t(\mathcal{D}_{jt}, \mathcal{T}_t) - \gamma((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t) \\ \varepsilon_{ijt} &= Z_t(\mathcal{D}_{it}, \mathcal{T}_t) \varepsilon_{jt} + Z_t(\mathcal{D}_{jt}, \mathcal{T}_t) \varepsilon_{it} + \varepsilon_{it} \varepsilon_{jt}, \end{aligned} \quad (2.8)$$

where G_t is an element of a weakly stationary time series process (G_t) of random, square integrable functions with mean $E(G_t((\delta, \delta'), \tau)) = 0$ and covariance function

$$\rho((\delta, \delta'), (\delta'', \delta'''), \tau) = \text{Cov}(G_t((\delta, \delta'), \tau), G_t((\delta'', \delta'''), \tau))$$

for all $(\delta, \delta'), (\delta'', \delta''') \in [a(\tau), b(\tau)]^2$ and $\tau \in \mathcal{I}_\tau$. The corresponding covariance operator, say \mathcal{R} , is defined as

$$(\mathcal{R}g)((\delta, \delta'), \tau) = \int_{[a(\tau), b(\tau)]^2} \rho((\delta, \delta'), (\delta'', \delta'''), \tau) g((\delta'', \delta'''), \tau) d(\delta'', \delta'''), \quad (2.9)$$

for all functions $g(\cdot, \cdot, \tau) \in L_2[[a(\tau), b(\tau)]^2]$. Note that, it is necessary to delete the diagonal terms, for which $i = j$, since for these we have $E(\varepsilon_{iit} | \mathcal{D}_{it}, \mathcal{T}_t) = E(\varepsilon_{it} \varepsilon_{it}) = \sigma_\varepsilon^2 > 0$, which would introduce a systematic bias. This peculiarity is considered in the definition of the index sets in the right column of Table 2.1.

Common notation (see Eq. (2.10))	(i) Mean function μ (see Eq. (2.3))	(ii) Covariance function γ_u (see Eq. (2.7))
$Y_{it} \in \mathbb{R}$	$P_{it} \in \mathbb{R}$	$C_{ijt} \in \mathbb{R}$
$m(X_{it}, Q_t)$	$\mu(\mathcal{D}_{it}, \mathcal{T}_t)$	$\gamma((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t)$
$W_t(X_{it}, Q_t)$	$Z_t(\mathcal{D}_{it}, \mathcal{T}_t)$	$G_t((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t)$
ν_{it}	ϵ_{it}	ϵ_{ijt}
n	n	$n^2 - n\mathbb{I}(u = 0)$
$i \in \{1, \dots, n\}$	$i \in \{1, \dots, n\}$	$ij \in \begin{cases} \{\{1, \dots, n\}^2 \mid i \neq j\}, & u = 0 \\ \{1, \dots, n\}^2, & u > 0 \end{cases}$
$t \in \{1, \dots, T\}$	$t \in \{1, \dots, T\}$	$t \in \{1, \dots, T - u\}$

Table 2.1: Unifying notation for the nonparametric regression models in Eq.'s (2.3) and (2.7). In the lower part of the table we are slightly abusing the notation. There, the rows shall be understood as re-definitions, i.e., as $n^{new} := n^2 - n\mathbb{I}(u = 0)$ and $i^{new} := ij$, but without the usage of the confounding superscripts “new”; \mathbb{I} denotes the indicator function.

2.3.1 Unifying regression model

The notation for our nonparametric regression problems in Eq.'s (2.3) and (2.7) is rather cumbersome. A long-winded notation, might be useful for implementation and specific interpretations, but confounds any theoretical consideration. Fortunately, the regression problems in Eq.'s (2.3) and (2.7) can be seen as particular versions of an interesting common nonparametric regression problem, since their error components

$$\begin{aligned} Z_t(\mathcal{D}_{it}, \mathcal{T}_t) + \epsilon_{it} & \quad (\text{see Eq. (2.3)}) \text{ and} \\ G_t((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t) + \epsilon_{ijt} & \quad (\text{see Eq. (2.7)}) \end{aligned}$$

consist both of a functional part and a scalar part. Therefore, we define in Table 2.1 a unifying common notation, which shall act as a clarifying reference point.

This unifying notation allows us to write the nonparametric regression models of Eq.'s (2.3) and (2.7) more conveniently as

$$Y_{it} = m(X_{it}, Q_t) + \eta_{it}, \quad (2.10)$$

where the nonstandard error term $\eta_{it} = W_t(X_{it}, Q_t) + \nu_{it}$ has a random, mean

zero, square integrable functional part W_t and a random, mean zero, scalar part ν_{it} , both with finite variances. This composition of η_{it} makes it a heteroscedastic and autocorrelated error term, where the heteroscedasticity is a function of the prediction points $X_{it} \in \mathcal{C}_X(q)$ and $Q_t \in \mathcal{C}_Q$ with $\mathcal{C}_X(q) \subset \mathbb{R}^{d_1}$ and $\mathcal{C}_Q \subset \mathbb{R}^{d_2}$, where we define $d = d_1 + d_2$.

For our specific real data problem we have

$$\mathcal{C}_X(q) = \begin{cases} [a(q), b(q)] \subset \mathbb{R} & \text{if } m \triangleq \mu \\ [a(q), b(q)]^2 \subset \mathbb{R}^2 & \text{if } m \triangleq \gamma \end{cases}$$

and $\mathcal{C}_Q = \mathcal{I}_{\mathcal{T}} \subset \mathbb{R}$ in both cases.

As already mentioned above, the error term η_{it} actually bears two different types of autocovariances, which have to be considered in our theoretical work. On the one hand, there are within-function covariances, on the other hand, there are temporal between-function covariances. In the following we discuss these different types of autocovariances in more detail.

Let's define the conditional covariance operator of $W_t(\cdot, q)$, and the corresponding conditional covariance function as

$$(\Psi w)(x, q) = \int_{\mathcal{C}(q)} \psi((x, x'), q) w(x', q) dx' \quad (2.11)$$

$$\psi((x, x'), q) = \text{Cov}(W_t(x, q), W_t(x', q)) \quad (2.12)$$

with $x, x' \in \mathcal{C}_X(q)$, $q \in \mathcal{C}_Q$, $w(\cdot, q) \in L_2[\mathcal{C}_X(q)]$, and $W_t(\cdot, q) \in L_2[\mathcal{C}_X(q)]$. The covariance operator Ψ is the analogue to the specific covariance operators Γ and \mathcal{R} in Eq.'s (2.5) and (2.9).

Note that the covariance function $\psi((\cdot, \cdot), q)$ inherits its smoothness properties from the random functions $W(\cdot, q) \in L_2[\mathcal{C}_X(q)]$. In fact, it are these smoothness properties, which lead to so-called within-function covariances, since necessarily $\psi((x, x'), q) / (\sqrt{\psi((x, x), q)\psi((x', x'), q)}) \rightarrow 1$ as $\|x - x'\|_2 \rightarrow 0$.

The temporal between-function covariances are related to our assumption that $W_t(\cdot, q)$ comes from a time series of weakly stationary random functions. In order to formalize our assumption on the strength of the temporal dependencies, we also need to introduce the conditional autocovariance function, say ψ_u , which

describes the covariances between random functions $W_t(\cdot, q)$ and $W_{t+u}(\cdot, q')$ with $u \in \{1, 2, \dots\}$. We define

$$\psi_u((x, x'), (q, q')) = \text{Cov}(W_t(x, q), W_{t+u}(x', q')), \quad (2.13)$$

with $x \in \mathcal{C}_X(q)$, $x' \in \mathcal{C}_X(q')$, and $q, q' \in \mathcal{C}_Q$. Furthermore, we do not assume any specific time series model, but a short-range dependence, which we formalize using a geometrically bounded autocovariance function

$$|\psi_u((x, x'), (q, q'))| \leq C_\psi r^u, \quad (2.14)$$

where $0 < C_\psi < \infty$ and $0 < r < 1$. This assumption includes the important case of functional ARMA processes, so-called autoregressive Hilbertian processes as discussed, e.g., in Bosq (2000).

Interpretation of the nonparametric regression model (2.10) can be done, as usual, on a conditional basis, i.e., $m(X_{it}, Q_t) = E(Y_{it}|X_{it}, Q_t)$. Note, however, the different stochastic natures of the prediction points X_{it} and Q_t ; while the former has an hourly stochastic the latter has a daily stochastic. The consideration of these different stochastics leads to a (slightly) nonstandard theory for the nonparametric local linear estimator, since we cannot assume that the prediction points X_{it} and Q_t are drawn from a common multivariate distribution function.

Before we introduce the local linear estimator for the nonparametric function m , we want emphasize that the above discussed within-function covariances have nothing to with our time series context. They are typical for functional error terms and have to be considered also for the case of cross sectional functional data. To the best of our knowledge, this particular composition of the error term differs from situations studied so far in the literature on multivariate local linear regression.

2.3.2 The estimator

In order to introduce the local linear estimator, we define the following vectors and (partitioned) matrices:

	Dimensions:
$Y_t = (Y_{1t}, \dots, Y_{nt})^T$	$n \times 1$
$Y = (Y_1, \dots, Y_T)^T$	$Tn \times 1$
$X_t = (X_{1t}, \dots, X_{nt})^T$	$n \times d_1$
$X = (X_1, \dots, X_T)^T$	$Tn \times d_1$
$\vec{Q}_t = \iota_n \otimes Q_t$	$n \times d_2$
$Q = (\vec{Q}_1, \dots, \vec{Q}_T)^T$	$Tn \times d_2$
$(X, Q) = ((X_1, \vec{Q}_1), \dots, (X_T, \vec{Q}_T))^T$	$Tn \times d$

with $\iota_n \in \mathbb{R}^n$ being a column vector full of ones and “ \otimes ” denoting the Kronecker product. Furthermore, remember that $d = d_1 + d_2$ and that for $m \triangleq \mu$ we have $d_1 = d_2 = 1$ and for $m \triangleq \gamma$ we have $d_1 = 2$ and $d_2 = 1$.

Using these vector notations and our definitions for the unifying regression model in Eq. (2.10), a typical diagonal element of the $Tn \times Tn$ conditional covariance matrix $Cov(Y|(X, Q))$ is given by $\psi((X_{it}, X_{it}), Q_t) + \sigma_v^2(X_{it}, Q_t)$, while a typical off-diagonal element is given by $\psi_u((X_{it}, X_{jt+u}), (Q_t, Q_{t+u}))$.

We analyze the following multivariate nonparametric local linear estimator:

$$\hat{m}(x, q; H) = e_1^T ((X_x, Q_q)^T W_{xq} (X_x, Q_q))^{-1} (X_x, Q_q)^T W_{xq} Y \quad (2.15)$$

for all $x \in \mathcal{C}_X(q)$ and $q \in \mathcal{C}_Q$. The $(d+1) \times 1$ vector e_1 denotes the unit vector $e_1 = (1, 0, \dots, 0)^T$. The $Tn \times (d+1)$ matrix (X_x, Q_q) denotes the regressor matrix, centered around x , plus an additional column vector of ones, and is defined as $(X_x, Q_q) = (\iota_{Tn}, (X - \iota_{Tn} \otimes x), (Q - \iota_{Tn} \otimes q))$. The $Tn \times Tn$ matrix W_{xq} is a diagonal weighting matrix with typical diagonal elements $K_H((X_{it}, Q_t) - (x, q))$, where $K_H(u)$ is defined as $K_H(u) = |H|^{1/2} K(H^{1/2}u)$ with $u \in \mathbb{R}^d$ and K being a d -variate symmetric probability density function (pdf) having compact support.

For the sake of preciseness, we define K as a product kernel such that

$$K(u) = K_X((u_1, \dots, u_{d_1})) \cdot K_Q((u_{d_1+1}, \dots, u_d))$$

with $K_X((u_1, \dots, u_{d_1})) = \prod_{i=1}^{d_1} \kappa(u_i)$, $K_Q((u_{d_1+1}, \dots, u_d)) = \prod_{j=(d_1+1)}^d \kappa(u_j)$, and κ being a standard univariate kernel such as the Epanechnikov kernel. The bandwidth matrix H is a symmetric $d \times d$ block diagonal bandwidth matrix defined as

$$H = \begin{pmatrix} H_X & 0 \\ 0 & H_Q \end{pmatrix},$$

where the $d_1 \times d_1$ sub-matrix H_X has elements $[H_X]_{ij} = h_{X,ij}^2$ with $i, j \in \{1, \dots, d_1\}$ and the $d_2 \times d_2$ sub-matrix H_Q has elements $[H_Q]_{kl} = h_{Q,kl}^2$ with $k, l \in \{1, \dots, d_2\}$.

The estimator in Eq. (2.15) is simply a weighted least squares estimator. A detailed theoretical consideration of the classical version of this estimator can be found in Ruppert & Wand (1994).

2.3.3 Random design of the prediction points

Distributional assumptions on the prediction points have to be made with care, since they need to guarantee that the prediction points $(X_{it}, Q_t) \in \mathbb{R}^d$ become dense within the considered domain as the sample size T becomes large. Usually, this is assured by the simple assumption that the underlying d -variate probability density function (pdf) is bounded away from zero everywhere within the considered domain. Here, this is more involved due to the specific nature of the prediction points \mathcal{D}_{it} and \mathcal{T}_t . Below, we begin with the notationally simpler case $m \triangleq \mu$ for which the prediction points are given by $X_{it} = \mathcal{D}_{it}$ and $Q_t = \mathcal{T}_t$.

We model the phenomenon of clustered electricity demand values by the assumption that the conditional random variable (rv) $\mathcal{D}_{it}|A_t, B_t, \mathcal{T}_t = \tau$ has realizations within the random sub interval $[A_t, B_t] \subset [a(\tau), b(\tau)]$, where $a(\tau)$ and $b(\tau)$ are unobserved, deterministic, two times continuously differentiable functions with $0 < a(\tau) < b(\tau) < \infty$.

The random interval $[A_t, B_t]$ is used to describe the clustering of electricity

demand, where we make the simplifying assumption that $B_t = A_t + c(\tau)$ with $0 < c(\tau) < \infty$ being an unobserved, deterministic, two times continuously differentiable function. That is, the temperature dependent range of electricity demand $c(\tau) \in \mathbb{R}$ is randomly shifted by the conditional random variable $A_t | \mathcal{T}_t = \tau$ with pdf $f_A(\alpha | \mathcal{T}_t = \tau) > 0$ for all $\alpha \in [a(\tau), b(\tau) - c(\tau)]$ and zero otherwise. These random shifts describe the random electricity infeeds from RES; remember the definition of \mathcal{D}_{it} in Section 2.2.

A realistic assumption on the random variable of daily air temperature includes a smooth, time varying mean temperature component and allows for autocorrelation [see, e.g., Campbell & Diebold (2005)]. Correspondingly, we model \mathcal{T}_t by a time-point specific pdf $f_{\mathcal{T}}(\tau | t/T)$, which varies smoothly with t , and assume autocovariance-stationarity for the centered times series $(\mathcal{T}_t - E(\mathcal{T}_t))$ with a geometrically bounded autocovariance function $|Cov(\mathcal{T}_t, \mathcal{T}_{t+u})| \leq C_{\mathcal{T}} r^u$, where $0 < C_{\mathcal{T}} < \infty$, and $0 < r < 1$.

Altogether, we model the pdf of the conditional random variable of (residual) electricity demand $\mathcal{D}_{it} | A_t = \alpha, \mathcal{T}_t = \tau$ of a specific day t as

$$f_{\mathcal{D}}(d | A_t = \alpha, \mathcal{T}_t = \tau) > 0 \text{ for all } d \in [A_t, B_t] \text{ and zero otherwise,}$$

which describes the underlying density of the n clustered prediction points of a specific day t . Without conditioning on A_t , we have

$$f_{\mathcal{D}}(d | \mathcal{T}_t = \tau) = \int_{a(\tau)}^{b(\tau)-c(\tau)} f_{\mathcal{D}}(d | A_t = \alpha, \mathcal{T}_t = \tau) f_A(\alpha | \mathcal{T}_t = \tau) d\alpha$$

with $f_{\mathcal{D}}(d | \mathcal{T}_t = \tau) > 0$ for all $d \in [a(\tau), b(\tau)]$ and zero otherwise.

The distributional assumptions for the case $m \triangleq \gamma$, with prediction points $(X_{it}, Q_t) = ((\mathcal{D}_{it}, \mathcal{D}_{jt}), \mathcal{T}_t)$, follow directly from our above discussions by generalizing the univariate random variable \mathcal{D}_{it} to the bivariate random variable: $\mathcal{D}_{it} := (\mathcal{D}_{it}, \mathcal{D}_{jt})$. For our theoretical considerations, we will use the unifying notation $f_X(x | Q_t = q)$ in order to denote the d_1 variate conditional pdf of the random prediction points $X_{it} | Q_t = q$. Furthermore, $f_Q(q | t/T)$ shall denote the d_2 variate conditional pdf of Q_t given a specific day t .

For our specific real data problem we have

$$f_X(x|Q_t = q) = \begin{cases} f_{\mathcal{D}}(d|\mathcal{T}_t = \tau) & \text{if } m \triangleq \mu \\ f_{\mathcal{D}}(\mathbf{d}|\mathcal{T}_t = \tau) & \text{if } m \triangleq \gamma \end{cases}$$

and $f_Q(q|t/T) = f_{\mathcal{T}}(\tau|t/T)$ in both cases.

2.3.4 Boundary estimators

An interesting peculiarity of our empirical study are the random and deterministic intervals of the random variable electricity demand $\mathcal{D}_{it} \in [A_t, B_t] \subset [a(\tau), b(\tau)]$. The estimation of the interval bounds is related with active research topics such as the estimation of frontiers and density boundaries. The simplest estimators are extremum estimators, for example, A_t and B_t can be estimated consistently by $A_t^{\min} = \min\{\mathcal{D}_{t1}, \dots, \mathcal{D}_{tn}\}$ $B_t^{\max} = \max\{\mathcal{D}_{t1}, \dots, \mathcal{D}_{tn}\}$ as long as $n = n(T) \rightarrow \infty$ with $T \rightarrow \infty$; see Gijbels & Peng (2000) for a consideration of similar estimators.

For the estimation of the bounds $a(\tau)$ and $b(\tau)$ we take advantage of our assumption that $a(\tau)$ and $b(\tau)$ are smooth, two times continuously differentiable, bounds. This allows us to estimate these bounds consistently as $T \rightarrow \infty$ even if $n(T)$ remains bounded. However, the scatter plot in Figure 2.2 suggests a nonstandard assumption on the shape of $a(\tau)$ and $b(\tau)$, which excludes the usage of classical boundary estimators such as free disposal hull (FDH) estimators or data envelope estimators [see, e.g., Deprins *et al.* (1984) and Kneip *et al.* (1998)].

Instead, we use nonparametric local linear regression in order to estimate the bounds $a(\tau)$ and $b(\tau)$. On the one hand, this allows us to estimate arbitrary smooth boundary functions; on the other hand, it seamlessly fits to our unifying nonparametric regression problem in Eq. (2.15). We use the deterministic frontier regression model proposed by Martins-Filho & Yao (2007), which can be formulated for our case as

$$A_t^{\min} = a(\mathcal{T}_t) R_t^a \quad \text{and} \quad B_t^{\max} = b(\mathcal{T}_t) R_t^b, \quad (2.16)$$

where the multiplicative error components R_t^a and R_t^b are assumed to have realizations $R_t^a \in [1, \infty)$ and $R_t^b \in [0, 1]$. If $R_t^a = 1$ and $R_t^b = 1$, then the corresponding

observations A_s^{\min} and B_t^{\max} lie exactly on the boundary functions with A_s^{\min} and B_t^{\max} being the above introduced extremum estimators.

The first two moments of the multiplicative error terms R_t^a and R_t^b are assumed to exist and to be independent of \mathcal{T}_t such that $E(R_t^a|\mathcal{T}_t = \tau) = \mu_{aR}$, $E(R_t^b|\mathcal{T}_t = \tau) = \mu_{bR}$, $Var(R_t^a|\mathcal{T}_t = \tau) = \sigma_{aR}^2$, and $Var(R_t^b|\mathcal{T}_t = \tau) = \sigma_{bR}^2$. These assumptions are needed to ensure identifiability of the boundary functions. In contrast to the assumptions of Martins-Filho & Yao (2007), we also allow for autocorrelated errors with geometrically decreasing bounds on the autocovariances $|Cov(R_t^a, R_{t+u}^a)| \leq C_a r^u$ and $|Cov(R_t^b, R_{t+u}^b)| \leq C_b r^u$, where $0 < C_a, C_b < \infty$, and $0 < r < 1$.

Martins-Filho & Yao (2007) propose a three-step estimation procedure. The first two steps consist of two separate nonparametric local linear regressions, which aim to estimate the shape of the frontier functions. The third estimation step is done in order to estimate the location of the frontier functions. Particularly, the second regression problem is prone to outlier problems, since it involves smoothing the squared residuals from the first regression. In contrast, we propose a simpler more robust, two step estimation procedure, which avoids the need to smooth squared residuals.

First, we estimate the shapes, say $m_a(\tau)$ and $m_b(\tau)$, of the log-transformed boundary functions $\log[a(\tau)]$ and $\log[b(\tau)]$ using the following two nonparametric regression problems

$$\log(A_t^{\min}) = m_a(\mathcal{T}_t) + \eta_t^a \quad \text{and} \quad \log(B_t^{\max}) = m_b(\mathcal{T}_t) + \eta_t^b, \quad (2.17)$$

where $m_a(\mathcal{T}_t) = [\log(a(\mathcal{T}_t)) + E(\log(R_t^a))]$, $\eta_t^a = [\log(R_t^a) - E(\log(R_t^a))]$, $m_b(\mathcal{T}_t) = [\log(a(\mathcal{T}_t)) + E(\log(R_t^b))]$, and $\eta_t^b = [\log(R_t^b) - E(\log(R_t^b))]$ with $E(\eta_t^a|\mathcal{T}_t = \tau) = 0$, $E(\eta_t^b|\mathcal{T}_t = \tau) = 0$, $\eta_t^a \in [0, \infty)$, and $\eta_t^b \in (-\infty, 0]$.

The final second step concerns the estimation of $c_a = E(\log(R_t^a))$ and $c_b = E(\log(R_t^b))$, which determines the locations of the boundary functions; note that by construction $c_a \geq 0$ and $c_b \leq 0$. In this step, we follow Martins-Filho &

Yao (2007) and estimate c_a and c_b using the following extremum estimators

$$\begin{aligned}\hat{c}_a &= (-1) \cdot \min_{t \in \{1, \dots, T\}} \left(\log \left(A_t^{\min} \right) - \hat{m}_a(\mathcal{T}_t) \right) \\ \hat{c}_b &= (-1) \cdot \max_{s \in \{1, \dots, T\}} \left(\log \left(B_s^{\min} \right) - \hat{m}_b(\mathcal{T}_s) \right),\end{aligned}$$

where $|\hat{c}_a|$ denotes the absolute value of the largest residual of all observations below the regression function \hat{m}_a ; vice versa for $|\hat{c}_b|$. This is equivalent to a normalization such that $R_s^a = 1$ and $R_t^b = 1$ for those observations s and t , which fulfill the above maximization and minimization problems. The estimates of the boundary functions are then defined as $\hat{a}(\tau) = \exp(\hat{m}_a(\tau) - \hat{c}_a)$ and $\hat{b}(\tau) = \exp(\hat{m}_b(\tau) - \hat{c}_b)$.

The models in Eq. (2.17) are standard nonparametric regression problems and, in particular, are special cases of our unifying nonparametric regression model (2.10) with $m \triangleq m_a$, $m \triangleq m_b$ and error terms, $\eta_{it} \triangleq \eta_t^a$ and $\eta_{it} \triangleq \eta_t^b$. Our theoretical considerations of the estimator (2.15) include this case as a special case; see also our discussion of Theorem 2.4.1 in Section 2.4.

2.4 Technical assumptions & main results

In this section we present our main theoretical results. The asymptotic conditional bias and variance of the estimator $\hat{m}(x; H)$ are given in Theorem 2.4.1. Consistency of the estimated eigencomponents is stated in Theorem 2.4.2.

For simplicity, we consider only interior points $x \in \{z \in \mathcal{C}_X(q) | B_{z, H_X} \subset \mathcal{C}_X(q)\}$ and interior points $q \in \{r \in \mathcal{C}_Q | B_{r, H_Q} \subset \mathcal{C}_Q\}$, where B_{z, H_X} is the H_X -ball centered around z , defined as $B_{z, H_X} = \{x \in \mathbb{R}^{d_1} : H_X^{-1/2}(x - z) \in \text{supp}(K_X)\}$ with $\text{supp}(K_X) = [-1, 1]^{d_1}$; correspondingly for B_{r, H_Q} . The consideration of boundary points does not add any new insights as it is straightforward to generalize our results for the case of boundary points using the theory of Ruppert & Wand (1994).

Our technical assumptions are in line with the standard assumptions in the literature on nonparametric regression and functional principal component analysis. The following list contains assumptions that are adopted from the papers of Ruppert & Wand (1994) and Kneip & Utikal (2001):

- (A0) The dependent random variable Y_{it} comes from the data generating process as defined in Eq. (2.10). The random design of the prediction points X_{it} and Q_t is as discussed in Section 2.3.3.
- (A1) The kernel function K is assumed to be a compactly supported, symmetric, d -variate density function, in order to assure that $\int_{[-1,1]^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du = 0$ for all non negative integers l_1, \dots, l_d such that their sum is odd. Furthermore, it is assumed that $\int_{[-1,1]^d} uu' K(u) du = \mu_2(K) I_d$ with $\mu_2(K) \neq 0$ and I_d being the $d \times d$ dimensional identity matrix, $R(K) = \int_{[-1,1]^d} K(u)^2 du$, and $R(K_Q) = \int_{[-1,1]^{d_2}} K_Q(w)^2 dw$.
- (A2) The pdf $f_Q(q|s)$ is continuously differentiable at any point $q \in \mathcal{C}_Q$ and any point s with $0 \leq s \leq 1$. Furthermore, $f_Q(q, s) > 0$ for all $q \in \mathcal{C}_Q$ and $0 \leq s \leq 1$. The covariance functions $\psi((x, x'), q)$ and $\psi_u((x, x''), (q, q''))$ are continuously differentiable for all $u \in \{1, 2, \dots\}$, all $x, x' \in \mathcal{C}_X(q)$, all $x'' \in \mathcal{C}_X(q'')$, and all $q, q'' \in \mathcal{C}_Q$. Furthermore, all second-order derivatives of m are continuous.
- (A3) The sequence of bandwidth matrices $H_X = H_{X, Tn}$ is such that $(Tn)^{-1} |H_X|^{-1/2}$ and all elements of H_X tend to zero as $Tn \rightarrow \infty$, where H_X remains symmetric and positive definite. Correspondingly, the sequence of bandwidth matrices $H_Q = H_{Q, T}$ is such that $T^{-1} |H_Q|^{-1/2}$ and all elements of H_Q tend to zero as $T \rightarrow \infty$, where H_Q remains symmetric and positive definite. Furthermore, the ratios of the largest and smallest eigenvalues of H_X and H_Q are assumed to be bounded for all Tn and all T . The sequence of $n = n(T) \geq 2$ is assumed to be $n(T) \sim T^\beta$ with $0 \leq \beta < \infty$. (We write “ $a_n \sim b_n$ ” in order to denote that a_n and b_n are asymptotically equivalent up to some positive constant $0 < c < \infty$, i.e., that $\lim_{n \rightarrow \infty} (a_n/b_n) = c$.)
- (B1) The k th eigenvalue $\lambda_k(\tau)$ of the matrix $M(\tau)$ shall be such that $\min_{s=1, \dots, n; s \neq r} |\lambda_k(\tau) - \lambda_s(\tau)| \geq n_e C_{3,r}$, where the $n_e \times n_e$ dimensional matrix $M(\tau)$ is defined below in Definition 2.4.1. This assumption is an adopted version of assumption A2 in Kneip & Utikal (2001).

The assumption that the number of discretization points n is a sequence $n = n(T) \sim T^\beta$ (Assumption A3) is actually not necessary for our asymptotic results;

there it suffices to assume that $n \geq 2$. It is just an auxiliary assumption, which allows us to interpret the influence of different scenarios for the sequence $n(T)$ in a simple manner.

An integral part of our empirical study is the comparison of the electricity market situations one year before and one year after Germany's nuclear moratorium on March 14, 2011. The most obvious advantage of such a year-wise consideration is that the estimation results can be easily contrasted. A further advantage, which is particularly important in our case, is that year wise considerations yield to year-wise averages of all time-varying variables that are not explicitly considered by our model (2.1), such as the steady nonlinear integration of renewable energy sources and the expansion of the grid infrastructure, etc. The explicit consideration of these variables is practically impossible due to the course of dimensionality in nonparametric regression problems.

Correspondingly, we are considering a year-wise fill-in asymptotic as $T \rightarrow \infty$ with T being the number of price auctions per year. Our theoretical results can be interpreted for the case in which the number of traded spot prices within each auction $n = n(T)$ remains bounded as well as when $n(T)$ diverges with T .

Of course, the given auction design at the EEX, does not allow for a simple increase of the sample size, but this is a not an unusual empirical problem. Nevertheless, in principle it is possible to organize electricity markets with more than one auction per day. For example, at the Australian electricity market $n = 6$ five-minute-wise spot prices are settled simultaneously within half-hourly auctions, such that $T = 365 \cdot 48$ [Wolak (2000)].

Theorem 2.4.1. *Conditional asymptotic bias and variance of the d -variate local linear estimator $\hat{m}(x, q; H)$, as $T \rightarrow \infty$ and $n = n(T) \geq 2$, with x and q being interior points of $\mathcal{C}_X(q)$ and \mathcal{C}_Q .*

(i) *Conditional bias:*

$$\begin{aligned} E(\hat{m}(x, q; H) - m(x) | (X, Q)) &= \frac{1}{2} \mu_2(K) \operatorname{tr} \{ H \mathcal{H}_m(x, q) \} + o_p(\operatorname{tr}(H)) \\ &= \mathcal{O}_p(\operatorname{tr}(H)), \end{aligned}$$

where $\mathcal{H}_m(x, q)$ is the $d \times d$ Hessian matrix of the regression function m at

the point (x, q) .

(ii) *Conditional variance:*

$$\begin{aligned}
\text{Var}(\hat{m}(x, q; H)|(X, Q)) &= \\
&= (Tn)^{-1}|H|^{-1/2} \left\{ \frac{R(K) (\psi((x, x), q) + \sigma_v^2(x, q))}{f_{XQ,T}(x, q)} \right\} (1 + \mathcal{O}_p(\text{tr}(H^{1/2}))) \\
&\quad + T^{-1}C_f(x, q) \left[\left(\frac{n-1}{n} \right) |H_Q|^{-1/2} R(K_Q)\psi((x, x), q) + C_r(x, q) \right] \\
&\quad + \mathcal{O}_p(T^{-1} \text{tr}(H^{1/2})) \\
&= \mathcal{O}_p((Tn)^{-1}|H|^{-1/2} + T^{-1}|H_Q|^{-1/2}),
\end{aligned}$$

where $C_f(x, q)$ is a positive constant $C_f(x, q) < 1$, defined as

$$C_f(x, q) = \frac{f_{XQ,T}^2(x, q)}{(f_{XQ,T}(x, q))^2} = \frac{T^{-1} \sum_{t=1}^T (f_X(x|q)f_Q(q|t/T))^2}{(T^{-1} \sum_{t=1}^T f_X(x|q)f_Q(q|t/T))^2},$$

and $C_r(x, q)$ is a non-negative constant $C_r(x, q) \leq \frac{2C_{\psi^r}}{1-r} < \infty$, defined as $C_r(x, q) = 2 \sum_{u=1}^{T-1} \psi_u((x, x), (q, q))$.

Discussion

1. Part (i) of Theorem 2.4.1 corresponds to the classical asymptotic conditional bias result of Ruppert & Wand (1994). Also, the first summand in the variance part (ii) of Theorem 2.4.1, which considers the variances $\text{Var}(\eta_{it}|(X, Q)) = \psi_0((x, x), q) + \sigma_v^2(x, q)$, is in line with the results in Ruppert & Wand (1994). However, the second and third summands, which consider the covariances $\text{Cov}(\eta_{it}, \eta_{js}|(X, Q))$, with $i \neq j$ and $t \neq s$, differs from the results in Ruppert & Wand (1994) and comes from the particular composition of the error term $\eta_{it} = W_t(X_{it}, Q_t) + \nu_{it}$ of the unifying regression model in Eq. (2.10).

The pure cross sectional case is included in Theorem 2.4.1 by the limiting value of $C_r(x, q) \rightarrow 0$ as $r \rightarrow 0$, where r quantifies the speed at which the autocovariances decay as $u = |t - s| \rightarrow \infty$.

The case of sparse functional data as in Yao *et al.* (2005), where n is assumed to be an independent discrete random variable, say N , with $E(N) < \infty$ and realizations in $\{2, 3, \dots\}$, is also a special case of our asymptotic analysis. Here, the quantities “ n ” in part (ii) of Theorem 2.4.1 simply have to be replaced by $E(N)$.

In the case that $m \triangleq \mu$, Theorem 2.4.1 is also valid if $n \geq 1$. But if $m \triangleq \gamma$ we need $n \geq 2$, otherwise it is impossible to compute the raw covariances in Eq. (2.6).

2. Our particular assumptions on the random design of the prediction points in Section 2.3.3 lead to a further non-standard variance component in part (ii) of Theorem 2.4.1: The scaling factor $0 < C_f(x, q) < 1$ would be equal to one in the case of classical assumptions on the random design of the prediction points.
3. The additional effect of having to estimate the mean function μ before being able to estimate the covariance function γ is negligible under our smoothness assumptions (Assumption A2), where we assume the existence of at least two derivatives for μ as well as γ . We do not discuss this issue in detail, since Wang *et al.* (2008a) already discuss this exhaustively under similar conditions.
4. The variance components $\psi((x, x'), q)$ and $\sigma_\nu^2(x, q)$ of part (ii) of Theorem 2.4.1 have different meanings depending on whether $m \triangleq \mu$ or $m \triangleq \gamma$. In order to facilitate interpretation, we provide Proposition 2.4.1.

Proposition 2.4.1. *Case wise definitions of $\psi((x, x'), q)$ and $\sigma_\nu^2(x, q)$.*

(i) *Case $m \triangleq \mu$:*

$$\begin{aligned} \psi((x, x'), q) &\triangleq \gamma((\delta, \delta'), \tau) = \text{Cov}(Z_t(\delta, \tau), Z_t(\delta', \tau)) \\ \sigma_\nu^2(x, q) &\triangleq \sigma_\epsilon^2 = \text{Var}(\epsilon_{it}), \end{aligned}$$

where $\delta, \delta' \in [a(\tau), b(\tau)]$ and $\tau \in \mathcal{I}_T$.

(ii) *Case $m \triangleq \gamma$:*

$$\begin{aligned}
\psi((x, x'), q) &\triangleq \rho((\delta, \delta'), (\delta'', \delta'''), \tau) = \text{Cov}(G_t((\delta, \delta'), \tau), G_t((\delta'', \delta'''), \tau)) \\
\sigma_\nu^2(x, q) &\triangleq \sigma_\varepsilon^2((\delta, \delta'), \tau) = \text{Var}(\varepsilon_{ijt} | (\mathcal{D}_{it}, \mathcal{D}_{jt}, \mathcal{T}_t) = (\delta, \delta', \tau)) \\
&= \gamma((\delta, \delta), \tau) + \gamma((\delta', \delta'), \tau) + 4\sigma_\varepsilon^2,
\end{aligned}$$

where $\delta, \delta', \delta'', \delta''' \in [a(\tau), b(\tau)]$ and $\tau \in \mathcal{I}_\tau$.

2.4.1 Optimal bandwidth selection

In the following, we optimize the bandwidth matrix with respect to the classical mean integrated squared error (MISE) loss function. Here and in what follows, we focus on the simpler problem of choosing one global bandwidth variable h , with $H_X = hI_{d_1}$ and $H_Q = hI_{d_2}$, since the use of a single global bandwidth variable yields satisfactory and stable results in our empirical analysis. In this case the MISE of $\hat{m}(x, q; h)$ is given by

$$\begin{aligned}
&\text{MISE}\{\hat{m}(x, q; h) | (X, Q)\} = \tag{2.18} \\
&= E \left[\int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (\hat{m}(x, q; h) - m(x, q))^2 f_{XQ,T}(x, q) dx dq | (X, Q) \right] \\
&= (Tn)^{-1} h^{-d} R(K) \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (\psi((x, x), q) + \sigma_\nu^2(x, q)) dx dq \\
&\quad + T^{-1} h^{-d_2} R(K_Q) \left(\frac{n-1}{n} \right) \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (C_f(x, q) \psi((x, x), q) f_{XQ,T}(x, q)) dx dq \\
&\quad + \frac{1}{4} h^4 (\mu_2(K))^2 \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (\text{tr} \{ \mathcal{H}_m(x, q) \})^2 f_{XQ,T}(x, q) dx dq \\
&\quad + o_p((Tn)^{-1} h^{-d} + T^{-1} h^{-d_2} + h^4),
\end{aligned}$$

where $(\text{tr} \{ \mathcal{H}_m(x, q) \})^2 = \left(\sum_{i=1}^{d_1} \frac{\partial^2 m}{\partial x_i \partial x_i}(x, q) + \sum_{j=d_1+1}^d \frac{\partial^2 m}{\partial q_j \partial q_j}(x, q) \right)^2$.

For the further minimization of the MISE with respect to h we face the problem that it remains unclear, which of the two variance rates $\mathcal{O}_p((Tn)^{-1} h^{-d})$ and $\mathcal{O}_p(T^{-1} h^{-d_2})$ dominates the other; except, of course, for the case of n being a constant, then $\mathcal{O}_p(T^{-1} h^{-d_2}) = o_p((Tn)^{-1} h^{-d})$, since $d > d_2$. In the following we differentiate between the two possible cases:

- If $\mathcal{O}_p(T^{-1} h^{-d_2}) = o_p((Tn)^{-1} h^{-d})$, e.g., if $n = n(T)$ remains constant as $T \rightarrow \infty$.

$$h_{\text{MISE}} = \left[\frac{dR(K) \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (\psi((x, x), q) + \sigma_v^2(x, q)) dx dq}{(\mu_2(K))^2 \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (\text{tr} \{\mathcal{H}_m(x, q)\})^2 f_{XQ,T}(x, q) dx dq Tn} \right]^{1/(d+4)}, \quad (2.19)$$

which corresponds to the usual optimal bandwidth result for d -variate nonparametric regression problems.

- If $\mathcal{O}_p((Tn)^{-1}h^{-d}) = o_p(T^{-1}h^{-d_2})$, i.e., if $n = n(T)$ diverges sufficiently fast as $T \rightarrow \infty$.

$$h_{\text{MISE}} = \left[\frac{d_2 R(K_Q) \binom{n-1}{n} \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (C_f(x, q) \psi((x, x), q) f_{XQ,T}(x, q)) dx dq}{(\mu_2(K))^2 \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(q)} (\text{tr} \{\mathcal{H}_m(x, q)\})^2 f_{XQ,T}(x, q) dx dq T} \right]^{1/(d_2+4)}, \quad (2.20)$$

which is surprising, since this rate is the optimal bandwidth result for a d_2 -variate nonparametric regression problem with $d > d_2$.

These MISE-optimal bandwidth results are still not exhaustive, but direct to further research.

Generally, an optimal bandwidth for estimating a nonparametric covariance function usually is not the optimal bandwidth for estimating its eigencomponents. In fact, Theorem 2.4.2 below suggests to choose a bandwidth $h \sim (Tn)^{-\delta}$ with $(1/4)d \leq \delta < 1$.

2.4.2 Estimation of the eigencomponents

For our asymptotic analysis of the estimated eigenvalues $\hat{\lambda}_k(\tau)$ and eigenfunctions $\hat{p}_k(\cdot, \tau)$, we take the point of view of a *discretized* covariance functions $\gamma(\cdot, \cdot, \tau)$. This is a relevant point of view, since any practical implementation of FPCA actually is based on discretized versions of the covariance functions. We define the following discretized, conditional covariance functions:

Definition 2.4.1. Let $\hat{M}(\tau)$ and $M(\tau)$ be $n_e \times n_e$ discretization matrices of the covariance functions $\hat{\gamma}(\cdot, \cdot, \tau)$ and $\gamma(\cdot, \cdot, \tau)$ such that

$\hat{M}(\tau)$ has typical elements $\hat{M}_{ij} = \hat{\gamma}(e_i, e_j, \tau; h) \in \mathbb{R}$ and

$M(\tau)$ has typical elements $M_{ij} = \gamma(e_i, e_j, \tau) \in \mathbb{R}$,

where e_i and e_j with $i, j \in \{1, \dots, n_e\}$ are evaluation points suitably chosen such that they build, e.g., a regular grid within the domain $[a(\tau), b(\tau)]^2$. The number of discretization points n_e is assumed to be constant such that $0 < n_e < \infty$.

The corresponding empirical eigenvalues and eigenvectors will be denoted by $\hat{\lambda}_k(\tau) \in \mathbb{R}$ and $\hat{\mathbf{p}}_k(\tau) \in \mathbb{R}^{n_e}$. Furthermore, we regard the more simple case of a global bandwidth as already considered above.

Theorem 2.4.2. *In-probability convergence rates of the empirical eigenvalues $\hat{\lambda}_k(\tau)$ and eigenvectors $\hat{\mathbf{p}}_k(\tau)$.*

Let $n \geq 2$ and $|H|^{1/2} = h^d \sim (Tn)^{-\delta}$ with $\frac{1}{4}d \leq \delta < 1$. Then

(i) *Eigenvalues:*

$$\begin{aligned} \hat{\lambda}_k(\tau) - \lambda_k(\tau) &= \mathbf{p}_k(\tau)^T [\hat{M}(\tau) - M(\tau)] \mathbf{p}_k(\tau) + \mathcal{O}_p((Tn)^{-1}) \\ &= \mathcal{O}_p((Tn)^{-1/2}) \end{aligned}$$

(ii) *Eigenvectors:*

$$\|\hat{\mathbf{p}}_k(\tau) - \mathbf{p}_k(\tau)\|_2 = \mathcal{O}_p((Tn)^{-1/2}),$$

where $\|\cdot\|_2$ denotes the Euclidean norm on \mathbb{R}^{n_e} .

2.4.3 Estimation of the principal component scores

In this section we discuss the problem of finding a K dimensional approximation for the latent price functions \mathcal{P}_t . With our estimates $\hat{\mu}$ and \hat{p}_k at hand, this approximation problem reduces to a parametric estimation problem of the temperature dependent principal component scores $\beta_{tk}(\mathcal{T}_t)$. We point to some estimation problems induced by the consideration of the non standard random design of electricity demand \mathcal{D}_{it} . The arguments within this section lack of mathematical rigor and have to be understood as a first general consideration of the problem.

Given our assumptions on model (2.3) the best K dimensional approximation is given by

$$\mathcal{P}_t(d, \mathcal{T}_t) \approx \mu(d, \mathcal{T}_t) + \sum_{k=1}^K \beta_{tk}(\mathcal{T}_t) p_k(d, \mathcal{T}_t), \quad (2.21)$$

with $d \in [a(\tau), b(\tau)]$, where the approximation error goes against zero as $K \rightarrow \infty$. Due to the well-known best basis property of the eigenfunctions, usually a low number of, e.g., $K = 3$ is sufficient to yield satisfactory approximations in practical problems [see Section 8.2.3 in Ramsay & Silverman (2005)].

Note that the approximation problem in Eq. (2.21) has to be read conditionally for a given (unobserved) realization of the random function \mathcal{P}_t ; consequently, the score parameters $\beta_{tk}(\mathcal{T}_t)$ are fixed values. The only stochastic source, which we have to consider in the estimation of the parameters $\beta_{tk}(\mathcal{T}_t)$ comes from the error term ϵ_{it} of Eq. (2.3), since we observe the noisy discretization points P_{it} instead of $\mathcal{P}_t(d, \mathcal{T}_t)$. A prerequisite for a consistent estimation of the score parameter $\beta_{tk}(\mathcal{T}_t)$ is that the number n of discretization points per function is large, i.e., we henceforth assume that $n = n(T) \rightarrow \infty$ as $T \rightarrow \infty$.

The classical way to approach the approximation problem in Eq. (2.21) is to plug in the corresponding estimates $\hat{\mu}$ and \hat{p}_k and to approximate the integrals $\beta_{tk}(\mathcal{T}_t) = \int (\mathcal{P}_t(x, \mathcal{T}_t) - \mu(x, \mathcal{T}_t)) p_k(x, \mathcal{T}_t) dx$ by $\sum_{i=1}^n (P_{it} - \hat{\mu}(\mathcal{D}_{it}, \mathcal{T}_t)) \hat{p}_k(\mathcal{D}_{it}, \mathcal{T}_t) (\mathcal{D}_{i-1t} - \mathcal{D}_{it})$ with $\mathcal{D}_{0t} = \hat{a}(\mathcal{T}_t)$. However, this procedure is inappropriate in our case as we face the problem that the random variable \mathcal{D}_{it} has only realizations within the subinterval $[A_t, B_t] \subseteq [a(\mathcal{T}_t), b(\mathcal{T}_t)]$. The problem is that the eigenfunctions are not necessarily orthogonal to each other when considered only on the subintervals $[A_t, B_t]$.

The ordinary least squares estimator accounts for non-orthogonal basis functions (regressors); therefore, it is convenient to state the approximation problem in Eq. (2.21) as a regression problem

$$(P_{it} - \mu(\mathcal{D}_{it}, \mathcal{T}_t)) = \sum_{k=1}^K \beta_{tk}(\mathcal{T}_t) p_k(\mathcal{D}_{it}, \mathcal{T}_t) + \epsilon_{it}^*, \quad (2.22)$$

where ϵ_{it}^* is an augmented error term defined as $\epsilon_{it}^* = \sum_{l=K+1}^{\infty} \beta_{tl}(\mathcal{T}_t) p_l(\mathcal{D}_{it}, \mathcal{T}_t) + \epsilon_{it}$

with mean value $E(\epsilon_{it}^*) = \sum_{l=K+1}^{\infty} \beta_{il}(\mathcal{T}_t) p_l(\mathcal{D}_{it}, \mathcal{T}_t) \neq 0$. After plugging in the estimates $\hat{\mu}$ and \hat{p}_k we can try to estimate the score parameter $\beta_{tk}(\mathcal{T}_t)$ using the corresponding ordinary least squares estimator, say $\hat{\beta}_{tk}(\mathcal{T}_t)$.

Obviously, the estimator $\hat{\beta}_{tk}(\mathcal{T}_t)$ suffers from an omitted variable bias. Whether the order of magnitude of this bias may be negligible or not depends on how fast the sequence of the β_{il} 's converges against zero, and of course on the choice of K . Indeed, very often the first three eigenfunctions allow for very accurate approximations; in such cases a choice of, e.g., $K = 5$ shall lead to a reasonable low omitted variables bias.

Additionally, the restriction that \mathcal{D}_{it} has only realizations within the subinterval $[A_t, B_t]$ can induce a multicollinearity problem. As the price functions are rather smooth objects, the first eigenfunctions will be rather smooth, too. This has the unpleasant effect that the first eigenfunctions considered within the subintervals $[A_t, B_t]$ resemble simple linear functions, which are likely to have an approximate linear relationship with each other. This is particularly true if the subintervals $[A_t, B_t]$ are relatively small in comparison to the intervals $[a(\mathcal{T}_t), b(\mathcal{T}_t)]$. The model fit of the regression problem in Eq. (2.22) within the subintervals $[A_t, B_t]$ is not affected by this multicollinearity problem. But it is a problematic issue, if prediction of the parts outside of the subinterval is of interest.

Experiences with our data set show that there the multicollinearity problem is a much worse problem than the omitted variable bias; see also Figure 2.6. However, there may be possibilities so improve the imprecise estimation of the score parameters by using the knowledge that price functions are monotonically increasing, since this imposes additional restriction on the parameters.

Yao *et al.* (2005) consider the case of truly sparse functional data with only a few discretization points per function and propose to estimate the principal component scores by their conditional expectation given the observed discretization points. This approach works well as long as the discretization points are scattered over the whole domain, but leads to unsatisfactory results in our case.

2.5 Application

In this section we apply our previously discussed estimation procedure to our data set introduced above in Section 2.2. We use a slightly modified version of the plug-in estimator of h_{MISE} suggested by Ruppert *et al.* (1995) in order to approximate the unknown quantities in Eq. (2.19). For the estimation of the eigencomponents, we simply use the same plug-in bandwidth.

The usefulness of our model is demonstrated by contrasting the estimation results for the time span one year before and one year after Germany's nuclear moratorium on March 14, 2011. In order to differentiate between the estimated quantities for the two different time spans we henceforth use self-explanatory superscripts $t.\text{sp} \in \{\text{bef}, \text{aft}\}$ such as, e.g., $\hat{\mu}^{\text{bef}}$ and $\hat{\mu}^{\text{aft}}$.

Germany's nuclear moratorium on March 14, 2011 describes a natural experiment, in which eight nuclear power plants were permanently phased-out [Nestle (2012)]. As a price reaction to this treatment, the micro economic merit order model predicts a horizontal left shift of the monotonically increasing merit order curves. Given that demand for electricity is nearly inelastic, this shift of the merit order curve should be reflected by the price functions.

In fact, the estimated mean function $\hat{\mu}^{\text{bef}}(d, \tau; h)$ with $(d, \tau) \in \mathcal{S}^{\text{bef}}$ is throughout beneath of the estimated mean function $\hat{\mu}^{\text{aft}}(d, \tau; h)$ with $(d, \tau) \in \mathcal{S}^{\text{aft}}$. Compare to this Figure 2.3, where the graphs of $\hat{\mu}^{\text{bef}}$ and $\hat{\mu}^{\text{aft}}$ are plotted as a red and green shaded contour plots. Note that the plots refer to peak price hours only. The two corresponding estimates of the mean functions, which refer to off-peak price hours, are not separately plotted, due to reasons of space.

This increase in the mean price function only potentially leads as well to higher electricity spot prices. Ultimately, it depends on the level of electricity demand whether the mean electricity spot prices increase or not. A first rough impression of the differences in (residual) electricity demand before and after the moratorium can be gained by considering the estimated lower and upper boundaries $[\hat{a}^{\text{bef}}(\tau; h), \hat{b}^{\text{bef}}(\tau; h)]$ and $[\hat{a}^{\text{aft}}(\tau; h), \hat{b}^{\text{aft}}(\tau; h)]$, which define the boundaries of the red and green shaded regions in Figure 2.3. Specifically, during the cold winter days the range of (residual) electricity demand after the moratorium is considerably lower than before the moratorium.

This shift has to be explained mainly by the reduction of gross electricity demand; remember the definition of (residual) electricity demand in Section 2.2. The ratio between mean gross electricity demand one year before and one year after the moratorium is about $\bar{\mathcal{G}}^{\text{aft}} \approx 0.9 \bar{\mathcal{G}}^{\text{bef}}$. The corresponding ratio conditional on days with temperatures below zero degree Celsius is about $\bar{\mathcal{G}}^{\text{aft}} | \mathcal{T}_t < 0 \approx 0.8 \bar{\mathcal{G}}^{\text{bef}} | \mathcal{T}_t < 0$. Additionally the electricity infeeds from renewable energy sources increased, but this effect is two decimal powers smaller than the reduction of gross electricity demand and therefore negligible.

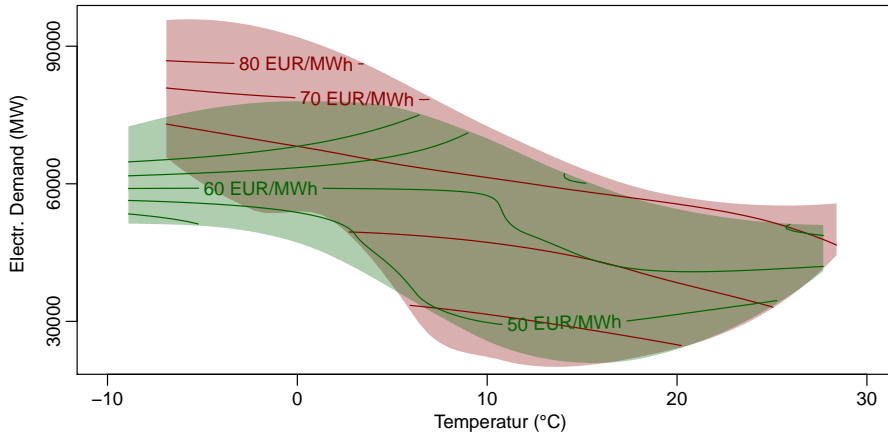


Figure 2.3: Contour plots of the estimated mean functions one year before the nuclear moratorium on March 14, 2011 (red shaded) and one year after (green shaded). The contour lines correspond to the following mean price levels: 40 EUR/MWh, 50 EUR/MWh, 60 EUR/MWh, 70 EUR/MWh, and 80 EUR/MWh.

The tempering effect of a reduced electricity demand is visualized by the contour lines in Figure 2.3. For each of the estimated mean functions $\hat{\mu}^{\text{bef}}$ and $\hat{\mu}^{\text{aft}}$, there are five contour lines, which correspond to the price levels: 40 EUR/MWh, 50 EUR/MWh, 60 EUR/MWh, 70 EUR/MWh, and 80 EUR/MWh. These contour lines show that the ranges of mean electricity prices before and after the moratorium are actually about the same.

In fact, taking the averages over all discretization points, say $\hat{\mu}_{it}^{\text{bef}} = \hat{\mu}^{\text{bef}}(\mathcal{D}_{it}^{\text{bef}}, \mathcal{T}_t^{\text{bef}}; h)$ and $\hat{\mu}_{it}^{\text{aft}} = \hat{\mu}^{\text{aft}}(\mathcal{D}_{it}^{\text{aft}}, \mathcal{T}_t^{\text{aft}}; h)$, leads to rather comparable mean values. In Table 2.2 we report these averages along with the average values over the discretization points of the two estimated mean functions that refer to off-peak hours. In order

to contrast also the estimated autocovariance functions $\hat{\gamma}^{\text{bef}}$ and $\hat{\gamma}^{\text{aft}}$ we take a look at the corresponding eigenfunctions \hat{p}_k^{bef} and \hat{p}_k^{aft} . A direct visualization of the estimated covariance functions is not practicable as these are mappings from \mathbb{R}^3 to \mathbb{R} . We abstain from a rotation of the eigenfunctions, since the application of the widely accepted VARIMAX rotation scheme lead to nearly the same shapes of the eigenfunctions and more advanced rotation schemes are not within the scope of this paper.

The empirical covariance function $\hat{\gamma}^{\text{t.sp}}$ can be approximated by the first K ordered, empirical eigencomponents, such that

$$\hat{\gamma}^{\text{t.sp}}(d, d', \tau; h) \approx \sum_{k=1}^K \hat{\lambda}_k^{\text{t.sp}}(\tau) \hat{p}_k^{\text{t.sp}}(d, \tau) \hat{p}_k^{\text{t.sp}}(d', \tau),$$

where the eigenfunction $\hat{p}_k^{\text{t.sp}}$ can be approximated from the vector $\hat{\mathbf{p}}_k$ using, e.g., B-spline basis expansions. The accuracy of the approximation is quantified by the cumulative variance criterion: $\sum_{k=1}^K \hat{\lambda}_k^{\text{t.sp}}(\tau) / \sum_{l=1}^L \hat{\lambda}_l^{\text{t.sp}}(\tau)$, where L denotes the number of positive eigenvalues. We choose $K = 2$, which allows for an approximation with 95% (or more) accuracy for both time spans and all temperature values τ .

In Figure 2.4 the eigenfunctions are plotted as univariate functions of electricity demand for different fixed temperature values τ . Eigenfunctions are only determined up to sign changes, but in order to get a clearly arranged plot, we show the eigenfunctions for one common direction only.

It is important to plot eigenfunctions at a meaningful scale in order to get a visual impression of their relative importance. The eigenfunctions are principal (pairwise orthonormal) variance directions of the random price functions. There-

MEAN-PRICES	Peak hours	Off-peak hours
Before Moratorium	57 EUR/MWh	42 EUR/MWh
After Moratorium	61 EUR/MWh	46 EUR/MWh

Table 2.2: Empirical mean values over all discretization points of the four estimated mean functions. The two mean functions, which refer to peak price hours, are plotted in Figure 2.3; the two other mean functions, which refer to off-peak price hours, are not plotted.

fore, a natural choice for a scaling parameter for the k th eigenfunction is the square root of the k th eigenvalue. This is equivalent to scaling the k th eigenfunction by one standard deviation of the random functions with respect to the corresponding variance direction.

Correspondingly, in Figure 2.4 we plot the graphs of the scaled eigenfunctions $\sqrt{\hat{\lambda}_k^{\text{t.sp}}(\tau)} \hat{p}_k^{\text{t.sp}}(d, \tau)$ with $d \in [\hat{a}^{\text{t.sp}}(\tau), \hat{b}^{\text{t.sp}}(\tau)]$ and τ equal to different temperature quantiles; in particular, we use the 0.025, 0.05, \dots , and 0.975 temperature quantiles.

Conditional eigenvalues for given temperature values, can be approximated by the eigenvalues of the $n_e \times n_e$ dimensional discretization matrices $\hat{M}^{\text{t.sp}}(\tau)$ with typical elements $\hat{M}^{\text{t.sp}}(\tau)_{ij} = \hat{\gamma}^{\text{t.sp}}(e_i, e_j, \tau; h)$, where e_i and e_j are evaluation values for the interval $[\hat{a}^{\text{t.sp}}(\tau; h), \hat{b}^{\text{t.sp}}(\tau; h)]$ with $i, j \in \{1, \dots, n_e\}$. We used the following discretization scheme with $n_e = 50$ evaluation points: $d_i = \hat{a}^{\text{t.sp}}(\tau) + (i - 1) \cdot [(\hat{b}^{\text{t.sp}}(\tau) - \hat{a}^{\text{t.sp}}(\tau)) / (n_e - 1)]$.

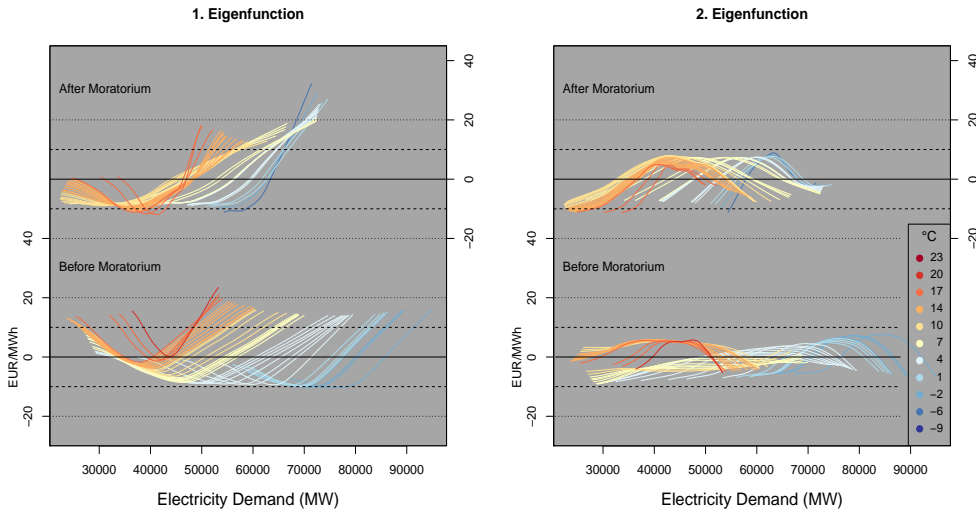


Figure 2.4: First and second eigenfunctions plotted as univariate functions of electricity demand for fixed temperature quantiles. (Selected temperature quantiles are: 0.025, 0.05, \dots , and 0.975.) The lower/upper sets of eigenfunctions refer to the time spans one year before/after Germany's nuclear moratorium on March 14, 2011.

The temperature dependent variance shares are visualized by the amount of how much the conditional eigenfunctions in Figure 2.4 deviate from the zero lines,

where the conditional variance shares are defined as $\hat{\lambda}_k^{\text{t.sp}}(\tau) / \sum_{l=1}^L \hat{\lambda}_l^{\text{t.sp}}(\tau)$ with $k \in \{1, 2\}$ and L being the number of positive eigenvalues. In both time spans, the first conditional eigenfunction explains between 60% and 90% of the total variance depending on the conditioning temperature values. The second conditional eigenfunctions explain still between 10% and 40% of the total variance; see also Figure 2.5.

A more eye-catching visualization of the temperature dependent variance shares is given in Figure 2.5. There the graphs of the dependent variance shares visualize an interesting peculiarity. In both time spans, the variance shares for the first/second eigenfunctions show global minima/maxima at about 10 degrees Celsius before the moratorium and about 15 degrees Celsius after the moratorium. This phenomenon has to be explained by seasonal changes in the compositions of the power plant portfolios.

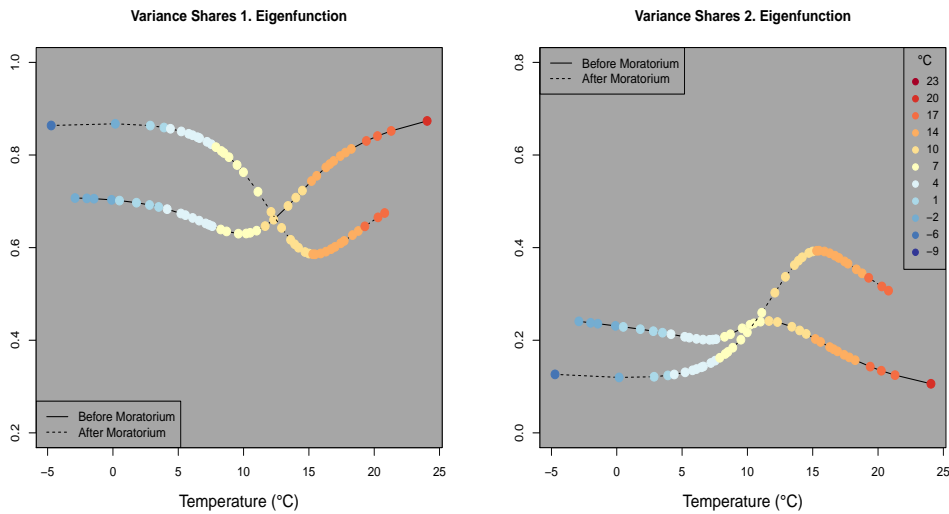


Figure 2.5: Visualization of the temperature dependent variance shares explained by the first (left panel) and the second (right panel) eigenfunctions.

Generally, the shapes of the eigenfunctions are more or less comparable, but a profound empirical analysis of the eigenfunctions demands for advanced rotation schemes, which are out of the scope of this paper.

Any specific interpretation of the eigenfunctions has to be done with care. For example, the first conditional eigenfunctions with respect to very cold tempera-

tures shows steeper slopes for the time span after the moratorium than before; see the left panel of Figure 2.4. It suggests itself to attribute this finding to the effects of the nuclear moratorium. However, the underlying reason is given by surprisingly cold winter-days in February 2012, which caused relatively high electricity spot prices. (In both time spans February is the coldest month, but the average temperature in February 2012 was minus three degrees Celsius below that of February 2011.)

In order to compare the eigenfunctions across the two time spans it is also of interest whether the eigenfunctions before the moratorium span the same space than those after the moratorium. Different eigenspaces across the time spans may point to structural changes in the underlying power plant portfolios. A very simple and intuitive way to check for differences in the eigenspaces is given by regression analysis. We regress discretized versions of the first and second conditional eigenfunctions of the time span before the moratorium on the first two conditional eigenfunctions of the time span after the moratorium. If the resulting R^2 values of the regressions are low, we have a descriptive hint on a difference between the corresponding eigenspaces. Of course, these regressions have to be done for common temperature values τ and overlapping parts of the intervals, i.e., $[\max(\hat{a}^{\text{bef}}(\tau; h), \hat{a}^{\text{aft}}(\tau; h)), \min(\hat{b}^{\text{bef}}(\tau; h), \hat{b}^{\text{aft}}(\tau; h))]$. The resulting R^2 values are all greater or equal to 0.96, which indicate that the eigenspaces conditionally on common temperature values are almost the same across the time periods.

In the remainder of this section we discuss the results from fitting the unobserved price functions \mathcal{P}_t . This goes without comparisons of the two time spans of our data set such that we can abstain from using the superscripts “t.sp”, “bef”, and “aft”.

Fitting the unobserved price functions

As mentioned in the introduction, we face a somewhat borderline case of sparse functional data. In fact, within the random subintervals $[A_t, B_t]$ the observed realizations of the discretization points \mathcal{D}_{it} are relatively densely sampled, but outside of these intervals there are no observations at all. This interesting phe-

nomenon is visualized Figure 2.6, where we plot the graphs of the functionals

$$\hat{\mathcal{P}}_t(d, \mathcal{T}_t) = \hat{\mu}(d, \mathcal{T}_t) + \sum_{k=1}^2 \hat{\beta}_{tk}(\mathcal{T}_t) \hat{p}_k(d, \mathcal{T}_t)$$

with $d \in [\hat{a}(\mathcal{T}_t), \hat{b}(\mathcal{T}_t)]$ and $\hat{\beta}_{tk}(\mathcal{T}_t)$ being the ordinary least squares estimate of $\beta_{tk}(\mathcal{T}_t)$ as discussed in Section 2.4.3. The black colored parts of the graphs indicate the regions of the random subintervals $[A_t, B_t]$ on which the discretization points are actually observed, where A_t and B_t are estimated by the extremum estimators A_t^{\min} and B_t^{\min} as discussed in Section 2.3.4.

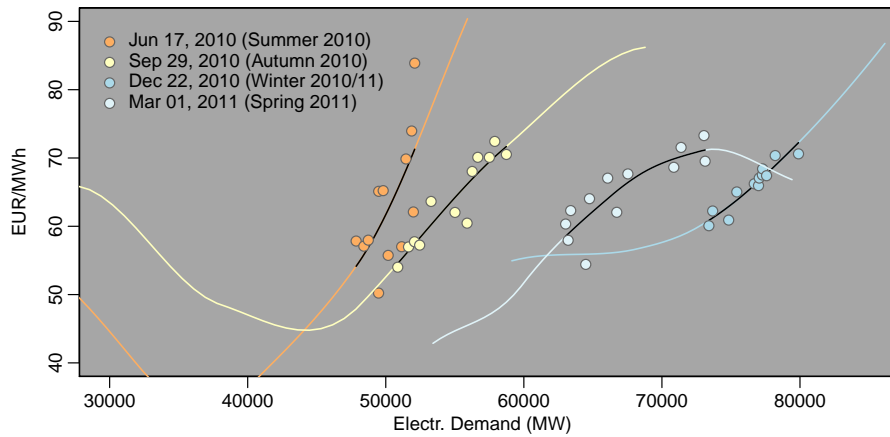


Figure 2.6: Four fitted price functions, where the days are chosen to have season-specific median temperatures for the time span one year before the moratorium (from March 15, 2010 to March 14, 2011). The same days are also emphasized in Figure 2.2.

Figure 2.6 complements our discussion of the ordinary least squares estimator $\hat{\beta}_{tk}(\mathcal{T}_t)$ in Section 2.4.3. The four fitted price functions are representative examples for whole data set: The model fit of the parts within the subintervals $[A_t^{\min}, B_t^{\max}]$ are throughout satisfactory. (The R^2 value for the whole data set is given by 0.92.) However, the prediction of the parts outside of the subintervals $[A_t^{\min}, B_t^{\max}]$ is generally not reliable, except for some exceptions such as shown by the graph of the fitted price function for December 22, 2010.

This visualizes nicely the problem of multicollinearity induced from observing

realizations of electricity demand only within subintervals of the whole domain. A numerical document for the four fitted price functions in Figure 2.6 can be given by the R^2 values from regressing the first eigenfunctions on the corresponding second eigenfunctions—under consideration of the respective subintervals. The R^2 value associated with the winter day December 22, 2010 is given by 0.43. By contrast, the R^2 values of the other three regressions are greater or equal 0.87 and point to an approximate linear relationship between the corresponding sub-parts of the eigenfunctions.

2.6 Conclusion

In this paper we propose to model hourly electricity spot prices as noisy discretization points of smooth random price functions, where the random price functions are specified as bivariate functions of the fundamental prediction variables electricity demand and air temperature. Theoretical foundation of this functional data point of view on electricity spot prices is given by the merit order model—the most important pricing model for electricity spot prices.

The consideration of the peculiar random design of the prediction variables induces problems similar to the case of sparse functional data. Within this environment, multivariate nonparametric regression is an important tool for functional principal component analysis. Our asymptotic bias and variance results for the multivariate local linear estimator account for the temporal dependencies between the random price functions, as well as the dependencies between spot prices coming from the same price functions.

Our theoretical considerations are concluded by a real data application. There, we apply our model to German electricity market data one year before and one year after Germany's nuclear phase-out on March 14, 2011 and contrast the estimation results. The real data study clearly confirms the usefulness of our functional data point of view on electricity spot prices.

2.7 Appendix: Proofs

In this section we give proofs our propositions and theorems. In order to facilitate readability we present a rather detailed version of the proof.

2.7.1 Proof of Proposition 2.4.1

The result for conditional variance of the error ε_{ijt} in part (ii) of Proposition 2.4.1:

$$\begin{aligned}\sigma_\varepsilon^2((\delta, \delta'), \tau) &= \text{Var}(\varepsilon_{ijt} | (\mathcal{D}_{it}, \mathcal{D}_{jt}, \mathcal{T}_t) = (\delta, \delta', \tau)) \\ &= \gamma((\delta, \delta), \tau) + \gamma((\delta', \delta'), \tau) + 4\sigma_\varepsilon^2\end{aligned}$$

can be seen by applying the conditional variance operator on the definition of ε_{ijt} in Eq. (2.8) and using formula (2) for the variance of products of mutually independent variables proposed by Goodman (1962).

All other results of Proposition 2.4.1 follow directly from our assumptions on the data-generating processes (2.3) and (2.7).

2.7.2 Proof of Theorem 2.4.1

Our proof of Theorem 2.4.1 follows that of Ruppert & Wand (1994), where the main difference is given by our Lemmas 2.7.1-2.7.3 below in the proof of Theorem 2.4.1, part (ii).

Proof of Theorem 2.4.1, part (i):

Let $x \in \mathcal{C}_X(q)$ and $q \in \mathcal{C}_Q$ be an interior points. Using d -variate Taylor-expansion, the conditional bias of the estimator $\hat{m}(x, q; H)$, defined in Eq. (2.15), can be written as

$$\begin{aligned}& E(\hat{m}(x, q; H) - m(x, q) | (X, Q)) \\ &= \frac{1}{2} e_1^T ((Tn)^{-1} (X_x, Q_q)^T W_{xq} (X_x, Q_q))^{-1} (Tn)^{-1} (X_x, Q_q)^T W_{xq} \left(\frac{1}{2} \mathcal{Q}_m(x, q) + R_m(x, q) \right),\end{aligned}\tag{2.23}$$

where $\mathcal{Q}_m(x, q)$ is a $Tn \times 1$ vector defined as $\mathcal{Q}_m(x, q) =$

$$(((X_{11}, Q_1) - (x, q))^T \mathcal{H}_m(x, q) ((X_{11}, Q_1) - (x, q)), \dots, ((X_{nT}, Q_T) - (x, q))^T \mathcal{H}_m(x, q) ((X_{nT}, Q_T) - (x, q)))^T$$

with $\mathcal{H}_m(x, q)$ being the $d \times d$ Hessian matrix of the regression function m at (x, q) and $R_m(x, q)$ is a $Tn \times 1$ vector of remainder terms defined as $R_m(x, q) = (o(((X_{11}, Q_1) - (x, q))^T((X_{11}, Q_1) - (x, q))), \dots, o(((X_{nT}, Q_T) - (x, q))^T((X_{nT}, Q_T) - (x, q))))^T$

Next, we derive asymptotic approximations for the two $(d + 1) \times (d + 1)$ matrices $((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1}$ and $(Tn)^{-1}(X_x, Q_q)^T W_{xq} \mathcal{Q}_m(x, q)$ of the right hand side of Eq. (2.23). Using standard procedures from kernel density estimation it is easy to derive that $(Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q) =$

$$\begin{pmatrix} f_{XQ,T}(x, q) + \mathcal{O}_p(\text{tr}(H_X)) + \mathcal{O}_p(\text{tr}(H_Q)) & \mu_2(K) D_{f_{XQ,T}}(x, q)^T H + o_p(\mathbf{1}^T H) \\ \mu_2(K) H D_{f_{XQ,T}}(x, q) + o_p(H \mathbf{1}) & \mu_2(K) H f_{XQ,T}(x, q) + o_p(H) \end{pmatrix},$$

where $f_{XQ,T}(x, q) = f_X(x|q) T^{-1} \sum_{t=1}^T f_Q(q|t/T)$, $H = \begin{pmatrix} H_X & 0 \\ 0 & H_Q \end{pmatrix}$, and $\mathbf{1}$ is a generic vector or matrix having appropriate dimensions and with all elements equal to one.

Standard results on inversion of block matrices yield

$$\left((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q) \right)^{-1} = \quad (2.24)$$

$$\begin{pmatrix} (f_{XQ,T}(x, q))^{-1} + \mathcal{O}_p(\text{tr}(H_X)) + \mathcal{O}_p(\text{tr}(H_Q)) & -D_{f_{XQ,T}}(x, q)^T (f_{XQ,T}(x, q))^{-2} + o_p(\mathbf{1}^T H^2) \\ -D_{f_{XQ,T}}(x, q) (f_{XQ,T}(x, q))^{-2} + o_p(H^2 \mathbf{1}) & (\mu_2(K) H f_{XQ,T}(x, q))^{-1} + o_p(H) \end{pmatrix}.$$

The second $(d + 1) \times (d + 1)$ matrix, namely $(Tn)^{-1}(X_x, Q_q)^T W_{xq} \mathcal{Q}_m(x, q)$, can be partitioned as following:

$$(Tn)^{-1} X_x^T W_x \mathcal{Q}_m(x, q) = \begin{pmatrix} \text{upper element} \\ \text{lower bloc} \end{pmatrix},$$

where the **upper element** is a scalar variable equal to

$$\begin{aligned}
& (Tn)^{-1} \sum_{it} K_{X,H_X}(X_{it} - x) K_{Q,H_Q}(Q_t - q) (X_{it} - x, Q_t - q) \mathcal{H}_m(x, q) ((X_{it} - x)^T, (Q_t - q)^T)^T \\
&= \int_{[-1,1]^{d_2}} \int_{[-1,1]^{d_1}} K_X(u) K_Q(v) (u^T H_X^{1/2}, v^T H_Q^{1/2}) \mathcal{H}_m(x, q) (H_X^{1/2} u, H_Q^{1/2} v)^T \cdot \\
&\quad \cdot f_{XQ,T}(x + H_X^{1/2} u, q + H_Q^{1/2} v) dudv + o_p(\mathbf{1}^T H \mathbf{1}) \\
&= \mathcal{O}_p(\mathbf{1}^T H \mathbf{1})
\end{aligned} \tag{2.25}$$

and the $d \times 1$ dimensional **lower bloc** is equal to

$$\begin{aligned}
& (Tn)^{-1} \sum_{it} \{ K_{H_X}(X_{it} - x) K_{H_Q}(Q_t - q) (X_{it} - x, Q_t - q) \mathcal{H}_m(x, q) ((X_{it} - x)^T, (Q_t - q)^T)^T \} \\
&\quad \cdot ((X_{it} - x)^T, (Q_t - q)^T)^T \\
&= \mathcal{O}_p(H^{3/2} \mathbf{1}).
\end{aligned} \tag{2.26}$$

Remember that at the beginning of Section 2.3.2 we define X_{it} and Q_t as $1 \times d_1$ and $1 \times d_2$ row-vectors.

Using the above derived expressions for the matrices $((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1}$ (see Eq. (2.24)) and $(Tn)^{-1}(X_x, Q_q)^T W_{xq} \mathcal{Q}_m(x, q)$ (see Eq.'s (2.25) and (2.26)), we can compute the first summand of the conditional bias expression in Eq. (2.23), which yields the scalar variable

$$\begin{aligned}
& \frac{1}{2} e_1^T ((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1} (Tn)^{-1}(X_x, Q_q)^T W_{xq} \mathcal{Q}_m(x, q) = \\
& \quad \frac{1}{2} (f_{XQ,T}(x, q))^{-1} \int_{[-1,1]^{d_2}} \int_{[-1,1]^{d_1}} K_X(u) K_Q(v) (u^T H_X^{1/2}, v^T H_Q^{1/2}) \mathcal{H}_m(x, q) (H_X^{1/2} u, H_Q^{1/2} v)^T \\
& \quad \cdot f_{XQ,T}(x + H_X^{1/2} u, q + H_Q^{1/2} v) dudv + o_p(\text{tr}(H)).
\end{aligned}$$

Furthermore, it is easily seen that the second summand of the conditional bias expression in Eq. (2.23), which holds the remainder term, is given by

$$\frac{1}{2} e_1^T ((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1} (Tn)^{-1}(X_x, Q_q)^T W_{xq} R_m(x, q) = o_p(\text{tr}(H)).$$

Summation of the two latter expressions and Taylor expansion of $f_{XQ,T}(\cdot, \cdot)$ around (x, q) yields the asymptotic approximation of the conditional bias

$$\begin{aligned}
E(\hat{m}(x, q; H) - m(x, q)|X) &= \\
&= \frac{1}{2}(f_{XQ,T}(x, q))^{-1} \int_{[-1,1]^{d_2}} \int_{[-1,1]^{d_1}} K_X(u)K_Q(v)(u^T H_X^{1/2}, v^T H_Q^{1/2})\mathcal{H}_m(x, q)(H_X^{1/2}u, H_Q^{1/2}v)^T \\
&\quad \cdot [f_{XQ,T}(x, q) + \mathcal{O}(\text{tr}(H_X^{1/2})) + \mathcal{O}(\text{tr}(H_Q^{1/2}))] dudv + o_p(\text{tr}(H)).
\end{aligned}$$

Using that $K((u, v)^T) = K_X(u)K_Q(v)$ and that H is a block diagonal matrix with diagonal blocks H_X and H_Q we can integrate with respect to $(u, v)^T$. This leads to the following compact expression: $E(\hat{m}(x, q; H) - m(x, q)|X) =$

$$= \frac{1}{2} \mu_2(K) \text{tr} \{H\mathcal{H}_m(x)\} + o_p(\text{tr}(H)) = \mathcal{O}_p(\text{tr}(H)),$$

which shows our bias statement of Theorem 2.4.1, part (i). \square

Proof of Theorem 2.4.1, part (ii): In the following we derive the conditional variance of the d -variate local linear estimator $\text{Var}(\hat{m}(x, q; H)|X) =$

$$\begin{aligned}
&= e_1^T ((X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1} \\
&\quad (X_x, Q_q)^T W_{xq} \cdot \text{Cov}(Y|(X, Q)) \cdot W_{xq}(X_x, Q_q) \\
&\quad ((X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1} e_1
\end{aligned} \tag{2.27}$$

$$\begin{aligned}
&= e_1^T ((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1} \\
&\quad ((Tn)^{-2}(X_x, Q_q)^T W_{xq} \cdot \text{Cov}(Y|(X, Q)) \cdot W_{xq}(X_x, Q_q)) \\
&\quad ((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q))^{-1} e_1,
\end{aligned}$$

where $\text{Cov}(Y|X)$ is the $Tn \times Tn$ matrix with typical elements

$$\begin{aligned}
&\text{Cov}(Y_{it}, Y_{js}|X_{it}, X_{js}, Q_t, Q_s) \\
&= \psi_{|t-s|}((X_{it}, X_{js}), (Q_t, Q_s)) + \sigma_\nu^2(X_{it}, Q_t)\mathbb{I}(i = j \text{ and } t = s)
\end{aligned}$$

with $\mathbb{I}(\text{true}) = 1$ and zero otherwise.

We begin with analyzing the $(d+1) \times (d+1)$ matrix

$$(Tn)^{-2}(X_x, Q_q)^T W_{xq} \cdot \text{Cov}(Y|(X, Q)) \cdot W_{xq}(X_x, Q_q)$$

using the following three Lemmas 2.7.1-2.7.3.

Lemma 2.7.1. *The upper-left scalar (block) of the matrix $(Tn)^{-2}(X_x, Q_q)^T W_{xq} Cov(Y|(X, Q)) W_{xq}(X_x, Q_q)$ is given by*

$$\begin{aligned}
& (Tn)^{-2} \mathbf{1}^T W_{xq} Cov(Y|(X, Q)) W_{xq} \mathbf{1} \\
&= (Tn)^{-1} f_{XQ,T}(x, q) |H|^{-1/2} R(K) \left(\psi((x, x), q) + \sigma_v^2(x, q) \right) \left(1 + \mathcal{O}_p(\text{tr}(H^{1/2})) \right) \\
&+ T^{-1} f_{XQ,T}^2(x, q) \left[\left(\frac{n-1}{n} \right) |H_Q|^{-1/2} R(K_Q) \psi((x, x), q) + C_r(x, q) \right] \\
&+ \mathcal{O}_p(T^{-1} \text{tr}(H^{1/2})) + \mathcal{O}_p(T^{-2}) \\
&= \mathcal{O}_p((Tn)^{-1} |H|^{-1/2}) + \mathcal{O}_p(T^{-1} |H_Q|^{-1/2}),
\end{aligned}$$

where $f_{XQ,T}(x, q) = T^{-1} \sum_{t=1}^T f_X(x|q) f(q|t/T)$, $f_{XQ,T}^2(x, q) = T^{-1} \sum_{t=1}^T (f_X(x|q) f(q|t/T))^2$, and $C_r(x, q)$ is a non-negative constant such that $0 \leq C_r(x, q) \leq \frac{2C_{\psi r}}{1-r}$, defined as $C_r(x, q) = 2 \sum_{u=1}^{T-1} \psi_u((x, x), (q, q))$.

Lemma 2.7.2. *The $1 \times d$ dimensional upper-right block of the matrix $(Tn)^{-2}(X_x, Q_q)^T W_{xq} Cov(Y|(X, Q)) W_{xq}(X_x, Q_q)$ is given by*

$$\begin{aligned}
& (Tn)^{-2} \mathbf{1}^T W_{xq} Cov(Y|(X, Q)) W_{xq} \begin{pmatrix} (X_{11} - x, Q_1 - q) \\ \vdots \\ (X_{nT} - x, Q_T - q) \end{pmatrix} \\
&= (Tn)^{-1} f_{XQ,T}(x, q) |H|^{-1/2} (\mathbf{1}^T H^{1/2}) R(K) \left(\psi((x, x), q) + \sigma_v^2(x, q) \right) \left(1 + \mathcal{O}_p(\text{tr}(H^{1/2})) \right) \\
&+ T^{-1} f_{XQ,T}^2(x, q) (\mathbf{1}^T H^{1/2}) \left[\left(\frac{n-1}{n} \right) |H_Q|^{-1/2} R(K_Q) \psi((x, x), q) + C_r(x, q) \right] \\
&+ \mathcal{O}_p(T^{-1} \text{tr}(H^{1/2})) + \mathcal{O}_p(T^{-2}) \\
&= \mathcal{O}_p((Tn)^{-1} |H|^{-1/2} (\mathbf{1}^T H^{1/2})) + \mathcal{O}_p(T^{-1} (\mathbf{1}^T H^{1/2}) |H_Q|^{-1/2}).
\end{aligned}$$

The $d \times 1$ dimensional lower-left block of the matrix $(Tn)^{-2} X_x^T W_x Cov(Y|X) W_x X_x$ is simply the transposed version of the latter result.

Lemma 2.7.3. *The $d \times d$ lower-right block of the matrix*

$(Tn)^{-2}(X_x, Q_q)^T W_{xq} Cov(Y|(X, Q)) W_{xq}(X_x, Q_q)$ is given by

$$\begin{aligned}
& (Tn)^{-2} \left(((X_{11} - x)^T, (Q_1 - q)^T)^T, \dots, ((X_{nT} - x)^T, (Q_T - q)^T)^T \right) \\
& \cdot W_{xq} Cov(Y|(X, Q)) W_{xq} \begin{pmatrix} (X_{11} - x, Q_1 - q) \\ \vdots \\ (X_{nT} - x, Q_T - q) \end{pmatrix} \\
& = (Tn)^{-1} f_{XQ,T}(x, q) |H|^{-1/2} HR(K) \left(\psi((x, x), q) + \sigma_\nu^2(x, q) \right) \left(1 + \mathcal{O}_p(tr(H^{1/2})) \right) \\
& + T^{-1} f_{XQ,T}^2(x, q) H \left[\left(\frac{n-1}{n} \right) |H_Q|^{-1/2} R(K_Q) \psi((x, x), q) + C_r(x, q) \right] \\
& + \mathcal{O}_p(T^{-1} tr(H^{1/2})) + \mathcal{O}_p(T^{-2}) \\
& = \mathcal{O}_p((Tn)^{-1} |H|^{-1/2} H) + \mathcal{O}_p(T^{-1} H |H_Q|^{-1/2}).
\end{aligned}$$

Using the approximations for the bloc-elements of the matrix $(Tn)^{-2}(X_x, Q_q)^T W_{xq} Cov(Y|(X, Q)) W_{xq}(X_x, Q_q)$, given by the Lemmas 2.7.1-2.7.3, and the approximation for the matrix $\left((Tn)^{-1}(X_x, Q_q)^T W_{xq}(X_x, Q_q) \right)^{-1}$, given in Eq. (2.24), we can approximate the conditional variance of the bivariate local linear estimator, given in Eq. (2.27). Some tedious yet straightforward matrix algebra leads to $Var(\hat{m}(x, q; H)|(X, Q)) =$

$$\begin{aligned}
& (Tn)^{-1} |H|^{-1/2} \left\{ \frac{R(K) (\psi((x, x), q) + \sigma_\nu^2(x, q))}{f_{XQ,T}(x, q)} \right\} \left(1 + \mathcal{O}_p(tr(H^{1/2})) \right) \\
& + T^{-1} C_f(x, q) \left[\left(\frac{n-1}{n} \right) |H_Q|^{-1/2} R(K_Q) \psi((x, x), q) + C_r(x, q) \right] \\
& + \mathcal{O}_p(T^{-1} tr(H^{1/2})) + \mathcal{O}_p(T^{-2}),
\end{aligned}$$

where $C_f(x, q)$ is a non-negative constant $0 < C_f(x, q) < 1$ for all $x \in \mathcal{C}_X(q)$ and $q \in \mathcal{C}_Q$, defined as

$$C_f(x, q) = \frac{f_{XQ,T}^2(x, q)}{(f_{XQ,T}(x, q))^2} = \frac{\sum_{t=1}^T (f_X(x|q) f_Q(q|t/T))^2}{(\sum_{t=1}^T f_X(x|q) f_Q(q|t/T))^2}.$$

Which of the above terms $\mathcal{O}_p(T^{-1} tr(H^{1/2}))$ and $\mathcal{O}_p(T^{-2})$ dominates the other depends on the order of magnitude of $tr(H^{1/2})$. The assumption that $(Tn)^{-1} |H|^{-1/2} \rightarrow 0$ as $Tn \rightarrow \infty$ and that $T^{-1} |H_Q|^{-1/2} \rightarrow 0$ as $T \rightarrow \infty$ (Assumption A3) leads to $|H|^{-1/2} \sim (Tn)^\delta$ and $|H_Q|^{-1/2} \sim T^{\delta_Q}$ for some δ and δ_Q with $0 < \delta < 1$ and

$0 < \delta_Q < 1$. Using the assumption that $n = n(T) \sim T^\beta$, with $\beta \geq 0$, it can be easily derived that $\text{tr}(H^{1/2}) \sim d_1 T^{(\delta_Q - (1+\beta)\delta)d_1} + d_2^{-\delta_Q/d_2}$. From this it follows that the term $\mathcal{O}(T^{-1}\text{tr}(H^{1/2}))$ always dominates the term $\mathcal{O}(T^{-2})$, since it is always the case that $-1 < -\delta_Q/d_2 < 0$. \square

Next we proof Lemma 2.7.1; the proofs of Lemmas 2.7.2 and 2.7.3 can be done correspondingly. Lemma 2.7.1 can be shown by considering the convergences of the mean and the variance of $(Tn)^{-2}\mathbf{1}^T W_{xq} \text{Cov}(Y|(X, Q)) W_{xq} \mathbf{1}$ against zero as $T \rightarrow \infty$. It is convenient to split the sum such that $(Tn)^{-2}\mathbf{1}^T W_{xq} \text{Cov}(Y|(X, Q)) W_{xq} \mathbf{1} = M_1 + M_2$ with

$$\begin{aligned} M_1 &= (Tn)^{-2} \sum_{it} (K_{X, H_X}(X_{it} - x) K_{Q, H_Q}(Q_t - q))^2 \text{Var}(Y_{it}|X_{it}, Q_t) & (2.28) \\ M_2 &= (Tn)^{-2} \sum_{ij} \sum_{\substack{ts \\ t \neq s}} K_{X, H_X}(X_{it} - x) K_{Q, H_Q}(Q_t - q) \text{Cov}(Y_{it}, Y_{js}|X_{it}, X_{js}, Q_t, Q_s) \\ &\quad \cdot K_{X, H_X}(X_{js} - x) K_{Q, H_Q}(Q_s - q) \\ &\quad + (Tn)^{-2} \sum_{ij} \sum_{\substack{t \\ i \neq j}} K_{X, H_X}(X_{it} - x) [K_{Q, H_Q}(Q_t - q)]^2 \text{Cov}(Y_{it}, Y_{jt}|X_{it}, X_{jt}, Q_t) \\ &\quad \cdot K_{X, H_X}(X_{jt} - x). & (2.29) \end{aligned}$$

Note that the factor $[K_{Q, H_Q}(Q_t - q)]^2$ in sum M_2 leads to the nonparametric convergence rate with respect to T .

First, we compute of the mean $E(M_1) =$

$$(Tn)^{-1} \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(v)} (K_{X, H_X}(u - x) K_{Q, H_Q}(v - q))^2 \text{Var}(Y_{1t}|u, v) f_{XQ, T}(u, v) du dv$$

with $\text{Var}(Y_{1t}|u, v) = \psi((u, u), v) + \sigma_\nu^2(u, v)$, substituting $H_X^{-1/2}(u - x) = w$ and $H_Q^{-1/2}(v - q) = z$, Taylor-expansions around (x, q) , and using the usual arguments leads to

$$\begin{aligned} E(M_1) &= (Tn)^{-1} |H|^{-1/2} f_{XQ, T}(x, q) R(K) \left(\psi((x, x), q) + \sigma_\nu^2(x, q) \right) & (2.30) \\ &\quad + \mathcal{O}((Tn)^{-1} |H|^{-1/2} \text{tr}(H^{1/2})), \end{aligned}$$

where $|H|^{-1/2} = |H_X|^{-1/2} |H_Q|^{-1/2}$, $R(K) := \int_{[-1, 1]^d} K((w, z)^T)^2 d(w, z)^T$ and $f_{XQ, T}(x, q) = f_X(x|q) T^{-1} \sum_{t=1}^T f_Q(q|t/T)$.

Second, we compute the mean $E(M_2) =$

$$\begin{aligned} & \frac{n^2}{(Tn)^2} \sum_{\substack{ts \\ t \neq s}} \int_{\mathcal{C}_Q} \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(v)} \int_{\mathcal{C}_X(\tilde{v})} K_{X,H_X}((u-x)) K_{Q,H_Q}((v-q)) \psi_{|t-s|}((u,\tilde{u}), (v,\tilde{v})) \\ & \cdot K_{X,H_X}((\tilde{u}-x)) K_{Q,H_Q}((\tilde{v}-q)) f_X(u|v) f_Q(v|t/T) f_X(\tilde{u}|\tilde{v}) f_Q(\tilde{v}|s/T) \, dud\tilde{u}dv d\tilde{v} \\ & + \frac{n^2-n}{(Tn)^2} \sum_t \int_{\mathcal{C}_Q} \int_{\mathcal{C}_X(v)} \int_{\mathcal{C}_X(v)} K_{X,H_X}((u-x)) [K_{Q,H_Q}((v-q))]^2 \psi((u,\tilde{u}), v) \\ & \cdot K_{X,H_X}((\tilde{u}-x)) f_X(u|v) f_Q(v|t/T) f_X(\tilde{u}|v) f_Q(v|t/T) \, dud\tilde{u}dv \end{aligned}$$

Substituting $H_X^{-1/2}(u-x) = w$, $H_X^{-1/2}(\tilde{u}-x) = \tilde{w}$, $H_Q^{-1/2}(v-q) = z$, and $H_Q^{-1/2}(\tilde{v}-q) = \tilde{z}$, Taylor-expansions around (x, q) , using the short-range dependence assumption that $\psi_{|t-s|}((x, x), q) \leq C_\psi r^{|t-s|}$, and using the usual arguments leads to the following approximation of $E(M_2) =$

$$\begin{aligned} & = \frac{1}{T^2} \sum_{\substack{ts \\ t \neq s}} [\psi_{|t-s|}((x, x), (q, q)) f_X(x|q) f_Q(q|t/T) f_X(x|q) f_Q(q|s/T) + \mathcal{O}(r^{|t-s|} \text{tr}(H^{1/2}))] \\ & + \frac{n-1}{T^2 n} \sum_t |H_Q|^{-1/2} R(K_Q) \cdot \psi((x, x), q) \cdot ((f_X(x|q) f_Q(q|t/T))^2 + \mathcal{O}(\text{tr}(H^{1/2}))) \end{aligned}$$

Further Taylor-expansion of $f_Q(q, s/T)$ around $f_Q(q, t/T)$ and using that $\sum_{t \neq s} \psi_{|t-s|}((x, x), (q, q)) = \mathcal{O}(T)$ yields to $E(M_2)$

$$\begin{aligned} & = \frac{1}{T^2} \left(T \left\{ \left(\frac{1}{T} \sum_{t=1}^T (f_X(x|q) f_Q(q|t/T))^2 \right) C_r(x, q) + \mathcal{O}(T^{-1}) \right\} + \mathcal{O}(T \text{tr}(H^{1/2})) \right) \\ & + \frac{n-1}{T^2 n} \left(T \left\{ \left(\frac{1}{T} \sum_{t=1}^T (f_X(x|q) f_Q(q|t/T))^2 \right) |H_Q|^{-1/2} R(K_Q) \psi((x, x), q) \right\} + \mathcal{O}(T \text{tr}(H^{1/2})) \right), \end{aligned}$$

where $C_r(x, q)$ is a non-negative constant with $0 \leq C_r(x, q) \leq \frac{2C_\psi r}{1-r} < \infty$, defined as $C_r(x, q) = 2 \sum_{u=1}^{T-1} \psi_u((x, x), (q, q))$. Further simplification leads to

$$\begin{aligned} E(M_2) & = \frac{1}{T} f_{XQ,T}^2(x, q) C_r(x, q) + \mathcal{O}(T^{-2}) + \mathcal{O}(T^{-1} \text{tr}(H^{1/2})) \\ & + \frac{1}{T} f_{XQ,T}^2(x, q) \left(\frac{n-1}{n} \right) |H_Q|^{-1/2} R(K_Q) \psi((x, x), q) + \mathcal{O}(T^{-1} \text{tr}(H^{1/2})), \end{aligned}$$

where $f_{XQ,T}^2(x, q) = T^{-1} \sum_{t=1}^T (f_X(x|q) f_Q(q|t/T))^2$.

Finally, we can derive that $E(M_2) =$

$$\begin{aligned} &= \frac{1}{T} f_{QX,T}^2(x, q) \left[\left(\frac{n-1}{n} \right) |H_Q|^{-1/2} R(K_Q) \psi((x, x), q) + C_r(x, q) \right] \\ &\quad + \mathcal{O}(T^{-1} \text{tr}(H^{1/2})) + \mathcal{O}(T^{-2}). \end{aligned} \quad (2.31)$$

Eq.'s (2.30) and (2.31) together describe the speed of convergence of the mean of $(Tn)^{-2} \mathbf{1}^T W_{xq} \text{Cov}(Y|(X, Q)) W_{xq} \mathbf{1}$ against zero as $T \rightarrow \infty$. In order to complete the proof, it remains for us to show that the variance converges against zero as well. Here, we do not have to derive precise convergence rates, but only need to show that $\text{Var}((Tn)^{-2} \mathbf{1}^T W_{xq} \text{Cov}(Y|(X, Q)) W_{xq} \mathbf{1}) = o(1)$. This follows from standard arguments on averages over weakly dependent random variables.

Lemmas 2.7.2 and 2.7.3 differ from Lemma 2.7.1 only with respect to the additional factors $\mathbf{1}^T H^{1/2}$ and H . These come in due to the usual substitution step for the additional data parts $(X_{it} - x, Q_t - q)$. \square

2.7.3 Proof of Theorem 2.4.2

In the following we proof Theorem 2.4.2.

For the asymptotic behavior of the eigenvalues $\hat{\lambda}_k$ of the matrix \hat{M} we proof the following claims

$$\hat{\lambda}_k(\tau) - \lambda_k(\tau) = \mathbf{p}_k(\tau)^T [\hat{M}(\tau) - M(\tau)] \mathbf{p}_k(\tau) + \mathcal{O}_p(T^{-1}) \quad \text{and} \quad (2.32)$$

$$= \mathcal{O}_p(T^{-1/2}). \quad (2.33)$$

For the asymptotic behavior of the eigenvectors $\hat{\mathbf{p}}_k(\tau)$ of the matrix $\hat{M}(\tau)$ we proof the claim that

$$\|\hat{\mathbf{p}}_k(\tau) - \mathbf{p}_k(\tau)\|_2 = \mathcal{O}_p(T^{-1/2}), \quad (2.34)$$

where $\|\cdot\|_2$ is the Euclidean norm on \mathbb{R}^{n_e} .

We start with our proofs regarding the convergence results for the eigenvalues

in Eq.'s (2.32) and (2.33). Using Eq.'s (20)-(22) of Kneip & Utikal (2001) we can derive for the difference between the matrices $\hat{M}(\tau)$ and $M(\tau)$:

$$\begin{aligned}
\|(\hat{M}(\tau) - M(\tau))\|_2^2 &= \sup_{\|v\|_2=1} (v^T(\hat{M}(\tau) - M(\tau))^T(\hat{M}(\tau) - M(\tau))v) \\
&= \sup_{\|v\|_2=1} \text{tr} \left((\hat{M}(\tau) - M(\tau))^T(\hat{M}(\tau) - M(\tau))vv^T \right) \\
&\leq \text{tr} \left((\hat{M}(\tau) - M(\tau))^T(\hat{M}(\tau) - M(\tau)) \right) \\
&= \sum_{ij} (\hat{M}(\tau)_{ij} - M(\tau)_{ij})^2 \\
&= \mathcal{O}_p \left(n_e^2 \text{tr}(H)^2 \right) \\
\|(\hat{M}(\tau) - M(\tau))\|_2 &= \mathcal{O}_p \left(n_e \text{tr}(H) \right).
\end{aligned}$$

For our further computations we use part a) of the Lemma A.1. in Kneip & Utikal (2001), where for simplicity we consider only the case of eigenvalues with multiplicity one:

Lemma A.1. a) of Kneip & Utikal (2001)

$$\begin{aligned}
\hat{\lambda}(\tau)_k - \lambda(\tau)_k &= \mathbf{p}(\tau)_r^T (\hat{M}(\tau) - M(\tau)) \mathbf{p}(\tau)_r + R(\tau)_1 \\
|R(\tau)_1| &\leq \frac{6 \|(\hat{M}(\tau) - M(\tau))\|_2^2}{\min_{\lambda(\tau) \in EG(M(\tau)), \lambda(\tau) \neq \lambda(\tau)_k} |\lambda(\tau) - \lambda(\tau)_k|},
\end{aligned}$$

where $EG(M(\tau)) = \{\lambda_1(\tau), \dots, \lambda_{n_e}(\tau)\}$ with $\lambda_1(\tau) > \dots > \lambda_{n_e}(\tau)$.

Using Assumption B1, i.e., $\min_{s=1, \dots, n; s \neq r} |\lambda_k(\tau) - \lambda_s(\tau)| \geq n_e C_{3,r}$, the remainder term $R_1(\tau)$ can be simplified such that:

$$\hat{\lambda}_k(\tau) - \lambda_k(\tau) = \mathbf{p}_k(\tau)^T (\hat{M}(\tau) - M(\tau)) \mathbf{p}_k(\tau) + \mathcal{O}_p(n_e \text{tr}(H)^2),$$

where $\text{tr}(H)^2 = dh^4$. Using the assumption that $|H|^{1/2} = h^d \sim (Tn)^{-\delta}$ with $\frac{1}{4}d \leq \delta < 1$, it can be shown that $\text{tr}(H)^2 = \mathcal{O}((Tn)^{-1})$. This shows Claim (2.32). \square

A more raw quantification of the estimation error is given in the following:

$$\begin{aligned}
E\left(\mathbf{p}_k(\tau)^T[\hat{M}(\tau) - M(\tau)]\mathbf{p}_k(\tau)\right) &= \mathbf{p}_k(\tau)^T E(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau) \\
&= \text{tr}(\mathbf{p}_k(\tau)^T E(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau)) \\
&\leq \text{tr}(E(\hat{M}(\tau) - M(\tau))) \\
&= \sum_{i=1}^{n_e} E(\hat{M}(\tau)_{ii} - M(\tau)_{ii}) \\
&= n_e(O_p(\text{tr}(H))),
\end{aligned}$$

which yields $\mathbf{p}_k(\tau)^T(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau) = O_p(n_e \text{tr}(H))$. By the same arguments as above, we have $\text{tr}(H) = \mathcal{O}((Tn)^{-1/2})$. This shows the Claim (2.33). \square

In the following we show Claim (2.34), which is based on part b) of the Lemma A1 of Kneip & Utikal (2001):

$$\begin{aligned}
\hat{\mathbf{p}}_k(\tau) - \mathbf{p}_k(\tau) &= -S_k(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau) + R_2(\tau) \\
S_k(\tau) &= \sum_{s \neq r} \frac{1}{\lambda_s(\tau) - \lambda_k(\tau)} \mathbf{p}_s(\tau)\mathbf{p}_s(\tau)^T \\
R_2(\tau) &\leq \frac{6 \|\hat{M}(\tau) - M(\tau)\|_2^2}{\min_{\lambda(\tau) \in EG(M(\tau)), \lambda(\tau) \neq \lambda_k(\tau)} |\lambda(\tau) - \lambda_k(\tau)|^2}.
\end{aligned}$$

This allows us to formalize the following approximation:

$$\|\hat{\mathbf{p}}_k(\tau) - \mathbf{p}_k(\tau)\| \leq \|S_k(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau)\|_2 + \|R_2(\tau)\|$$

Using Assumption B1, i.e., $\min_{s=1, \dots, n; s \neq r} |\lambda_k(\tau) - \lambda_s(\tau)| \geq n_e C_{3,r}$, and our above derived approximation $\|\hat{M}(\tau) - M(\tau)\|_2^2 = \mathcal{O}_p(\text{tr}(H)^2)$ it is easily seen that:

$$\|R_2(\tau)\| = \mathcal{O}_p(\text{tr}(H)^2).$$

Furthermore:

$$\begin{aligned}
\|S_k(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau)\|_2 &= \left\| \left(\sum_{s \neq r} \frac{1}{\lambda_s(\tau) - \lambda_k(\tau)} \mathbf{p}_s(\tau) \mathbf{p}_s(\tau)^T \right) (\hat{M}(\tau) - M(\tau)) \mathbf{p}_k(\tau) \right\|_2 \\
&\leq \left\| \left(\frac{1}{l_k} \sum_{s \neq r} \mathbf{p}_s(\tau) \mathbf{p}_s(\tau)^T \right) (\hat{M}(\tau) - M(\tau)) \mathbf{p}_k(\tau) \right\|_2 \\
&= \frac{1}{l_k} \left\| \sum_{s \neq r} \mathbf{p}(\tau)_s \text{tr}(\mathbf{p}_s(\tau)^T (\hat{M}(\tau) - M(\tau)) \mathbf{p}_k(\tau)) \right\|_2 \\
&\leq \frac{1}{l_k} \left\| \text{tr}(\hat{M}(\tau) - M(\tau)) \sum_{s \neq r} \mathbf{p}_s(\tau) \right\|_2 \\
&= \frac{1}{l_k} \left\| \left(\sum_{i=1}^{n_e} [\hat{M}(\tau)_{ii} - M(\tau)_{ii}] \right) \sum_{s \neq r} \mathbf{p}_s(\tau) \right\|_2,
\end{aligned}$$

where $l_k = \min\{|\lambda_{r-1}(\tau) - \lambda_r(\tau)|, |\lambda_{r+1}(\tau) - \lambda_r(\tau)|\}$. With $l_k = \mathcal{O}(n_e)$ and $E(\hat{m}(x, q; H) - m(x, q)) = \mathcal{O}_p(\text{tr}(H))$ we have

$$\|S_k(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau)\|_2 = \mathcal{O}_p(\text{tr}(H)).$$

This shows Claim (2.34), since

$$\begin{aligned}
\|\hat{\mathbf{p}}_k(\tau) - \mathbf{p}_k(\tau)\| &\leq \|S_k(\hat{M}(\tau) - M(\tau))\mathbf{p}_k(\tau)\|_2 + \|R_2(\tau)\| \\
&= \mathcal{O}_p(\text{tr}(H)) + \mathcal{O}_p(\text{tr}(H)^2) \\
&= \mathcal{O}_p(\text{tr}(H)).
\end{aligned}$$

By the same arguments as above, we have $\text{tr}(H) = \mathcal{O}((Tn)^{-1/2})$. □

Chapter 3

The R-package phtt: Panel Data Analysis with Heterogeneous Time Trends

The R-package phtt provides estimation procedures for panel data with large dimensions n , T , and general forms of unobservable heterogeneous effects. Particularly, the estimation procedures are those of Bai (2009) and Kneip *et al.* (2012b), which complement one another very well: both models assume the unobservable heterogeneous effects to have a factor structure. The method of Bai (2009) assumes that the factors are stationary, whereas the method of Kneip *et al.* (2012b) allows the factors to be non-stationary. Additionally, the phtt package provides a wide range of dimensionality criteria in order to estimate the number of the unobserved factors simultaneously with the remaining model parameters.

3.1 Introduction

One of the main difficulties and at the same time appealing advantages of panel models is their need to deal with the problem of unobserved heterogeneity. Classical panel models, such as *fixed effects* or *random effects*, try to model unobserved heterogeneity using dummy variables or structural assumptions on the error term (see, e.g., Baltagi (2005)). In both cases the unobserved heterogeneity

is assumed to remain constant over time within each cross-sectional unit—apart from an eventual common time trend. This assumption might be reasonable for approximating panel data with fairly small temporal dimensions T ; however, for panel data with large T this assumption becomes implausible.

Nowadays, the availability of panel data with large cross-sectional dimensions n and large time dimensions T has triggered the development of a new class of panel data models. Recent discussions by Ahn *et al.* (2006), Pesaran (2006), Bai (2009), Bai *et al.* (2009), and Kneip *et al.* (2012b) have focused on advanced panel models for which the unobservable individual effects are allowed to have heterogeneous time trends that can be approximated by a factor structure. The basic form of this new class of panel models can be presented as follows:

$$y_{it} = \sum_{j=1}^P x_{itj} \beta_j + \nu_{it} + \epsilon_{it}, \quad \text{for } i \in \{1, \dots, n\}, \quad \text{and } t \in \{1, \dots, T\}, \quad (3.1)$$

where y_{it} is the dependent variable for each individual i at time t , x_{itj} is the j th element of the vector of explanatory variables $x_{it} \in \mathbb{R}^P$, and ϵ_{it} is the idiosyncratic error term. The time-varying individual effects $\nu_{it} \in \mathbb{R}$ of individual i for the time points $t \in \{1, \dots, T\}$ are assumed to have a d -dimensional factor structure. The following two specifications of the time-varying individual effects ν_{it} are implemented in the R package phtt:

$$\nu_{it} = \begin{cases} v_{it} &= \sum_{l=1}^d \lambda_{il} f_{lt}, & \text{for the model of Bai (2009),} \\ v_i(t) &= \sum_{l=1}^d \lambda_{il} f_l(t), & \text{for the model of Kneip } et al. (2012b). \end{cases} \quad (3.2)$$

Here, λ_{il} are unobserved individual loadings parameters, f_{lt} are unobserved common factors for the model of Bai (2009), $f_l(t)$ are the unobserved common factors for the model of Kneip *et al.* (2012b), and d is the unknown factor dimension. We consider the standard case of iid error terms ϵ_{it} with $E(\epsilon_{it}) = 0$ and $V(\epsilon_{it}) = \sigma^2$.

Note that the explicit consideration of an intercept in model (3.1) is not necessary but may facilitate interpretations. If x_{it} includes an intercept, the time-varying individual effects ν_{it} are centered around zero. If x_{it} does not include an intercept, the time-varying individual effects ν_{it} are centered around the overall mean.

Model (3.1) includes the classical panel data models with additive time-invariant individual effects and common time-specific effects. This model is obtained by choosing $d = 2$ with a first common factor $f_{1t} = 1$ for all $t \in \{1, \dots, T\}$ that has individual loadings parameters λ_{i1} , and a second common factor f_{2t} that has the same loadings parameter $\lambda_{i2} = 1$ for all $i \in \{1, \dots, n\}$.

An intrinsic problem of factor models lies in the fact that the true factors are only identifiable up to rotation. In order to ensure the uniqueness of these parameters, a number of d^2 restrictions are required. The usual normalization conditions are given by

- (a) $\frac{1}{T} \sum_{t=1}^T f_{lt}^2 = 1$ for all $l \in \{1, \dots, d\}$,
- (b) $\sum_{t=1}^T f_{lt} f_{kt} = 0$ for all $l, k \in \{1, \dots, d\}$ with $k \neq l$, and
- (c) $\sum_{i=1}^N \lambda_{il} \lambda_{ik} = 0$ for all $l, k \in \{1, \dots, d\}$ with $k \neq l$.

For the model of Kneip *et al.* (2012b), f_{lt} in conditions (a) and (b) has to be replaced by $f_l(t)$. As usual in factor models, a certain degree of indeterminacy remains, because the factors can only be determined up to sign changes.

Kneip *et al.* (2012b) consider the case in which the common factors $f_l(t)$ show relatively smooth patterns over time. This includes strongly positive auto-correlated stationary as well as non-stationary factors. The authors propose to approximate the time-varying individual effects $v_i(t)$ by smooth functions $\vartheta_i(t)$. In this way (3.1) becomes a semi-parametric model and its estimation is done using a two-step estimation procedure, explained in more detail in Section 3.2.

Alternatively, Bai (2009) proposes an iterated least squares approach to estimate (3.1) for stationary time-varying individual effects v_{it} such as ARMA or white noise processes. The estimators are the result of an iterative procedure solving a system of non-linear equations. However, Bai (2009) assumes the factor dimension d to be a known parameter, which is usually not the case. Therefore, the phtt package uses an algorithmic refinement of Bai's method proposed by Bada & Kneip (2010) in order to estimate the number of unobserved common factors d jointly with the remaining model parameters; see Section 3.4 for more details.

Besides the implementations of the methods proposed by Kneip *et al.* (2012b), Bai (2009), and Bada & Kneip (2010) the R package phtt comes with a wide range

of criteria (13 in total) for estimating the factor dimension d . The main functions of the phtt package are given in the following list:

- `KSS()` : Computes the estimators of the model parameters according to the method of Kneip *et al.* (2012b); see Section 3.2
- `Eup()` : Computes the estimators of the model parameters according to the method of Bai (2009) and Bada & Kneip (2010); see Section 3.4
- `OptDim()` : Allows for a comparison of the estimated factor dimensions \hat{d} obtained from many different panel criteria; see Section 3.3
- `checkSpecif()` : Tests whether to use a classical fixed effects panel model or a panel model with individual effects ν_{it} ; see Section 3.5.1

The functions are provided with `print()`-, `summary()`-, `plot()`-, `coef()`- and `residuals()`-methods.

Standard methods for estimating models for panel and longitudinal data are also implemented in the R R Development Core Team (2012) packages `plm` (Croissant & Millo, 2008), `nlme` (Pinheiro, Bates, DebRoy, Sarkar & R Core team, 2012), and `lme4` (Bates, Maechler & Bolker, 2012); see Croissant & Millo (2008) for an exhaustive comparison of these packages. Recently, Millo & Piras (2012) published the R package `splm` for spatial panel data models. The phtt package further extends the toolbox for statisticians and econometricians and provides the possibility of analyzing panel data in the case when the unobserved heterogeneity is time-varying.

To the best of our knowledge, the phtt package Bada & Liebl (2012) is the first software package that offers the estimation methods of Bai (2009) and Kneip *et al.* (2012b). Regarding the different dimensionality criteria (in total 13) that can be accessed via the function `OptDim()` only those of Bai & Ng (2002) are publicly available as MATLAB codes (The MathWorks Inc., 2012) from the homepage of Serena Ng (<http://www.columbia.edu/~sn2294/>).

To demonstrate the use of our functions, we re-explore the well known `Cigar` dataset, which is frequently used in the literature of panel models. The panel contains the amounts of cigarette consumption of $n = 46$ American states from

1963 to 1992 ($T = 30$) as well as data about the income per capita and cigarette prices in the same states during the same period (see, e.g., Baltagi & Levin (1986) for more details on the dataset).

We follow Baltagi & Li (2004), who estimate the following panel model:

$$\ln(\text{Consumption}_{it}) = \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + e_{it}. \quad (3.3)$$

Here, Consumption_{it} presents the sales of cigarettes (packs of cigarettes per capita), Price_{it} is the average real retail price of cigarettes, and Income_{it} is the real disposable income per capita. The index $i \in \{1, \dots, 46\}$ denotes the single states and the index $t \in \{1, \dots, 30\}$ denotes the year.

Baltagi & Li (2004) assume the error term e_{it} to be affected by time-varying spatial correlations between neighboring states. To estimate the model, the authors use a pre-defined $n \times n$ spatial weights matrix $W = \{\omega_{ij}\}_{i,j=1,\dots,n}$, where ω_{ij} is equal to one if state i and state j are neighboring states and zero else. If a state i has more than one neighboring state then the corresponding ω_{ij} 's are normalized to sum up to one.

However, the model of Baltagi & Li (2004) is very restrictive, since the assumptions on the structure of the error term e_{it} are fixed a priori. Instead, we apply the panel methods introduced above and allow for the state-cross-correlations in the error term e_{it} to be approximated from the data by a multidimensional factor structure such that

$$e_{it} = \sum_{l=1}^d \lambda_{il} f_{lt} + \epsilon_{it}.$$

The `Cigar` dataset can be obtained from the `pht` package using the function `data("Cigar")`. The panels of the variables $\ln(\text{Consumption}_{it})$, $\ln(\text{Price}_{it})$, and $\ln(\text{Income}_{it})$ are shown in Figure 3.1.

Section 3.2 is devoted to a short introduction of the method of Kneip *et al.* (2012b), which is appropriate for relatively smooth common factors $f_l(t)$. Section 3.3 presents the usage of the function `OptDim()`, which provides access to a wide range of panel dimensionality criteria recently discussed in the literature on factor models. Section 3.4 deals with the explanation as well as application

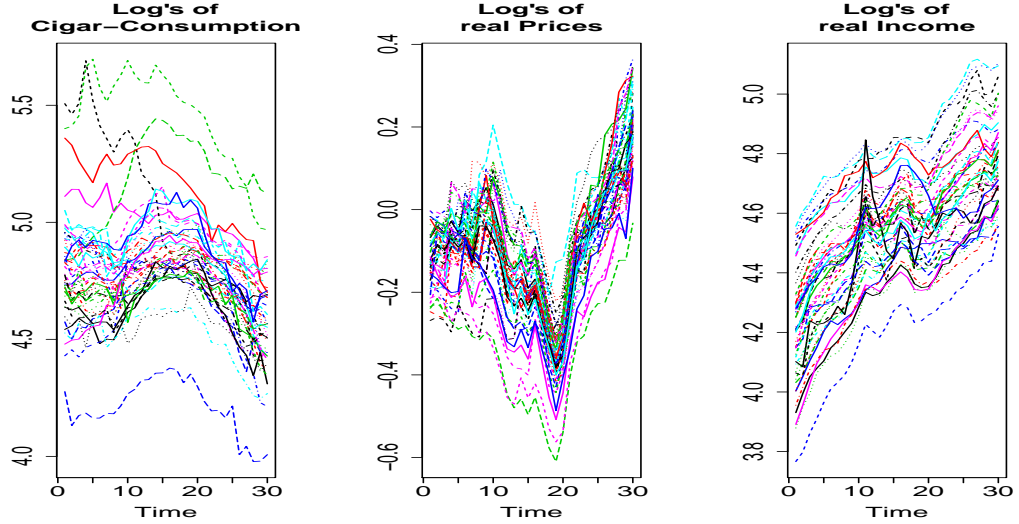


Figure 3.1: Plots of the dependent variable $\ln(\text{Consumption}_{it})$ and regressor variables $\ln(\text{Price}_{it})$ and $\ln(\text{Income}_{it})$.

of the panel method proposed by Bai (2009), which is appropriate for stationary and relatively unstructured common factors f_{it} .

3.2 Panel models for heterogeneity in time trends

The panel model proposed by Kneip *et al.* (2012b) can be presented as follows:

$$y_{it} = \sum_{j=1}^P x_{itj} \beta_j + v_i(t) + \epsilon_{it}, \quad (3.4)$$

where the time-varying individual effects $v_i(t)$ are parametrized in terms of common non-parametric basis functions $f_1(t), \dots, f_d(t)$ such that

$$v_i(t) = \sum_{l=1}^d \lambda_{il} f_l(t). \quad (3.5)$$

The asymptotic properties of this method rely on second order differences of $v_i(t)$, which apply for continuous functions as well as for classical discrete stochastic time series processes such as (S)AR(I)MA processes. Therefore, the functional

notation of the time-varying individual effects $v_i(t)$ and their underlying common factors $f_1(t), \dots, f_d(t)$ does not restrict them to a purely functional interpretation. The main idea of this approach is to approximate the time series of the time-varying individual effects $v_i(t)$ by smooth functions $\vartheta_i(t)$.

The estimation approach proposed by Kneip *et al.* (2012b) relies on a two-step procedure: first, estimates of the common slope parameters β_j and the time-varying individual effects $v_i(t)$ are obtained semi-parametrically. Second, functional principal component analysis is used to estimate the common factors $f_1(t), \dots, f_d(t)$, and to re-estimate the time-varying individual effects $v_i(t)$ more efficiently. In the following we describe both steps in more detail.

Step 1: The unobserved parameters β_j and $v_i(t)$ are estimated by the minimization of

$$\sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \left(y_{it} - \sum_{j=1}^P x_{itj} \beta_j - \vartheta_i(t) \right)^2 + \sum_{i=1}^n \kappa \int_1^T \frac{1}{T} \left(\vartheta_i^{(m)}(s) \right)^2 ds, \quad (3.6)$$

over all $\beta_j \in \mathbb{R}$ and all m -times continuously differentiable functions ϑ_i , where $\vartheta_i^{(m)}$ denotes the m th derivative of the function ϑ_i . A first approximation of $v_i(t)$ is then given by $\tilde{v}_i(t) := \hat{\vartheta}_i(t)$. Spline theory implies that any solution $\hat{\vartheta}_i(t)$ possesses an expansion in terms of a natural spline basis $z_1(t), \dots, z_T(t)$ such that $\hat{\vartheta}_i(t) = \sum_{s=1}^T \hat{\zeta}_{is} z_s(t)$; see, e.g., De Boor (2001b). Using the latter expression, we can rewrite (3.6) to formalize the following objective function:

$$S(\beta, \zeta) = \sum_{i=1}^n \left(\|Y_i - X_i \beta - Z \zeta_i\|^2 + \kappa \zeta_i^\top R \zeta_i \right), \quad (3.7)$$

where $Y_i = (y_{i1}, \dots, y_{iT})^\top$, $X_i = (x_{i1}^\top, \dots, x_{iT}^\top)^\top$, $x_{it} = (x_{it1}, \dots, x_{itp})^\top$, $\beta = (\beta_1, \dots, \beta_p)^\top$, $\zeta_i = (\zeta_{i1}, \dots, \zeta_{iT})^\top$, Z and R are $T \times T$ matrices with elements $\{z_s(t)\}_{s,t=1,\dots,T}$ and $\{\int z_s^{(m)}(t) z_k^{(m)}(t) dt\}_{s,k=1,\dots,T}$ respectively. κ is a preselected smoothing parameter to control the smoothness of $\hat{\vartheta}_i(t)$. We follow the usual choice of $m = 2$, which leads to cubic smoothing splines.

The semi-parametric estimators $\hat{\beta}, \hat{\zeta}_i = (\hat{\zeta}_{i1}, \dots, \hat{\zeta}_{iT})^\top$, and $\tilde{v}_i = (\tilde{v}_{i1}, \dots, \tilde{v}_{iT})^\top$ can be obtained by minimizing $S(\beta, \zeta)$ over all $\beta \in \mathbb{R}^p$ and $\zeta \in \mathbb{R}^{T \times n}$.

The solutions are given by

$$\hat{\beta} = \left(\sum_{i=1}^N X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1} \left(\sum_{i=1}^N X_i^\top (I - \mathcal{Z}_\kappa) Y_i \right), \quad (3.8)$$

$$\hat{\zeta}_i = (Z^\top Z + \kappa R)^{-1} Z^\top (Y_i - X_i \hat{\beta}), \text{ and} \quad (3.9)$$

$$\tilde{v}_i = \mathcal{Z}_\kappa (Y_i - X_i \hat{\beta}), \text{ where } \mathcal{Z}_\kappa = Z (Z^\top Z + \kappa R)^{-1} Z^\top. \quad (3.10)$$

Step 2: The common factors are obtained by the first d eigenvectors $\hat{\gamma}_1, \dots, \hat{\gamma}_d$ that correspond to the largest eigenvalues $\hat{\rho}_1, \dots, \hat{\rho}_d$ of the empirical covariance matrix

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \tilde{v}_i \tilde{v}_i^\top. \quad (3.11)$$

The estimator of the common factor $f_l(t)$ is then defined by the l th scaled eigenvector

$$\hat{f}_l(t) = \sqrt{T} \hat{\gamma}_{lt} \text{ for all } l \in \{1, \dots, d\}, \quad (3.12)$$

where $\hat{\gamma}_{lt}$ is the t th element of the eigenvector $\hat{\gamma}_l$. The scaling factor \sqrt{T} yields that $\hat{f}_l(t)$ satisfies the normalization condition $\frac{1}{T} \sum_{t=1}^T \hat{f}_l(t)^2 = 1$ as listed above in Section 3.1. The estimates of the individual loadings parameters λ_{il} are obtained by ordinary least squares regressions of $(Y_i - X_i \hat{\beta})$ on \hat{f}_l , where $\hat{f}_l = (\hat{f}_l(1), \dots, \hat{f}_l(T))'$. Recall from conditions (a) and (b) that $\hat{\lambda}_{il}$ can be calculated as follows:

$$\hat{\lambda}_{il} = \frac{1}{T} \hat{f}_l^\top (Y_i - X_i \hat{\beta}). \quad (3.13)$$

The time-varying individual effects $v_i(t)$ are re-estimated by $\hat{v}_i(t) := \sum_{l=1}^d \hat{\lambda}_{il} \hat{f}_l(t)$, where the factor dimension d can be determined, e.g., by the sequential testing procedure of Kneip *et al.* (2012b) or by any other dimensionality criteria. In Section 3.3 we introduce several such criteria.

To determine the optimal smoothing parameter κ_{opt} , Kneip *et al.* (2012b) propose the following cross validation (CV) criterion:

$$CV(\kappa) = \sum_{i=1}^n \|Y_i - X_i \hat{\beta}_{-i} - \sum_{l=1}^d \hat{\lambda}_{-i,l} \hat{f}_{-i,l}\|^2, \quad (3.14)$$

where $\hat{\beta}_{-i}$, $\hat{\lambda}_{-i,l}$, and $\hat{f}_{-i,l}$ are estimates of the parameters β , λ , and f_l based on

the dataset without the i th observation. Unfortunately, this criterion is computationally very costly and requires determining the factor dimension d in advance. To overcome this disadvantage, we propose a plug-in smoothing parameter that is discussed in more detail in Section 3.2.1.

Kneip *et al.* (2012b) derive the consistency of the estimators as $n, T \rightarrow \infty$ and show that the asymptotic distribution of common slope estimators is given by $\hat{\Sigma}_\beta^{-1/2}(\hat{\beta} - E_\epsilon(\hat{\beta})) \xrightarrow{d} \mathbf{N}(0, I)$, where

$$\hat{\Sigma}_\beta = \sigma^2 \left(\sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1} \left(\sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa)^2 X_i \right) \left(\sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1}. \quad (3.15)$$

A consistent estimator of σ^2 can be obtained by

$$\hat{\sigma}^2 = \frac{1}{(n-1)T} \sum_{i=1}^n \|Y_i - X_i \hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{i,l} \hat{f}_l\|^2. \quad (3.16)$$

3.2.1 Computational details

A problem that remains to be discussed is the determination of the smoothing parameter κ in (3.8), (3.12), and (3.13). Generally, it is possible to determine κ by the CV criterion in (3.14); however, for relatively large dimensions T and n cross validation is computationally very costly. Moreover, Kneip *et al.* (2012b) do not explain how the factor dimension d is to be specified during the optimization process, which is critical since \hat{d} itself depends on κ ; see (3.21) in Section 3.3.

We propose to determine the smoothing parameter κ by generalized cross validation (GCV). However, we cannot apply the classical GCV formulas as proposed, e.g., in Craven & Wahba (1978) since we do not know the parameters β and $v_i(t)$. Our computational algorithm for determining the GCV smoothing parameter κ_{GCV} is based on the method of Cao & Ramsay (2010), who propose optimizing objective functions of the form (3.7) by updating the parameters iteratively in a functional hierarchy. Formally, the iteration algorithm can be described as follows:

1. For given κ and β , we optimize (3.7) with respect to ζ_i to get

$$\hat{\zeta}_i = (Z'Z + \kappa R)^{-1} Z^\top (Y_i - X_i \beta). \quad (3.17)$$

2. By using (3.17), we minimize (3.7) with respect to β to get

$$\hat{\beta} = \left(\sum_{i=1}^N X_i^\top X_i \right)^{-1} \left(\sum_{i=1}^N X_i^\top (Y_i - Z \hat{\zeta}_i) \right) \quad (3.18)$$

3. Once (3.17) and (3.18) are obtained, we optimize the following GCV criterion to calculate κ_{GCV} :

$$\kappa_{GCV} = \arg \min_{\kappa} \frac{1}{\frac{n}{T} \text{tr}(I - \mathcal{Z}_\kappa)^2} \sum_{i=1}^n \|Y_i - X_i \hat{\beta} - \mathcal{Z}_\kappa (Y_i - X_i \hat{\beta})\|^2. \quad (3.19)$$

The program starts with initial estimates of β and κ and proceeds with steps 1, 2, and 3 in recurrence until convergence of all parameters, where the initial value $\hat{\beta}_{start}$ is defined in (3.45) and the initial value κ_{start} is the GCV-smoothing parameter of the residuals $Y_i - X_i \hat{\beta}_{start}$.

The advantage of this approach is that the inversion of the $P \times P$ matrix in (3.18) does not have to be updated during the iteration process. Moreover, the determination of the GCV-minimizer in (3.19) can be easily performed in R using the function `smooth.spline()`, which calls on a rapid C-routine.

The GCV smoothing parameter κ_{GCV} in (3.19) does not explicitly account for the factor structure of the time-varying individual effects $v_i(t)$ as formalized in (3.2). However, given the assumption of a factor structure, the goal is not to obtain optimal estimates of $v_i(t)$ but rather to obtain optimal estimates of the common factors $f_l(t)$, which implies that the optimal smoothing parameter κ_{opt} will be smaller than κ_{GCV} ; see Kneip *et al.* (2012b).

We use the GCV smoothing parameter κ_{GCV} as an upper bound for κ_{opt} , which provides a quick-and-dirty approximation. Alternatively, it is possible to optimize the CV criterion (3.14). In this case, the optimal smoothing parameter κ_{opt} is selected from the interval $(0, \kappa_{GCV})$ and the factor dimension d in (3.14) is estimated by (3.21) using the plug-in estimator $\kappa_{plug-in}$.

3.2.2 Application

This section is devoted to the application of the method above discussed, which is accessible through the function `KSS()`. In total, the function `KSS()` has the following arguments:

```
R> args(KSS)

function (formula, additive.effects = c("none", "individual",
    "time", "twoways"), consult.dim.crit = FALSE, d.max = NULL,
    sig2.hat = NULL, factor.dim = NULL, level = 0.01, spar = NULL,
    CV = FALSE, convergence = 1e-06, restrict.mode = c("restrict.factors",
    "restrict.loadings"), ...)
NULL
```

The argument `additive.effects` makes it possible to extend the model (3.4) for additional additive `individual`, `time`, or `twoways` effects as discussed in Section 3.5. If the logical argument `consult.dim.crit` is set to `TRUE` all dimensionality criteria discussed in Section 3.3 are computed and the user is asked to choose one of their results.

The arguments `d.max` and `sig2.hat` are required for the computation of some dimensionality criteria discussed in Section 3.3. If their default values are maintained, the function internally computes `d.max` = $\lfloor \min\{\sqrt{n}, \sqrt{T}\} \rfloor$ and `sig2.hat` as in (3.16), where $\lfloor x \rfloor$ indicates the integer part of x . The argument `level` allows us to adjust the significance level for the dimensionality testing procedure (3.21) of Kneip *et al.* (2012b); see Section 3.3.

The factor dimension d can also be pre-specified by the argument `factor.dim`. Recall from restriction (a) that $\frac{1}{T} \sum_{t=1}^T \hat{f}_l(t)^2 = 1$. Alternatively, it is possible to standardize the individual loadings parameters such that $\frac{1}{n} \sum_{i=1}^n \hat{\lambda}_{il} = 1$, which can be done by setting `restrict.mode = "restrict.loadings"`.

As an illustration we estimate the Cigarettes model (3.3) introduced in Section 3.1:

$$\begin{aligned} \ln(\text{Consumption}_{it}) &= \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + e_{it} \quad (3.20) \\ \text{with } e_{it} &= \sum_{l=1}^d \lambda_{il} f_l(t) + \epsilon_{it}, \end{aligned}$$

where $v_i(t) = \sum_{l=1}^d \lambda_{il} f_l(t)$. In the following lines of code we load the `Cigar` dataset and take logarithms of the three variables, `Consumptionit`, `Priceit/cpit` and `Incomeit/cpit`, where `cpit` is the consumer price index. Please note that we store the variables as $T \times n$ -matrices. This is necessary, because the `formula` argument of the `KSS()`-function takes the panel variables as matrices in which the number of rows has to be equal to the temporal dimension T and the number of columns has to be equal to the individual dimension n .

Note that the function `KSS()` is written for balanced panels, and eventually missing values have to be replaced in a pre-processing step by appropriate estimates.

```
R> library("phtt")
R> data("Cigar")
R> N <- 46
R> T <- 30
R> l.Consumption <- log(matrix(Cigar$sales, T, N))
R> cpi <- matrix(Cigar$cpi, T, N)
R> l.Price <- log(matrix(Cigar$price, T, N)/cpi)
R> l.Income <- log(matrix(Cigar$ndi, T, N)/cpi)
```

The model parameters β_1 , β_2 , the factors $f_l(t)$, the loadings parameters λ_{il} , and the factor dimension d can be estimated by the `KSS()`-function with its default arguments. Inferences about the slope parameters can be obtained by using the method `summary()`.

```
R> Cigar.KSS <- KSS(formula = l.Consumption ~ l.Price + l.Income)
R> (Cigar.KSS.summary <- summary(Cigar.KSS))
```

Call:

```
KSS.default(formula = l.Consumption ~ l.Price + l.Income)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.11	-0.01	0.00	0.01	0.12

Slope-Coefficients:

	Estimate	StdErr	z.value	Pr(>z)
(Intercept)	4.0600	0.1770	23.00	< 2.2e-16 ***
l.Price	-0.2600	0.0223	-11.70	< 2.2e-16 ***
l.Income	0.1550	0.0382	4.05	5.17e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type: none

Used Dimension of the Unobserved Factors: 6

Residual standard error: 0.000725 on 921 degrees of freedom

R-squared: 0.99

The effects of the log-real prices for cigarettes $\ln(\text{Price}_{it})$ and the log-real incomes $\ln(\text{Income}_{it})$ on the log-sales of cigarettes $\ln(\text{Consumption}_{it})$ are highly significant and in line with results in the literature. The summary output reports an estimated factor dimension of $\hat{d} = 6$. In order to get a visual impression of the six estimated common factors $\hat{f}_1(t), \dots, \hat{f}_6(t)$ and the estimated time-varying individual effects $\hat{v}_1(t), \dots, \hat{v}_n(t)$, we provide a `plot()`-method for the `KSS`-summary object.

```
R> plot(Cigar.KSS.summary)
```

The left panel of Figure 3.2 shows the six estimated common factors $\hat{f}_1(t), \dots, \hat{f}_6(t)$ and the right panel of Figure 3.2 shows the $n = 46$ estimated time-varying individ-

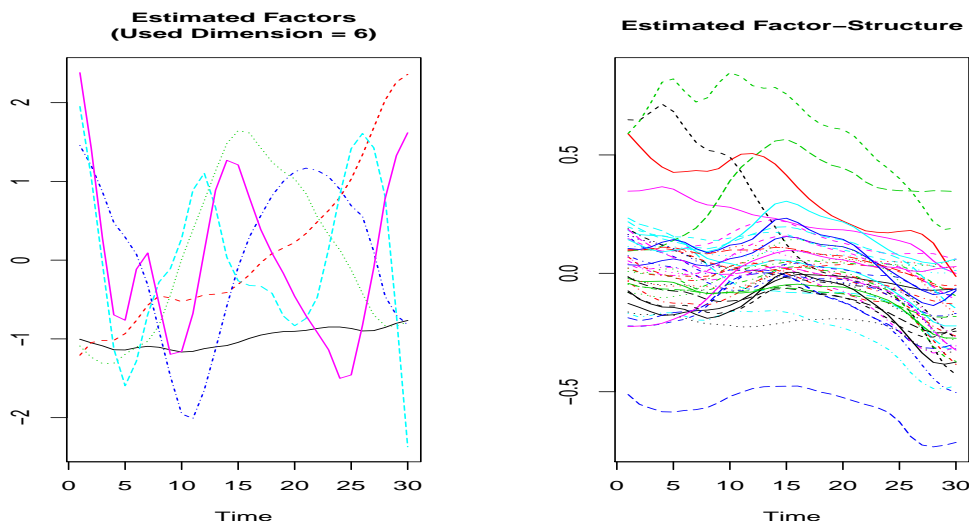


Figure 3.2: LEFT PANEL: Estimated factors $\hat{f}_1(t), \dots, \hat{f}_6(t)$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_1(t), \dots, \hat{v}_n(t)$.

ual effects $\hat{v}_1(t), \dots, \hat{v}_n(t)$. Obviously, there is one nearly time-invariant common factor; this suggests extending the model (3.20) by additive **individual** effects; see Section 3.5 for more details.

By setting the logical argument `consult.dim.crit=TRUE`, the user can choose from other dimensionality criteria, which are discussed in Section 3.3. Note that the consideration of different factor dimensions d would not alter the results for the slope parameters β since the estimation procedure of Kneip *et al.* (2012b) for the slope parameters β does not depend on the dimensionality parameter d .

3.3 Panel criteria for selecting the number of factors

In order to estimate the factor dimension d , Kneip *et al.* (2012b) propose a sequential testing procedure based on the following test statistic:

$$KSS(d) = \frac{n \sum_{r=d+1}^T \hat{\rho}_r - (n-1) \hat{\sigma}^2 \text{tr}(\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)}{\hat{\sigma}^2 \sqrt{2N \cdot \text{tr}((\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)^2)}} \stackrel{a}{\sim} N(0, 1), \quad (3.21)$$

where $\hat{\mathcal{P}}_d = I - \frac{1}{T} \sum_{l=1}^d f_l f_l^\top$ with $f_l = (f_l(1), \dots, f_l(T))^\top$, and

$$\hat{\sigma}^2 = \frac{1}{(n-1)tr((I - \mathcal{Z}_\kappa)^2)} \sum_{i=1}^n \|(I - \mathcal{Z}_\kappa)(Y_i - X_i \hat{\beta})\|^2. \quad (3.22)$$

The selection method can be described as follows: choose a significance level α (e.g., $\alpha = 1\%$) and begin with $H_0 : d = 0$. Test if $KSS(0) \leq z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of the standard normal distribution. If the null hypothesis can be rejected, go on with $d = 1, 2, 3, \dots$ until H_0 cannot be rejected.

The dimensionality criterion of Kneip *et al.* (2012b) can be used for stationary as well as non-stationary factors. However, this selection procedure has a tendency to ignore factors that are weakly auto-correlated. As a result, the number of factors can be underestimated.

More robust against this kind of underestimation are the criteria of Bai & Ng (2002). The basic idea of their approach consists simply of finding a suitable penalty term g_{nT} , which countersteers the undesired variance reduction caused by an increasing number of factors \hat{d} . Formally, \hat{d} can be obtained by minimizing the following criterion:

$$PC(l) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(l))^2 + l g_{nT} \quad (3.23)$$

for all $l \in \{1, 2, \dots\}$, where $\hat{y}_{it}(l)$ is the fitted variable for a given factor dimension l . To estimate consistently the dimension of stationary factors Bai & Ng (2002) propose specifying g_{nT} by one of the following penalty terms:

$$g_{nT}^{(PC1)} = \hat{\sigma}^2 \frac{(n+T)}{nT} \log\left(\frac{nT}{n+T}\right) \quad (3.24)$$

$$g_{nT}^{(PC2)} = \hat{\sigma}^2 \frac{(n+T)}{nT} \log(\min\{n, T\}) \quad (3.25)$$

$$g_{nT}^{(PC3)} = \hat{\sigma}^2 \frac{\log(\min\{n, T\})}{\min\{n, T\}}, \quad (3.26)$$

where $\hat{\sigma}^2$ is a consistent estimator of σ^2 , the variance of ϵ_{it} . The proposed criteria are denoted by PC1, PC2, and PC3, respectively. In practice, $\hat{\sigma}^2$ can be obtained

by

$$\hat{\sigma}^2(d_{max}) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(d_{max}))^2, \quad (3.27)$$

where d_{max} is an arbitrary maximal dimension that is larger than d . This kind of variance estimation can, however, be inappropriate in some cases, especially when (3.27) underestimates the true variance σ^2 . The latter can be the case, if the error terms are auto-correlated. To overcome this problem, Bai & Ng (2002) propose three additional criteria (IC1, IC2, and IC3):

$$IC(l) = \log \left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(l))^2 \right) + lg_{nT} \quad (3.28)$$

with

$$g_{nT}^{(IC1)} = \frac{(n+T)}{nT} \log \left(\frac{nT}{n+T} \right) \quad (3.29)$$

$$g_{nT}^{(IC2)} = \frac{(n+T)}{nT} \log(\min\{n, T\}) \quad (3.30)$$

$$g_{nT}^{(IC3)} = \frac{\log(\min\{n, T\})}{\min\{n, T\}}. \quad (3.31)$$

Under similar assumptions, Ahn & Horenstein (2013) propose selecting d by maximizing the ratio of adjacent eigenvalues (or the ratio of their growth rate). The criteria are referred to as *Eigenvalue Ratio* (ER) and *Growth Ratio* (GR) and defined as following:

$$ER = \frac{\hat{\rho}_l}{\hat{\rho}_{l+1}}$$

$$GR = \frac{\log(\sum_{r=l}^T \hat{\rho}_r / \sum_{r=l+1}^T \hat{\rho}_r)}{\log(\sum_{r=l+1}^T \hat{\rho}_r / \sum_{r=l+2}^T \hat{\rho}_r)}.$$

Note that the theory of the above dimensionality criteria PC1, PC2, PC3, IC1, IC2, IC3, ER, and GR is developed for stationary factors. In order to estimate the number of unit root factors, Bai (2004) proposes the following panel criteria:

$$IPC(l) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(l))^2 + lg_{nT} \quad (3.32)$$

where

$$g_{nT}^{(\text{IPC1})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T} \frac{(n+T)}{nT} \log\left(\frac{nT}{n+T}\right) \quad (3.33)$$

$$g_{nT}^{(\text{IPC2})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T} \frac{(n+T)}{nT} \log(\min\{n, T\}) \quad (3.34)$$

$$g_{nT}^{(\text{IPC3})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T} \frac{\log(\min\{n, T\})}{\min\{n, T\}}. \quad (3.35)$$

Alternatively, Onatski (2010) has introduced a threshold approach based on the empirical distribution of the sample covariance eigenvalues, which can be used for both stationary and non-stationary factors. The estimated dimension is obtained by

$$\hat{d} = \max\{l \leq d_{max} : \hat{\rho}_l - \hat{\rho}_{l-1} \geq \delta\},$$

where δ is a positive threshold, estimated iteratively from the data.

3.3.1 Application

The dimensionality criteria introduced above are implemented in the function `OptDim()`, which has the following arguments:

```
R> args(OptDim)
```

```
function (Obj, criteria = c("PC1", "PC2", "PC3", "IC1", "IC2",
  "IC3", "IPC1", "IPC2", "IPC3", "KSS.C", "ED", "ER", "GR"),
  standardize = FALSE, d.max, sig2.hat, spar, level = 0.01)
NULL
```

The desired criteria can be selected by one or several of the following character variables: "KSS.C", "PC1", "PC2", "PC3", "IC1", "IC2", "IC3", "ER", "GR", "IPC1", "IPC2", "IPC3", and "ED". The default significance level used for the "KSS"-criterion is `level = 0.01`. The values of d_{max} and $\hat{\sigma}^2$ can be specified externally by the arguments `d.max` and `sig2.hat`. By default, `d.max` is computed internally as `d.max = floor(min{sqrt(n), sqrt(T)})` and `sig2.hat` as in (3.22) and (3.27). The input variable can be standardized by choosing `standardize = TRUE`. In this

case, the calculation of the eigenvalues is based on the correlation matrix instead of the covariance matrix.

As an illustration, imagine that we are interested in the estimation of the factor dimension of the variable $\ln(\text{Consumption}_{it})$ with the dimensionality criterion "PC1". The function `OptDim()` requires a $T \times n$ matrix as input variable.

```
R> OptDim(Obj = l.Consumption, criteria = "PC1")
```

```
Call: OptDim.default(Obj = l.Consumption, criteria = "PC1")
```

```
-----
```

```
Criterion of Bai and Ng (2002):
```

```
PC1
```

```
7
```

`OptDim()` offers the possibility of comparing the result of different selection procedures by giving the corresponding criteria to the argument `criteria`. If the argument `criteria` is left unspecified, `OptDim()` automatically compares all 13 dimensionality selection procedures.

```
R> (OptDim.obj <- OptDim(Obj = l.Consumption, criteria = c("PC3", "ER", "GR",
+               "IPC1", "IPC2", "IPC3"), standardize = TRUE))
```

```
Call: OptDim.default(Obj = l.Consumption, criteria = c("PC3", "ER",
+               "GR", "IPC1", "IPC2", "IPC3"), standardize = TRUE)
```

```
-----
```

```
Criterion of Bai and Ng (2002):
```

```
PC3
```

```
8
```

```
-----
```


Criteria of Ahn and Horenstein (2008):

```
ER GR
  3  3
```

Criteria of Bai (2004):

```
IPC1 IPC2 IPC3
  3    3    2
```

In order to help users to choose the most appropriate dimensionality criterion for the data, `OptDim`-objects are provided with a `plot()`-method. This method displays, in descending order, the magnitude of the eigenvalues in percentage of the total variance and indicates where the given criteria detect the optimal dimension: see Figure 3.3.

```
R> plot(OptDim.obj)
```

In this regard, the function `KSS()` offers us the ability to compare the results of all dimensionality criteria and to select one of them. If the `KSS()`-argument `consult.dim = TRUE` the results of the dimensionality criteria are printed on the console of R and the user is asked to choose one of the results.

```
R> KSS(formula = l.Consumption ~ -1 + l.Price + l.Income, consult.dim = TRUE)
```

Results of Dimension-Estimations

-Bai:

```
PC1  PC2  PC3  IC1  IC2  IC3  IPC3  IPC2  IPC3
  7    6    7    5    5    5    3    3    3
```

-KSS:

```
KSS.C
```

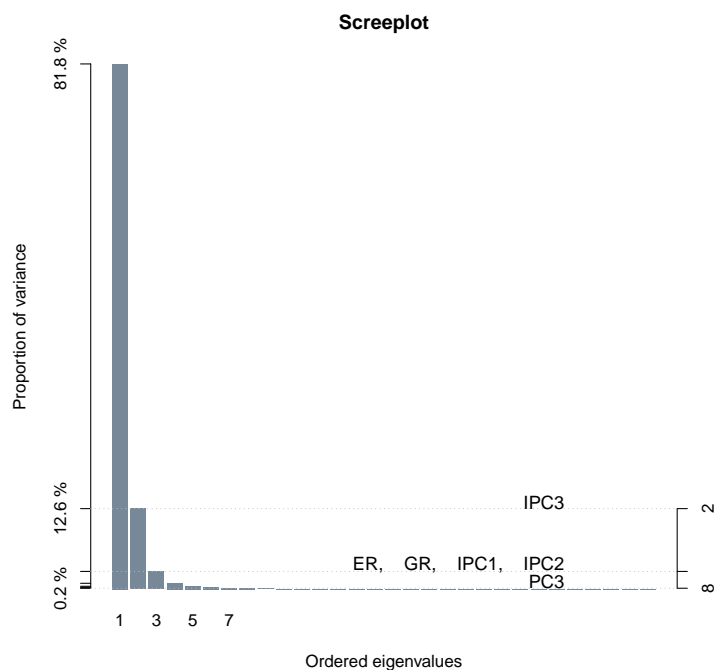


Figure 3.3: Scree plot produced by the `plot()`-method for `OptDim`-objects. Most of the dimensionality criteria (ER, GR, IPC1 and IPC2) suggest using the dimension $\hat{d} = 3$.

6

-Onatski:

ED

3

-RH:

ER GR

3 4

Please, choose one of the proposed integers:

After entering a number of factors, e.g., 6 we get the following feedback:

Used dimension of unobs. factor structure is: 6

Note that the maximum number of factors that can be given, cannot exceed the maximal number of all proposed factor dimensions (here maximal dimension would be 7). A higher dimension can be chosen using the argument `factor.dim`.

3.4 Panel models with stationary common factors

The panel model proposed by Bai (2009) can be presented as follows:

$$y_{it} = \sum_{j=1}^P x_{itj} \beta_j + v_{it} + \epsilon_{it}, \quad (3.36)$$

where

$$v_{it} = \sum_{l=1}^d \lambda_{il} f_{lt}. \quad (3.37)$$

Combining (3.36) with (3.37) and writing the model in matrix notation, we get

$$Y_i = X_i \beta + F \Lambda_i^\top + \epsilon_i \quad (3.38)$$

where $Y_i = (y_{i1}, \dots, y_{iT})^\top$, $X_i = (x_{i1}^\top, \dots, x_{iT}^\top)^\top$, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})^\top$, $\Lambda_i = (\lambda_1, \dots, \lambda_n)^\top$ and $F = (f_1, \dots, f_T)^\top$ with $\lambda_i = (\lambda_{i1}, \dots, \lambda_{id})$, $f_t = (f_{t1}, \dots, f_{dt})$, and $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})^\top$.

The asymptotic properties of Bai's method rely, among others, on the following assumption:

$$\frac{1}{T} F^\top F \xrightarrow{p} \Sigma_F \text{ as } T \rightarrow \infty, \quad (3.39)$$

where Σ_F is a fixed positive definite $d \times d$ matrix. This rules out a large class of non-stationary stochastic processes such as unit root processes.

3.4.1 Model with known number of factors d

Bai (2009) proposes to estimate the model parameters β , F and Λ_i by minimizing

the following least squares objective function:

$$S(\beta, F, \Lambda_i) = \sum_i^n \|Y_i - X_i\beta - F\Lambda_i^\top\|^2 \quad (3.40)$$

For each given F , the OLS estimator of β can be obtained by

$$\hat{\beta}(F) = \left(\sum_{i=1}^n X_i^\top \mathcal{P}_d X_i \right)^{-1} \left(\sum_{i=1}^n X_i^\top \mathcal{P}_d Y_i \right)$$

where $\mathcal{P}_d = I - F(F^\top F)^{-1}F^\top = I - FF^\top/T$. If β is known, F can be estimated by using the first d eigenvectors $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_d)$ corresponding to the first d eigenvalues of the empirical covariance matrix $\hat{\Sigma} = (nT)^{-1} \sum_{i=1}^n w_i w_i^\top$, where $w_i = Y_i - X_i\beta$. That is,

$$\hat{F}(\beta) = \sqrt{T}\hat{\gamma}.$$

The idea of Bai (2009) is to start with initial values for β or F and calculate the estimators iteratively. The method requires, however, the factor dimension d to be known, which is usually not the case in empirical applications.

3.4.2 Model with unknown number of factors d

Bada & Kneip (2010) propose an algorithmic refinement of the method of Bai (2009) in order to provide a joint estimation of the factor dimension d together with the other parameters β , F , and Λ_i . In this case, the optimization criterion can be defined as a penalized least squares objective function:

$$S(\beta, F, \Lambda_i, l) = \sum_i^N \|Y_i - X_i\beta - F\Lambda_i^\top\|^2 + lg_{nT} \quad (3.41)$$

The role of the additional term lg_{nT} is to pick up the optimal dimension \hat{d} , of the unobserved factor structure. The penalty factor g_{nT} can be chosen according to Bai & Ng (2002). Alternatively, g_{nT} can be replaced by the threshold δ proposed by Onatski (2010). The estimation algorithm is based on the parameter cascading method of Cao & Ramsay (2010) that can be described as follows:

1. Minimizing (3.41) with respect to Λ_i for each given β, F and d , we get

$$\hat{\Lambda}_i^\top(\beta, F, d) = F^\top (Y_i - X_i\beta) / T. \quad (3.42)$$

2. Introducing (3.42) in (3.41) and minimizing with respect to F for each given β and d , we get

$$\hat{F}(\beta, d) = \sqrt{T}\hat{\gamma}(\beta, d), \quad (3.43)$$

where $\hat{\gamma}(\beta, d)$ is a $T \times d$ matrix that contains the first d eigenvectors corresponding to the first d eigenvalues ρ_1, \dots, ρ_d of the covariance matrix $\hat{\Sigma} = (nT)^{-1} \sum_{i=1}^n w_i w_i^\top$ with $w_i = Y_i - X_i\beta$.

3. Reintegrating (3.43) and (3.42) in (3.41) and minimizing with respect to β for each given d , we get

$$\hat{\beta}(d) = \left(\sum_{i=1}^N X_i^\top X_i \right)^{-1} \left(\sum_{i=1}^N X_i^\top (Y_i - \hat{F}\hat{\Lambda}_i^\top(\hat{\beta}, d)) \right). \quad (3.44)$$

4. Optimizing (3.41) with respect to l given the results in (3.42), (3.43), and (3.44) allows us to select \hat{d} as

$$\hat{d} = \operatorname{argmin}_l \sum_i^N \|Y_i - X_i\hat{\beta} - \hat{F}\hat{\Lambda}_i^\top\|^2 + lg_{nT}, \quad \text{for all } l \in \{0, 1, \dots, d_{max}\}.$$

The final estimators are obtained by alternating between an inner iteration to optimize $\hat{\beta}(d)$, $\hat{F}(d)$, and $\hat{\Lambda}_i(d)$ for each given d and an outer iteration to select the optimal dimension \hat{d} . The updating process is repeated in its entirety till the convergence of all the parameters. This is why the estimators are called *entirely updated estimators* (Eup).

It is notable that the objective functions (3.41) and (3.40) are not globally convex. There is no guarantee that the iteration algorithm converges to the global optimum. Therefore, it is important to choose reasonable starting values \hat{d}_{start} and $\hat{\beta}_{start}$. We propose to select a large dimension d_{max} and to start the iteration

with the following estimate of β :

$$\hat{\beta}_{start} = \left(\sum_{i=1}^N X_i^\top (I - GG^\top) X_i \right)^{-1} \left(\sum_{i=1}^N X_i^\top (I - GG^\top) Y_i \right), \quad (3.45)$$

where G is the $T \times d_{max}$ matrix of the eigenvectors corresponding to the first d_{max} eigenvalues of the augmented covariance matrix

$$\Gamma^{Aug} = \frac{1}{nT} \sum_{i=1}^n (Y_i, X_i)(Y_i^\top, X_i^\top)^\top.$$

The intuition behind these starting estimates relies on the fact that the unobserved factors cannot escape from the space spanned by the eigenvectors G . The projection of X_i on the orthogonal complement of G in (3.45) eliminates the effect of a possible correlation between the observed regressors and unobserved factors, which can heavily distort the value of β^0 if it is neglected. Greenaway-McGrevy *et al.* (2012) give conditions under which (3.45) is a consistent estimator of β .

According to Bai (2009), the asymptotic distribution of the slope estimator $\hat{\beta}$ for known d and i.i.d. ϵ_{it} is:

$$\sqrt{nT}(\hat{\beta}(d) - \beta) \overset{a}{\sim} N(0, \Sigma_\beta),$$

where $\Sigma_\beta = D_0^{-1}\sigma^2$. Here, $\sigma^2 = \text{Var}(\epsilon_{it})$ and $D_0 = \text{plim}_{\frac{1}{nT}} \sum_{i=1}^n Z_i^\top Z_i$ with $Z_i = \mathcal{P}_d X_i - \frac{1}{n} \sum_{k=1}^n \mathcal{P}_d X_i a_{ik}$ and $a_{ik} = \Lambda_i (\frac{1}{n} \sum_{i=1}^n \Lambda_i^\top \Lambda_i)^{-1} \Lambda_k^\top$. Bada & Kneip (2010) show that if $\lim_{n,T \rightarrow \infty} P(\hat{d} = d) = 1$, the entirely updated estimator $\hat{\beta} = \hat{\beta}(\hat{d})$ will have the same asymptotic distribution as $\hat{\beta}(d)$. The asymptotic variance of the estimator $\hat{\beta}$ can be estimated as follows:

$$\hat{\Sigma}_\beta = \hat{\varsigma}^2 \left(\frac{1}{nT} \sum_{i=1}^n Z_i^\top Z_i \right)^{-1},$$

where $\hat{\varsigma}^2 = \frac{1}{nT} \sum_{i=1}^n \hat{\epsilon}_i^\top \hat{\epsilon}_i$ and $\hat{\epsilon}_i = Y_i - X_i \hat{\beta} - \hat{F} \hat{\Lambda}_i^\top$.

3.4.3 Application

The above described methods are implemented in the function `Eup()`, which takes the following arguments:

```
R> args(Eup)
```

```
function (formula, additive.effects = c("none", "individual",
    "time", "twoways"), dim.criterion = c("PC1", "PC2", "PC3",
    "IC1", "IC2", "IC3", "IPC1", "IPC2", "IPC3", "ED"), d.max = NULL,
    sig2.hat = NULL, factor.dim = NULL, double.iteration = TRUE,
    start.beta = NULL, max.iteration = 500, convergence = 1e-06,
    restrict.mode = c("restrict.factors", "restrict.loadings"),
    ...)
```

```
NULL
```

The argument `additive.effects` gives the possibility of extending the model (3.38) for additional additive effects as discussed in more detail in Section 3.5. The argument `dim.criterion` specifies the dimensionality criterion to be used if `factor.dim` is left unspecified and defaults to `dim.criterion = "PC1"`. The arguments `d.max` and `sig2.hat` are required for the computation of some dimensionality criteria discussed in Section 3.3. If their default values are maintained, the function internally computes `d.max` as $\min\{\sqrt{n}, \sqrt{T}\}$ and `sig2.hat` according to (3.27).

Setting the argument `double.iteration=FALSE` may speed up computations, because the updates of \hat{d} will be done simultaneously with \hat{F} without waiting for their inner convergences. However, in this case, the convergence of the parameters is less stable than in the default setting.

The argument `start.beta` allows us to give a vector of starting values for the slope parameters β_{start} . The maximal number of iteration and the convergence condition can be controlled by `max.iteration` and `convergence`. Finally, by choosing `restrict.mode = c("restrict.loadings")`, the restriction $\frac{1}{T} \sum_t f_{it}^2 = 1$ will be replaced by the restriction $\frac{1}{n} \sum_i \lambda_{il}^2 = 1$ for all $l \in \{1, \dots, d\}$.

In our application, we take first-order differences of the observed time series. This is because some factors show temporal trends, which can violate the sta-

tionarity condition (3.39); see Figure 3.2. We consider the following modified Cigarettes model:

$$\begin{aligned}\nabla \ln(\text{Consumption}_{it}) &= \beta_1 \nabla \ln(\text{Price}_{it}) + \beta_2 \nabla \ln(\text{Income}_{it}) + e_{it}, \\ \text{with } e_{it} &= \sum_{l=1}^d \lambda_{il} f_{lt} + \epsilon_{it},\end{aligned}$$

where $\nabla x_t = x_t - x_{t-1}$. In order to avoid notational mess, we use the same notation for the unobserved time-varying individual effects $v_{it} = \sum_{l=1}^d \lambda_{il} f_{lt}$ as above in (3.20). The ∇ -transformation can be easily performed in R using the standard `diff()`-function as follows:

```
R> d.l.Consumption <- diff(l.Consumption)
R> d.l.Price       <- diff(l.Price)
R> d.l.Income      <- diff(l.Income)
```

As previously mentioned for the `KSS()`-function, the `formula` argument of the `Eup()`-function takes balanced panel variables as $T \times n$ dimensional matrices, where the number of rows has to be equal to the temporal dimension T and the number of columns has to be equal to the individual dimension n .

```
R> (Cigar.Eup <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+                   dim.criterion = "PC3"))
```

Call:

```
Eup.default(formula = d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
             dim.criterion = "PC3")
```

Coeff(s) of the Observed Regressor(s) :

```
      d.l.Price d.l.Income
-0.3171044    0.1838808
```

Additive Effects Type: none

Dimension of the Unobserved Factors: 7

Number of iterations: 115

Inferences about the slope parameters can be obtained by using the method `summary()`.

```
R> summary(Cigar.Eup)
```

Call:

```
Eup.default(formula = d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
  dim.criterion = "PC3")
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.09340	-0.01170	0.00063	0.01260	0.07690

Slope-Coefficients:

	Estimate	Std.Err	Z value	Pr(>z)
d.l.Price	-0.3170	0.0237	-13.40	< 2.2e-16 ***
d.l.Income	0.1840	0.0372	4.95	7.48e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type: none

Dimension of the Unobserved Factors: 7

Residual standard error: 0.0006995 on 807 degrees of freedom,

R-squared: 0.78

The summary output reports that "PC3" detects 7 common factors. The effect of the log-real prices for cigarettes on the log-sales is negative and amounts to -0.317104 . The estimated effect of the real disposable log-income per capita is

0.183882, which is smaller than the effect estimated by the method of Kneip *et al.* (2012b).

The estimated factors \hat{f}_{it} as well as the individual effects \hat{v}_{it} can be plotted using the `plot()`-method for `summary.Eup`-objects. The corresponding graphics are shown in Figure 3.4.

```
R> plot(summary(Cigar.Eup))
```

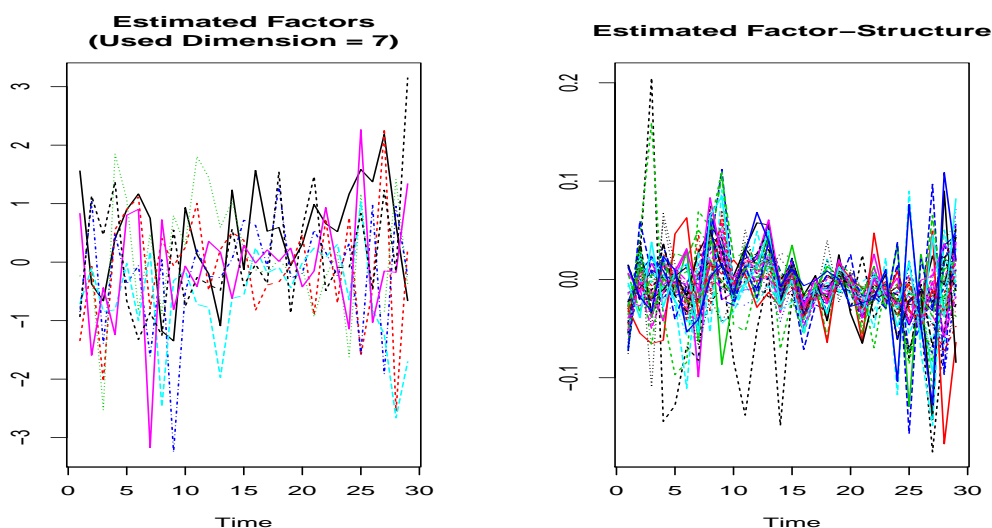


Figure 3.4: LEFT PANEL: Estimated factors $\hat{f}_{1t}, \dots, \hat{f}_{7t}$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_{1t}, \dots, \hat{v}_{nt}$.

3.5 Models with additive and interactive unobserved effects

Even though the classical additive "individual", "time", and "twoways" effects can be absorbed by the factor structure, there are good reasons to model them explicitly. On the one hand, if there are such effects in the true model, then neglecting them will result in non-efficient estimators; see Bai *et al.* (2009). On the other hand, additive effects can be very useful for interpretation purposes.

Consider now the following model:

$$y_{it} = \mu + \alpha_i + \theta_t + x_{it}^\top \beta + \nu_{it} + \epsilon_{it} \quad (3.46)$$

with

$$\nu_{it} = \begin{cases} v_{it} & = \sum_{l=1}^d \lambda_{il} f_{lt}, & \text{for the model of Bai (2009),} \\ v_i(t) & = \sum_{l=1}^d \lambda_{il} f_l(t), & \text{for the model of Kneip *et al.* (2012b),} \end{cases}$$

where α_i are time-constant individual effects and θ_t is a common time-varying effect.

In order to ensure identification of the additional additive effects α_i and θ_t , we need the following further restrictions:

$$(d) \sum_{i=1}^n \lambda_{il} = 0 \quad \text{for all } l \in \{1, \dots, d\}$$

$$(e) \sum_{t=1}^T f_{lt} = 0 \quad \text{for all } l \in \{1, \dots, d\}$$

$$(f) \sum_{i=1}^n \alpha_i = 0$$

$$(g) \sum_{t=1}^T \theta_t = 0$$

By using the classical within-transformations on the observed variables, we can eliminate the additive effects α_i and θ_t , such that

$$\dot{y}_{it} = \dot{x}_{it}^\top \beta + \nu_{it} + \dot{\epsilon}_{it},$$

where $\dot{y}_{it} = y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it} - \frac{1}{n} \sum_{i=1}^n y_{it} + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n y_{it}$, $\dot{x}_{it} = x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} - \frac{1}{n} \sum_{i=1}^n x_{it} + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n x_{it}$, and $\dot{\epsilon}_{it} = \epsilon_{it} - \frac{1}{T} \sum_{t=1}^T \epsilon_{it} - \frac{1}{n} \sum_{i=1}^n \epsilon_{it} + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n \epsilon_{it}$. Note that restrictions (d) and (e) insure that the transformation does not affect the time-varying individual effects ν_{it} .

The parameters β and ν_{it} can be estimated by the above introduced estimation procedures. All possible variants of (3.46) are implemented in the functions `KSS()` and `Eup()`. The appropriate model can be specified by the argument

```
additive.effects = c("none", "individual", "time", "twoways"):
```

$$\begin{array}{ll}
 \text{"none"} & y_{it} = \mu + x_{it}^{\top} \beta + \nu_{it} + \epsilon_{it} \\
 \text{"individual"} & y_{it} = \mu + \alpha_i + x_{it}^{\top} \beta + \nu_{it} + \epsilon_{it} \\
 \text{"time"} & y_{it} = \mu + \theta_t + x_{it}^{\top} \beta + \nu_{it} + \epsilon_{it} \\
 \text{"twoways"} & y_{it} = \mu + \alpha_i + \theta_t + x_{it}^{\top} \beta + \nu_{it} + \epsilon_{it}.
 \end{array}$$

The presence of μ can be controlled by `-1` in the `formula`-object: a formula with `-1` refers to a model without intercept. However, for identification purposes, if a `twoways` model is specified, the presence `-1` in the `formula` will be ignored.

As an illustration we continue with the application of the `KSS()`-function in Section 3.2. The left panel of Figure 3.2 shows that one of the six estimated factors is nearly time-invariant. This motivates us to augment the model (3.20) for a time-constant additive effects α_i . In this case it is convenient to use an intercept μ , which yields the following model:

$$\begin{aligned}
 \ln(\text{Consumption}_{it}) &= \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + \alpha_i + v_i(t) \quad (3.47) \\
 \text{where } v_i(t) &= \sum_{l=1}^d \lambda_{il} f_l(t).
 \end{aligned}$$

The estimation of the augmented model (3.47) can be done using the following lines of code.

```
R> Cigar2.KSS <- KSS(formula = l.Consumption ~ l.Price + l.Income,
+                   additive.effects = "individual")
R> (Cigar2.KSS.summary <- summary(Cigar2.KSS))
```

Call:

```
KSS.default(formula = l.Consumption ~ l.Price + l.Income, additive.effects = "in
```

Residuals:

Min	1Q	Median	3Q	Max
-0.11	-0.01	0.00	0.01	0.12

```

Slope-Coefficients:
              Estimate StdErr z.value   Pr(>z)
(Intercept)   4.0500  0.1760   23.10 < 2.2e-16 ***
1.Price       -0.2600  0.0222  -11.70 < 2.2e-16 ***
1.Income       0.1570  0.0381    4.11  3.88e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
Additive Effects Type: individual
```

```
Used Dimension of the Unobserved Factors: 5
```

```
Residual standard error: 0.000734 on 951 degrees of freedom
```

```
R-squared: 0.99
```

Again, the `plot()` method provides an useful visualization of the results.

```
R> plot(Cigar2.KSS.summary)
```

The "individual"-transformation of the data does not affect the estimation of the slope parameters but reduces the estimated dimension from $\hat{d} = 6$ to $\hat{d} = 5$. The remaining five common factors $\hat{f}_1, \dots, \hat{f}_5$ correspond to those of model (3.20); see the middle panel of Figure 3.5. The estimated time-constant state-specific effects α_i are shown in the left plot of Figure 3.5. The extraction of the α_i 's from the factor structure yields a denser set of time-varying individual effects \hat{v}_i shown in the right panel of Figure 3.5.

3.5.1 Specification tests

Model specification is an important step for any empirical analysis. The `phtt` package is equipped with two types of specification tests: the first is a Hausman-type test appropriate for the model of Bai *et al.* (2009); see Section 3.5.1.1. The second one examines the existence of a factor structure in Bai's model as well as in the model of Kneip *et al.* (2012b); see Section 3.5.1.2.

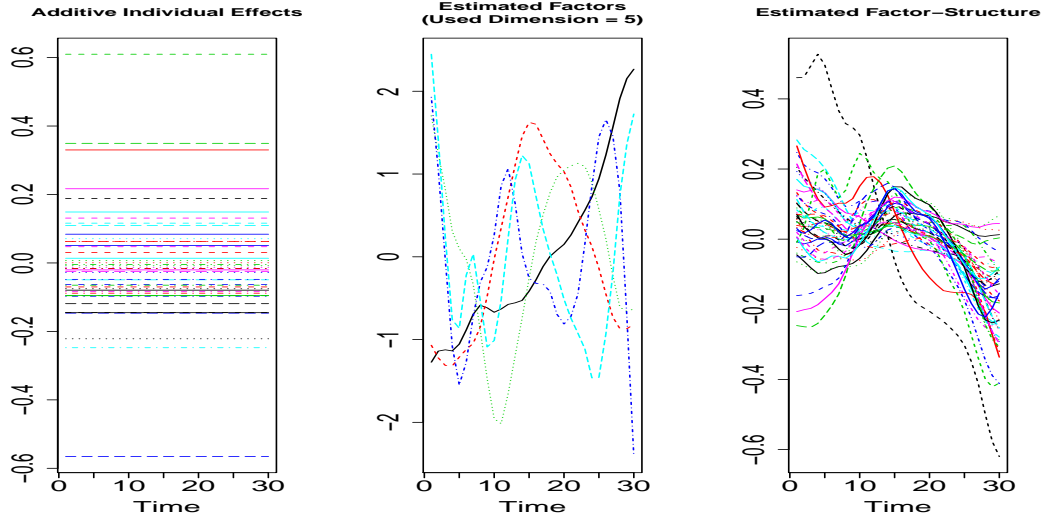


Figure 3.5: LEFT PANEL: Estimated time-constant state-specific effects $\hat{\alpha}_1, \dots, \hat{\alpha}_n$. MIDDLE PANEL: Estimated common factors $\hat{f}_1(t), \dots, \hat{f}_5(t)$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_1(t), \dots, \hat{v}_n(t)$.

3.5.1.1 Testing the sufficiency of classical additive effects

For the case in which there are only one or two factors ($1 \leq d \leq 2$), it is interesting to check whether or not these factors can be interpreted as classical "individual", "time", or "twoways" effects. Bai *et al.* (2009) considers the following testing problem:

$$\begin{aligned} H_0: & \quad v_{it} = \alpha_i + \theta_t \\ H_1: & \quad v_{it} = \sum_{l=1}^2 \lambda_{il} f_{lt} \end{aligned}$$

The model with factor structure, as described in Section 3.4, is consistent under both hypotheses. However, it is less efficient under H_0 than the classical within estimator, while the latter is inconsistent under H_1 if x_{it} and v_{it} are correlated. These conditions are favorable for applying the Hausman test:

$$J_{Bai} = nT \left(\hat{\beta} - \hat{\beta}_{within} \right) \Delta^{-1} \left(\hat{\beta} - \hat{\beta}_{within} \right) \stackrel{a}{\sim} \chi_P^2, \quad (3.48)$$

where $\hat{\beta}_{within}$ is the classical within least squares estimator, $\Delta = \text{Var} \left(\hat{\beta} - \hat{\beta}_{within} \right)$, P is the vector-dimension of β , and χ_P^2 is the χ^2 -distribution with P degree of freedom.

The null hypothesis H_0 can be rejected, if $J_{Bai} > \chi_{P,1-\alpha}^2$, where $\chi_{P,1-\alpha}^2$ is the $(1 - \alpha)$ -quantile of the χ^2 distribution with P degrees of freedom.

To calculate J_{Bai} we can replace Δ by its consistent estimator

$$\hat{\Delta} = \left(\left(\frac{1}{nT} \sum_{i=1}^n Z_i^\top Z_i \right)^{-1} - \left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \dot{x}_i^\top \dot{x}_i \right)^{-1} \right) \hat{\sigma}^2, \quad (3.49)$$

where

$$\hat{\sigma}^2 = \frac{1}{nT - (n + T)\hat{d} - P + 1} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - x_{it}^\top \hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{il} \hat{f}_{lt})^2. \quad (3.50)$$

The test is implemented in the function `checkSpecif()`, which takes the following arguments:

```
R> checkSpecif(obj1, obj2, level = 0.05)
```

The arguments `obj1` and `obj2` take both objects of class `Eup` produced by the function `Eup()`:

`obj1` Takes an `Eup`-object from an estimation with "individual", "time", or "twoways" effects and a factor dimension equal to $d = 0$; specified as `factor.dim = 0`.

`obj2` Takes an `Eup`-object from an estimation with "none"-effects and a positive factor dimension $1 \leq d \leq 2$:

`factor.dim=1` for testing "individual" or "time" effects.

`factor.dim=2` for testing "twoways" effects.

The argument `level` is used to specify the significance level.

However, the Hausman test of Bai *et al.* (2009) has a clear disadvantage. It is applicable only in situations of one or two factors ($1 \leq d \leq 2$). This is, e.g., not fulfilled in our demonstrations using the `Cigar` dataset, where the estimated factor dimension \hat{d} lies between six and seven; see Figures 3.2 and 3.4. The problem is that in such cases the matrix $\hat{\Delta}$ in (3.49) can become negative definite, which would yield a negative test statistic J_{Bai} in (3.48). If the test is applied in such situations, an error message is printed:

```
R> twoways.obj      <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+                          factor.dim = 0, additive.effects = "twoways")
R> not.twoways.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+                          factor.dim = 2, additive.effects = "none")
R> checkSpecif(obj1 = twoways.obj, obj2 = not.twoways.obj, level = 0.01)
```

```
Error in checkSpecif(obj1 = twoways.obj, obj2 = not.twoways.obj, level = 0.01)
```

```
The assumptions of the test are not fulfilled.
```

```
The (unobserved) true number of factors is probably greater than 2.
```

An alternative test for the sufficiency of a classical additive effects model is given by the following test proposed by Kneip *et al.* (2012b). This test can be applied for arbitrary factor dimensions d .

3.5.1.2 Testing the existence of common factors

This section is concerned with testing the existence of common factors. In contrast to the Hausman type statistic discussed above, the goal of this test is not merely to decide which model specification is more appropriate for the data, but rather to test in general the existence of common factors beyond the possible presence of additional classical "individual", "time", or "twoways" effects in the model.

This test relies on using the dimensionality criterion proposed by Kneip *et al.* (2012b) to test the following hypothesis after eliminating eventual additive "individual", "time", or "twoways" effects:

$$H_0: d = 0$$

$$H_1: d > 0$$

Under H_0 the slope parameters β can be estimated by the classical within estimation method. In this simple case, the dimensionality test of Kneip *et al.* (2012b) can be reduced to the following test statistic:

$$J_{KSS} = \frac{n \operatorname{tr}(\hat{\Sigma}_w) - (n-1)(T-1)\hat{\sigma}^2}{\sqrt{2n(T-1)}\hat{\sigma}^2} \underset{a}{\sim} N(0, 1),$$

where $\hat{\Sigma}_w$ is the covariance matrix of the within residuals. The reason for this simplification is that under H_0 there is no need for smoothing, which allows us to set $\kappa = 0$.

We reject $H_0: d = 0$ at a significance level α if $J_{KSS} > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of the standard normal distribution. It is important to note that the performance of the test depends heavily on the accuracy of the variance estimator $\hat{\sigma}^2$. We propose to use the variance estimators (3.16) or (3.50), which are consistent under both hypotheses as long as \hat{d} is greater than the unknown dimension d . Internally, the test procedure sets $\hat{d} = d.\max$.

This test can be performed for `Eup`- as well as for `KSS`-objects by using the function `checkSpecif()` leaving the second argument `obj2` unspecified. In the following we apply the test for both models:

```
R> Eup.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+               additive.effects = "twoways")
R> checkSpecif(Eup.obj, level = 0.01)
```

```
-----
Testing the Presence of Interactive Effects
Test of Kneip, Sickles, and Song (2012)
-----
```

H0: The factor dimension is equal to 0.

Test-Statistic	p-value	crit.-value	sig.-level
13.29	0.00	2.33	0.01

```
R> KSS.obj <- KSS(l.Consumption ~ -1 + l.Price + l.Income,
+               additive.effects = "twoways")
R> checkSpecif(KSS.obj, level = 0.01)
```

```
-----
Testing the Presence of Interactive Effects
Test of Kneip, Sickles, and Song (2012)
-----
```

H0: The factor dimension is equal to 0.

Test-Statistic	p-value	crit.-value	sig.-level
104229.55	0.00	2.33	0.01

The null hypothesis $H_0: d = 0$ can be rejected for both models at a significance level $\alpha = 0.01$.

3.6 Interpretation

This section is intended to outline an exemplary interpretation of the panel model (3.47), which is estimated by the function `KSS()` in Section 3.5. The interpretation of models estimated by the function `Eup()` can be done accordingly. For convenience sake we re-write the model (3.47) in the following:

$$\ln(\text{Consumption}_{it}) = \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + \alpha_i + v_i(t) + \varepsilon_{it},$$

$$\text{where } v_i(t) = \sum_{l=1}^d \lambda_{il} f_l(t).$$

A researcher, who chooses the panel models proposed by Kneip *et al.* (2012b) or Bai (2009) will probably find them attractive due to their ability to control for very general forms of unobserved heterogeneity. Beyond this a further great advantage of these models is that the time-varying individual effects $v_i(t)$ provide a valuable source of information about the *differences* between the individuals i . These differences are often of particular interest as, e.g., in the literature on stochastic frontier analysis.

The left panel of Figure 3.5 shows that the different states i have considerable different time-constant levels $\hat{\alpha}_i$ of cigarette consumption. A classical further econometric analysis could be to regress the additive individual effects $\hat{\alpha}_i$ on other time-constant variables, such as the general populations compositions, the cigarette taxes, etc.

The right panel of Figure 3.5 shows the five estimated common factors $\hat{f}_1(t), \dots, \hat{f}_5(t)$. It is a good practice to start the interpretation of the single common factors with an overview about their importance in describing the differences between the

$v_i(t)$'s, which is reflected in the variances of the individual loadings parameters $\hat{\lambda}_{il}$. A convenient depiction is the quantity of variance-shares of the individual loadings parameters on the total variance of the loadings parameters

$$\text{coef}(\text{Cigar2.KSS})\$\text{Var.shares.of.loadings.param}[l] = V(\hat{\lambda}_{il}) / \sum_{k=1}^d V(\hat{\lambda}_{ik}),$$

which is shown for all common functions $\hat{f}_1(t), \dots, \hat{f}_5(t)$ in the following table:

Common Factor	Share of total variance of $v_i(t)$
$\hat{f}_1(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[1]</code> = 66.32%
$\hat{f}_2(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[2]</code> = 24.28%
$\hat{f}_3(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[3]</code> = 5.98%
$\hat{f}_4(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[4]</code> = 1.92%
$\hat{f}_5(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[5]</code> = 1.50%

Table 3.1: List of the variance shares of the common factors $\hat{f}_1(t), \dots, \hat{f}_5(t)$.

The values in Table 3.1 suggest to focus on the first two common factors, which explain together about 90% of the total variance of the time-varying individual effects $\hat{v}_i(t)$.

The first two common factors

$$\begin{aligned} \text{coef}(\text{Cigar2.KSS})\$\text{Common.factors}[,1] &= \hat{f}_1(t) \quad \text{and} \\ \text{coef}(\text{Cigar2.KSS})\$\text{Common.factors}[,2] &= \hat{f}_2(t) \end{aligned}$$

are plotted as black solid and red dashed lines in the middle panel of Figure 3.5. Figure 3.6 visualizes the differences of the time-varying individual effects $v_i(t)$ in the direction of the first common factor (i.e: $\hat{\lambda}_{i1}\hat{f}_1(t)$) and in the direction of the second common factor (i.e: $\hat{\lambda}_{i2}\hat{f}_2(t)$). As for the time-constant individual effects $\hat{\alpha}_i$ a further econometric analysis could be to regress the individual loadings parameters $\hat{\lambda}_{i1}$ and $\hat{\lambda}_{i2}$ on other explanatory time-constant variables.

Generally, for both models proposed by Kneip *et al.* (2012b) and Bai (2009) the time-varying individual effects

$$\nu_{it} = \sum_{l=1}^d \lambda_{il} f_{lt}$$

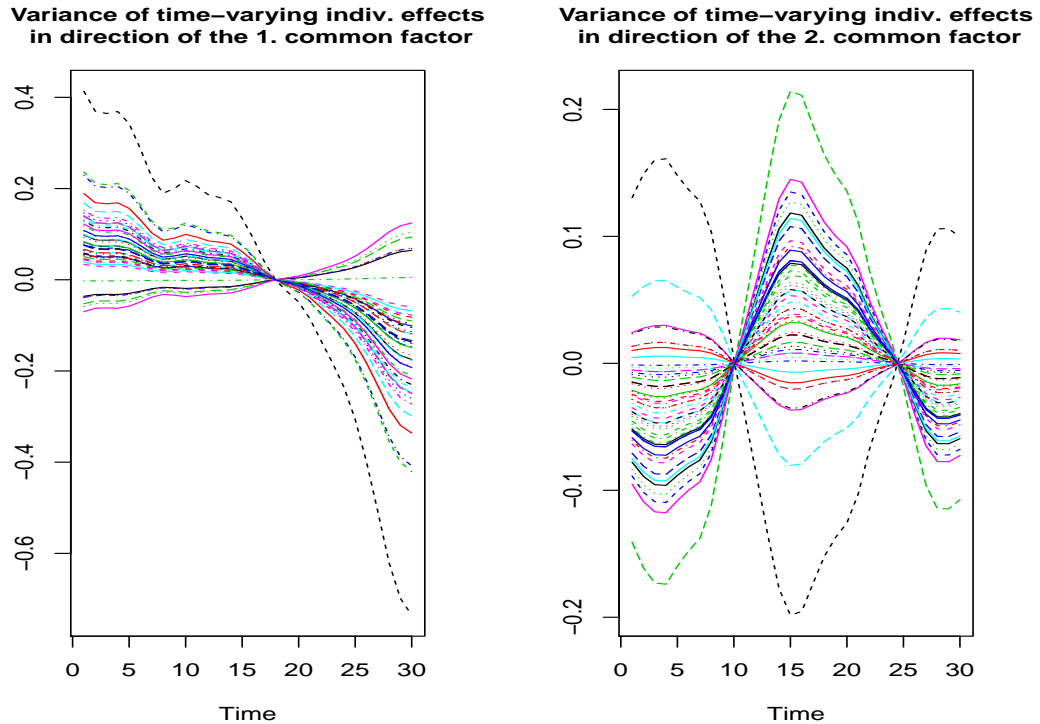


Figure 3.6: LEFT PANEL: Visualization of the differences of the time-varying individual effects $v_i(t)$ in the direction of the first factor $\hat{f}_1(t)$ (i.e: $\hat{\lambda}_{i1}\hat{f}_1(t)$). RIGHT PANEL: Visualization of the differences of the time-varying individual effects $v_i(t)$ in the direction of the second factor $\hat{f}_2(t)$ (i.e: $\hat{\lambda}_{i2}\hat{f}_2(t)$).

can be interpreted as it is usually done in the literature on factor models. An important topic that is not covered in this section is the rotation of the common factors. Often, the common factors f_l can be interpreted economically only after the application of an appropriate rotation scheme for the set of factors $\hat{f}_1, \dots, \hat{f}_d$. The latter can be done, e.g., using the function `varimax()` from the `stats` package. Sometimes it is also preferable to standardize the individual loadings parameters instead of the common factors as it is done, e.g., in Ahn *et al.* (2001). This can be done by choosing `restrict.mode = c("restrict.loadings")` in the functions `KSS()` and `Eup()` respectively.

3.7 Summary

This paper introduces the R package `phtt` for the new class of panel models proposed by Bai (2009) and Kneip *et al.* (2012b). The two main functions of the package are the `Eup()`-function for the estimation procedure proposed in Bai (2009) and the `KSS()`-function for the estimation procedure proposed in Kneip *et al.* (2012b). Both of the main functions are supported by several print, summary, and plot methods. While parts of the method of Bai (2009) are available for commercially available software packages, the estimation procedure proposed by Kneip *et al.* (2012b) is not available elsewhere. A further remarkable feature of our `phtt` package is the `OptDim()`-function, which provides an ease access to many different dimensionality criteria proposed in the literature on factor models. The usage of the functions is demonstrated by a real data applications.

Conclusion

The common theme underlying the chapters of this thesis is the use of functional data analysis as a valuable tool for statistical modeling of economic data. While many theoretical economic models build upon smooth functions, economic data are usually observed discretely with some additional uninformative noise components. We demonstrate that functional data analysis can bridge this gap in a most natural way. Consequently this thesis covers conceptual work, which shows the strength of functional data methods, when applied to economic contexts. Beyond this, this thesis covers methodological contributions, which are motivated from problems encountered in our applied work.

Regarding the conceptual part, a new functional data point of view on modeling and forecasting time series of hourly electricity spot prices is introduced. Motivated from the so-called merit order model we interpret electricity spot prices as noisy discretization points of smooth random price functions. In contrast to classical time series models, this approach provides a much more convenient way for the development of statistical models, which are well-interpretable in the context of electricity data.

With respect to the methodological contributions, we discuss in detail multivariate nonparametric regression as a tool for functional principal component analysis. Our theoretical considerations account for the temporal dependencies between the functional data as well as for dependencies between discretization points that are induced by the particular nature of functional data. As this contribution is motivated from the problem of modeling electricity spot prices, we consider also specific peculiarities in the random design of the prediction points electricity demand and air temperature.

In Chapter 1 a two-step method of estimating basis functions for the ran-

dom electricity price functions is presented. The initial estimation step is a pre-smoothing step, which takes into account the second estimation step, which aims for the statistical approximation of the underlying basis functions. The chapter concludes with an exhaustive real data study, which demonstrates the superior forecast performance of our model in comparison to many alternative forecast models for electricity spot prices.

In Chapter 2 we discuss multivariate nonparametric regression analysis as a tool for functional principal component analysis. Specifically, we consider the asymptotic properties of the multivariate local linear estimator when applied to smoothing discretization points of temporal dependent functional data. Beyond this, we extend our model for electricity spot prices introduced in Chapter 1 by considering further covariables. This yields a statical model, which is well-suited to for a comparison of different electricity market situations. The chapter concludes with a real data study, which contrasts the market situation one year before Germany's nuclear phase-out on March 14, 2011 and one year after.

In Chapter 3 we discuss a new type of panel data models that allows for complex, time-varying unobserved heterogeneity effects by the incorporation of latent factor models. An integral part of this chapter is the introduction of the statistical software package of Bada & Liebl (2013b), which contains estimation procedures for the models of Bai (2009) and Kneip *et al.* (2012a) that prototypical for this new type of panel data models. We extend these models for classical fixed effects, and introduce a new Hausman-type specification test. Particular attention is paid to the development of a reliable procedure to approximate the optimal smoothing parameter, which is needed in the estimation procedure proposed by Kneip *et al.* (2012a).

As the American economist and sociologist Thorstein Veblen (1857-1929) felicitously summed up, research makes two questions grow where only one grew before. Correspondingly, this thesis rises many new research questions and we name two of them: First, our work on electricity prices is based on the concept of smooth random price functions. However, it seems likely that the non-smooth deviations from these price functions also bear information, which can be of particular benefit in modeling extreme price events called price spikes. This is a challenging as well as active research topic and recent developments in the

literature on functional data analysis can contribute to this. Second, our discussion of the multicollinearity problem in the estimation the principal component scores points in Chapter 2 (pp. 69) to a possible solution, which is based on the non-restrictive assumption that the price functions are monotonically increasing functions. Future research should examine under which conditions this proposal actually provides a practical solution to this multicollinearity problem.

Danksagung

Zu guter Letzt bleibt mir noch, allen zu danken, die mich im Verlauf dieser Arbeit unterstützt haben.

Ohne Herrn Professor Dr. Alois Kneip wäre diese Arbeit nicht möglich gewesen. Sein Verständnis von Statistik hat mich maßgeblich geprägt. Ohne ihn und seine inspirierende Lehre hätte ich mich vermutlich nie diesem Themengebiet zugewandt. Dafür sowie für die zahlreichen Gespräche, Ratschläge und Ideen danke ich ihm ebenso wie für seine von Hilfsbereitschaft, Erfahrung und Weitsicht geprägte Betreuung.

Auch Herrn Professor Dr. Karl Mosler bin ich zu großem Dank verpflichtet. Er hat mich als Doktoranden in sein Team aufgenommen und mich in allen Entscheidungen tatkräftig unterstützt. Seine Anmerkungen sowie die mit ihm geführten Diskussionen lieferten wertvolle Beiträge für die vorliegende Arbeit. Für dies, seine große Hilfsbereitschaft, seine motivierende Betreuung sowie die vielen Möglichkeiten, an Konferenzen teilzunehmen, danke ich ihm.

Herrn Professor Dr. Marc Oliver Bettzüge danke ich für seine Bereitschaft, den Prüfungsvorsitz zu übernehmen. Er und sein Arbeitsteam haben mich mit ihrer konstruktiven Kritik und ihrem großen Wissen im Bereich der Energieökonomie sehr gefördert.

Des Weiteren danke ich Herrn Professor Dr. Pascal Sarda für seine Unterstützung und Betreuung während meines Forschungsaufenthaltes in Toulouse. Seine Geduld mit meinem äußerst rudimentären Französisch ist unübertroffen.

Herrn Professor Dr. Piotr Kokoszka danke für ich seine wohlwollenden Gutachten und den Glauben an meine Arbeit.

Meinem Freund, Koautor und “Leidensgenossen”, Oualid Bada, danke ich für die vielen, gemeinsam verbrachten Arbeitsstunden sowie die ungezählten

Gespräche und Diskussionen. All das werde ich sehr vermissen.

Schließlich bedanke ich mich auch noch bei all meinen Kollegen der beiden Lehrstühle von Herrn Professor Dr. Kneip und Herrn Professor Dr. Mosler für das in beiden Fällen stets freundschaftliche und fruchtbare Arbeitsklima.

Freilich haben mich noch wesentlich mehr Menschen im Laufe dieser Arbeit unterstützt. Auch bei ihnen möchte ich mich an dieser Stelle bedanken.

References

- AHN, S., HOON LEE, Y. & SCHMIDT, P. (2001). Gmm estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, **101**, 219–255. 132
- AHN, S., LEE, Y. & SCHMIDT, P. (2006). Panel data models with multiple time-varying individual effects. *Working Paper 0702, University of Crete, Department of Economics*. 96
- AHN, S.C. & HORENSTEIN, A.R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, *forthcoming*. 110
- ANTOCH, J., PRCHAL, L., ROSA, M. & SARDA, P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *Journal of Applied Statistics*, **37**, 2027–2041. 11, 16, 35
- ASH, R. & GARDNER, M. (1975). *Topics in stochastic processes*. Academic Press New York. 1
- BADA, O. & KNEIP, A. (2010). Panel data models with unobserved multiple time-varying effects to estimate the risk premium of corporate bonds. *Working Paper, University of Bonn*. 97, 98, 116, 118
- BADA, O. & LIEBL, D. (2012). *phtt: Panel Data Analysis with Heterogeneous Time Trends*. R package version 2.07. 98
- BADA, O. & LIEBL, D. (2013a). Panel data models with interactive fixed effects. *Journal of Statistical Software*,, accepted. 5

- BADA, O. & LIEBL, D. (2013b). *phtt: Panel data analysis with heterogeneous time trends*. R package version 2.07. 4, 6, 136
- BAI, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, **122**, 137–183. 110
- BAI, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, **77**, 1229–1279. 4, 5, 95, 96, 97, 98, 100, 115, 116, 118, 123, 130, 131, 133, 136
- BAI, J. & NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70**, 191–221. 98, 109, 110, 116
- BAI, J., KAO, C. & NG, S. (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics*, **149**, 82–99. 4, 96, 122, 125, 126, 127
- BALTAGI, B. (2005). *Econometric Analysis of Panel Data*. Wiley, 3rd edn. 95
- BALTAGI, B. & LEVIN, D. (1986). Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics*, 148–155. 99
- BALTAGI, B. & LI, D. (2004). *Prediction in the Panel Data Model with Spatial Correlation*. Springer-Verlag, 1st edn. 99
- BATES, D., MAECHLER, M. & BOLKER, B. (2012). *lme4: Linear Mixed-Effects Models Using S4 Classes*. R package version 0.999375-42. 98
- BENEDETTI, J.K. (1977). On the nonparametric estimation of regression functions. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 248–253. 18, 25
- BENKO, M., HÄRDLE, W. & KNEIP, A. (2009). Common functional principal components. *The Annals of Statistics*, **37**, 1–34. 1, 19, 30, 31
- BESSE, P. & RAMSAY, J. (1986). Principal components analysis of sampled functions. *Psychometrika*, **51**, 285–311. 50
- BORAK, S. & WERON, R. (2008). A semiparametric factor model for electricity forward curve dynamics. *Journal of Energy Markets*, **1**, 3–16. 12

- BOSCO, B., PARISIO, L., PELAGATTI, M. & FABIO, B. (2010). Long-run relations in european electricity prices. *Journal of Applied Econometrics*, **25**, 805–832. 16
- BOSQ, D. (2000). *Linear processes in function spaces: theory and applications*, vol. 149. Springer Verlag. 2, 58
- BROCKWELL, P.J. & DAVIS, R.A. (1991). *Time Series: Theory and Methods*. Springer, 2nd edn. 34
- BURGER, M., GRAEBER, B. & SCHINDLMAYR, G. (2008). *Managing Energy Risk: An Integrated View on Power and Other Energy Markets*. Wiley, 1st edn. 8, 14, 46, 52
- CAMPBELL, S.D. & DIEBOLD, F.X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, **100**, 6–16. 61
- CAO, J. & RAMSAY, J. (2010). Linear mixed-effects modeling by parameter cascading. *Journal of the American Statistical Association*, **105**, 365–374. 5, 103, 116
- CHRISTENSEN, T., HURN, S. & LINDSAY, K. (2009). It never rains but it pours: Modeling the persistence of spikes in electricity prices. *The Energy Journal*, **30**, 25–48. 14, 52
- CRAVEN, P. & WAHBA, G. (1978). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, **31**, 377–403. 4, 103
- CROISSANT, Y. & MILLO, G. (2008). Panel data econometrics in r: The plm package. *Journal of Statistical Software*, **27**, 1–43. 98
- DE BOOR, C. (2001a). *A Practical Guide to Splines*, vol. 27 of *Applied Mathematical Series*. Springer, revised edn. 19
- DE BOOR, C. (2001b). *A Practical Guide to Splines*, vol. 27 of *Applied Mathematical Series*. Springer-Verlag, revised edn. 101

- DEPRINS, D., SIMAR, L. & TULKENS, H. (1984). *Measuring labor inefficiency in post offices*. In: Marchand, M., Pestiau, P., Tulkens, H. (Eds.), *The Performance of Public Enterprises: Concepts and Measurements*. North-Holland, Amsterdam. 62
- ENGLE, R.F., GRANGER, C.W., RICE, J. & WEISS, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, **81**, 310–320. 48
- FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 1st edn. 23, 51
- FEBRERO-BANDE, M. & OVIEDO DE LA FUENTE, M. (2012). Statistical computing in functional data analysis: The R package *fda.usc*. *Journal of Statistical Software*, **51**, 1–28. 39
- FERRATY, F. & VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, Springer, 1st edn. 2, 11
- GERVINI, D. (2008). Robust Functional Estimation Using the Median and Spherical Principal Components. *Biometrika*, **95**, 587–600. 21
- GIJBELS, I. & PENG, L. (2000). Estimation of a support curve via order statistics. *Extremes*, **3**, 251–277. 62
- GNEITING, T. & RAFTERY, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378. 40
- GOODMAN, L. (1962). The variance of the product of k random variables. *Journal of the American Statistical Association*, **57**, 54–60. 82
- GRANGER, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438. 29

- GREENAWAY-MCGREY, R., HAN, C. & SUL, D. (2012). Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics (Forthcoming)*. 118
- HALL, P. & HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 109–126. 50
- HALL, P., MÜLLER, H.G. & WANG, J.L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, **34**, 1493–1517. 11, 22, 50, 51
- HÄRDLE, W. & TRÜCK, S. (2010). The dynamics of hourly electricity prices. *SFB 649 Discussion Papers*. 12, 36, 38
- HARVEY, A. & KOOPMAN, S.J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, **88**, 1228–1236. 48
- HAYS, S., SHEN, H. & HUANG, J.Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, **6**, 870–894. 11, 12
- HE, G., MÜLLER, H. & WANG, J. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, **85**, 54–77. 2
- HUISMAN, R. & DE JONG, C. (2003). Option pricing for power prices with spikes. *Energy Power Risk Management*, **7**, 12–16. 15, 36, 37
- HUISMAN, R., HUURMAN, C. & MAHIEU, R. (2007). Hourly electricity prices in day-ahead markets. *Energy Economics*, **29**, 240–248. 8, 15, 46
- KARAKATSANI, N. & BUNN, D. (2008). Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, **24**, 764–785. 8, 50

- KNEIP, A. & RAMSAY, J. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association*, **103**, 1155–1165. 2
- KNEIP, A. & UTIKAL, K.J. (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, **96**, 519–532. 64, 65, 91, 92
- KNEIP, A., PARK, B.U. & SIMAR, L. (1998). A note on the convergence of non-parametric dea estimators for production efficiency scores. *Econometric theory*, **14**, 783–793. 62
- KNEIP, A., SICKLES, R.C. & SONG, W. (2012a). A new panel data treatment for heterogeneity in time trends. *Econometric Theory*, **28**, 590–628. 4, 5, 136
- KNEIP, A., SICKLES, R.C. & SONG, W. (2012b). A new panel data treatment for heterogeneity in time trends. *Econometric Theory*, **28**, 590–628. 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 108, 109, 122, 123, 125, 128, 130, 131, 133
- KNITTEL, C. & ROBERTS, M. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, **27**, 791–817. 8, 15
- KOOPMAN, S., OOMS, M. & CARNERO, M. (2007). Periodic seasonal garch models for daily electricity spot prices. *Journal of the American Statistical Association*, **102**, 16–27. 8, 50
- KOSATER, P. & MOSLER, K. (2006). Can markov regime-switching models improve power-price forecasts? evidence from german daily power prices. *Applied Energy*, **83**, 943–958. 8, 15, 35, 37, 50
- LAU, A. & MCSHARRY, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *The Annals of Applied Statistics*, **4**, 1311–1341. 16, 35
- LI, Y. & HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, **38**, 3321–3351. 11, 22, 50, 51

- LIEBL, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. *The Annals of Applied Statistics*, forthcoming. 3, 46
- LIU, C., RAY, S., HOOKER, G. & FRIEDL, M. (2012). Functional factor analysis for periodic remote sensing data. *The Annals of Applied Statistics*, **6**, 601–624. 11
- LOCANTORE, N., MARRON, J.S., SIMPSON, D.G., TRIPOLI, N., ZHANG, J.T., COHEN, K.L., BOENTE, G., FRAIMAN, R., BRUMBACK, B. & CROUX, C. (1999). Robust principal component analysis for functional data. *Test*, **8**, 1–73. 21
- MARTINS-FILHO, C. & YAO, F. (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics*, **141**, 283–319. 62, 63
- MILLO, G. & PIRAS, G. (2012). splm: Spatial panel data models in r. *Journal of Statistical Software*, **47**, 1–38. 98
- NESTLE, U. (2012). Does the use of nuclear power lead to lower electricity prices? an analysis of the debate in germany with an international perspective. *Energy Policy*, **41**, 152–160. 4, 46, 51, 74
- OCKENFELS, A., GRIMM, V. & ZOETTL, G. (2008). Electricity market design. *Expertise commissioned by EEX for submission to the Saxon Exchange Supervisory Authority*. 13, 46, 51
- ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, **92**, 1004–1016. 111, 116
- PARK, B.U., MAMMEN, E., HÄRDLE, W. & BORAK, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, **104**, 284–298. 12, 36, 37, 38, 42
- PESARAN, H.M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, **74**, 967–1012. 96

- PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D. & R CORE TEAM (2012). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-103. 98
- R DEVELOPMENT CORE TEAM (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. 98
- RAMSAY, J.O. (1996). Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society. Series B (Methodological)*, 495–508. 2
- RAMSAY, J.O. & DALZELL, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 539–572. 2, 50
- RAMSAY, J.O. & RAMSEY, J.B. (2002). Functional data analysis of the dynamics of the monthly index of nondurable goods production. *Journal of Econometrics*, **107**, 327–344. 1
- RAMSAY, J.O. & SILVERMAN, B.W. (2005). *Functional Data Analysis*. Springer Series in Statistics, Springer, 2nd edn. 2, 11, 19, 22, 53, 72
- RAO, C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**, 1–17. 1
- RICE, J. & SILVERMAN, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, **53**, 233–243. 27, 48
- RUPPERT, D. & WAND, M. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, **22**, 1346–1370. 60, 64, 67, 82
- RUPPERT, D., SHEATHER, S.J. & WAND, M.P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, **90**, 1257–1270. 74

- STANISWALIS, J. & LEE, J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, **93**, 1403–1418. 11, 22
- THE MATHWORKS INC. (2012). *MATLAB – The Language of Technical Computing, Version 7.14.*. The MathWorks, Inc., Natick, Massachusetts. 98
- VILAR, J., CAO, R. & ANEIRO, G. (2012). Forecasting next-day electricity demand and price using nonparametric functional methods. *International Journal of Electrical Power & Energy Systems*, **39**, 48–55. 11, 36, 37, 39, 40, 41
- WANG, L., BROWN, L.D., CAI, T.T. & LEVINE, M. (2008a). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics*, **36**, 646–664. 68
- WANG, S., JANK, W. & SHMUELI, G. (2008b). Explaining and forecasting online auction prices and their dynamics using functional data analysis. *Journal of Business & Economic Statistics*, **26**. 1
- WERON, R., BIERBRAUER, M. & TRÜCK, S. (2004). Modeling electricity prices: Jump diffusion and regime switching. *Physica A: Statistical and Theoretical Physics*, **336**, 39–48. 8, 50
- WOLAK, F.A. (2000). *Market Design and Price Behavior in Restructured Electricity Markets: An International Comparison*. Deregulation and Interdependence in the Asia-Pacific Region, NBER-EASE Volume 8, University of Chicago Press. 66
- YAO, F., MÜLLER, H.G. & WANG, J.L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590. 11, 20, 22, 23, 26, 50, 55, 68, 73

Curriculum Vitae

Contact Information

Address: Dominik Liebl
University of Cologne
Department of Economic & Social Statistics
Meister-Ekkehart-Str. 9
50937 Cologne (Germany)

Email: liebl@statistik.uni-koeln.de

Current Positions

since 10/2010 **Research Assistant** at the Chair of Prof. Dr. Karl Mosler, *University Cologne*

since 03/2009 **Research Assistant** at the Chair of Prof. Dr. Alois Kneip, *University Bonn*

since 10/2008 **Doctoral Student** at the Cologne Graduate School in Management, Economics, and Social Sciences, University Cologne, Supervisors: Prof. Dr. Karl Mosler and Prof. Dr. Alois Kneip

Education

04/2010-08/2010	Visiting PhD Student at the Working group STAPH in Toulouse (Université Toulouse III, Paul Sabathier)
2008-2010	PhD Candidate at the University Cologne, supervisors: Prof. Dr. Karl Mosler and Prof. Dr. Alois Kneip
2008	Diploma Thesis , <i>The EM Algorithm and its Application</i> , Supervisor: Prof. Dr. Alois Kneip
2006-2007	Visiting Student , <i>Charles University Prague</i>
2002-2008	Diploma (M.Sc. equivalent) in Economics, <i>University Bonn</i> , Special emphasis on: Statistics
2001	Abitur Johannes Nepomuk Gymnasium, Rohr i. NB.

Academic & Professional Experience

10/2012-today	Teaching assignment , University Cologne
10/2012-02/2013	Teaching assignment , Kölner Journalistenschule
10/2010-today	Research & Teaching Assistant , University Cologne, Chair for Economic and Social Statistics, Prof. Dr. K. Mosler
03/2009-today	Research Assistant , University Bonn, Statistical Department, Prof. Dr. A. Kneip
10/2008-02/2009	Teaching Assistant , University Cologne, Chair for Empirical Social and Economic Research, Prof. Dr. H-J. Andreß
02/2007-03/2007	Internship at KPMG
09/2005-10/2005	Internship at e.on Bayern AG
02/2005-04/2005	Internship at KPMG

Publications

- Liebl, D. (2013). Modeling and Forecasting Electricity Spot Prices: A Functional Data Perspective, *Annals of Applied Statistics*, forthcoming
- Bada, O., Liebl, D. (2013). The R-package phtt: Panel Data Analysis with Heterogeneous Time Trends, *Journal of Statistical Software*, accepted

Miscellaneous

Language skills	English fluent; basic Spanish and French
Programming	R/S-plus, Matlab, Java and C++
Sports	3 times German Judo Champion (2004-2006), 7th in the Military Judo World Championships in Beijing (1997) Trainer of an inclusive Judo group

Cologne, August 28, 2013