Three Essays in Econometrics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Fabian Walders

aus Bonn

2018

Dekan:	Prof. Dr. Daniel Zimmer, LL.M.
Erstreferent:	Prof. Dr. Alois Kneip
Zweitreferent:	Prof. Dr. Christian Hafner
Tag der mündlichen Prüfung:	27.07.2018

To my parents.

Acknowledgments

Over the last four years I have greatly benefited from the support of many people. First and foremost, I want to thank my main advisor Alois Kneip for his enduring support and his intellectual guidance. In our discussions and meetings, I had the chance to benefit from his outstanding expertise and helpful thoughts on my research. I want to express my gratitude to Christian Hafner and Dominik Liebl. It has been inspiring and a pleasure to work with both of them. I am very glad that our joint projects constitute two chapters of this dissertation.
Further I would like to thank the Bonn Graduate School of Economics and the Institute for Financial Economics and Statistics for the unique opportunity as well as the material support they gave me. In particular I want to thank Silke Kinzig, Michael Vogt and Jörg Stoye as well as my colleagues Katja Mann, Robert Scherf, Jens Herold, Daniel Becker, Dominik Poß and Marina Khismatullina for their help. I owe particular gratitude to my friends and colleagues Thomas Graeber, Lukas Püttmann and Justus Winkelmann. I am glad I had the chance to spend the last four years with these exceptional and inspiring persons—it has been an honor to me.

I am supremely thankful to my sister Wibke and my girlfriend Nicole for their enduring help and the energy they gave me during my doctoral studies. Most importantly, I am grateful for my wonderful parents Angela and Dirk, who always encouraged me to find and pursue my intellectual goals. I am deeply indebted for the many years of their unconditional support: Thank you!

Contents

List of Figures

List of Tables

Introduction

This thesis consists of three self-contained essays in econometrics and statistics. It discusses methodological topics in semi- and nonparametric statistics as well as empirical questions in financial econometrics. My methodological contributions cover regression and times series prediction techniques, which are designed to deal with complex and high-dimensional data. The proposed statistical concepts are particularly useful for the analysis of risk in modern financial markets, which is of major importance for example in risk management, automated trading and for regulatory authorities. The empirical contribution of this thesis is an analysis of the role and dynamics of price and liquidity risk of equity and debt securities in developed capital markets. The applications in the first two chapters deal with risk on stock markets, while the third chapter deals with risk on bond markets.

Chapters 1 and 2 are concerned with systematic time heterogeneity in econometric models dealing with functional data. With these works I contribute in particular to the literature on *functional data analysis*, a branch of statistics which has become increasingly popular within the last two decades. The term functional data refers to the outcomes of random variables that are curve-valued, such as human growth curves, temperature curves or implied volatility surfaces of options on financial assets. Ramsay and Silverman (2005) offer a general introduction to the topic. Functional data analysis is useful for econometricians, for example when dealing with high frequency data from stock markets. Stock trading takes place in continuous time over a trading day and consequently the underlying population price process is most reasonably described as a continuous time stochastic process. Densely sampled observations from this process, potentially allow to infer curve-valued characteristics describing its moments or functional components and parameters of an underlying population model. Motivated by this observation, the first two chapters present statistical concepts allowing to deal with dynamic features and pricing effects of such characteristics.

CHAPTER 1. In the first chapter, which is based on joint work with Dominik Liebl, we propose a partial functional linear regression model for panel data.[1] The suggested semiparametric regression framework allows analyzing time-varying marginal effects of a functional covariate on a scalar response, in the presence of a vector-valued control variable. While the parameter of the control variable is completely time-varying, the parameter of the functional model component is assumed to change over a fixed number of unknown parameter regimes. This type of parameter instability is motivated by recent research on multivariate panel data models as in Vogt and Linton (2017) and Su et al. (2016). We suggest simple estimators for the regimes and the regression slopes and examine their asymptotic properties as the time and the cross section dimensions diverge simultaneously. Our statistical model is particularly helpful for exploring time heterogeneity in the pricing of idiosyncratic risk. In an empirical study of the US stock market, we use our framework to quantify when idiosyncratic risk premiums turn negative, a phenomenon which is known as the *idiosyncratic volatility puzzle*. Based on a functional measure of idiosyncratic volatility, our model identifies systematically negative risk premiums for almost half of the available trading days. The composition of the corresponding parameter regimes indicates a complex temporal pattern of the underlying risk pricing mechanism.

CHAPTER 2. The second chapter starts from a similar notion of stock price riskiness as the application in the first chapter. In this work I consider online prediction of spot volatility curves for intraday stock price returns. Such curves are considered to be random functions, governing the intraday variation of returns in continuous time. I assess the forecasting problem, which is of particular interest for automated trading, by means of a novel component model for functional volatility. The model incorporates an intraday shape pattern of volatility, which is supplemented by level shifts of the curves. As a key novelty, I allow for partially systematic deviations from the intraday shape pattern. I suggest a new distance metric learning procedure to uncover latent similarity of such deviations over trading days. The quantified distance notion can be used to forecast shape properties of the intraday volatility curve from a functional kernel regression. The accuracy of my prediction technique is examined in an empirical study of the German stock market. The results indicate a promising performance of the mechanism when compared to relevant competitors.

While the applications of the first two chapters focus on asset price risk in equity markets, the third chapter shifts the focus to liquidity risk of debt securities. Markets for corporate bonds are typically less centralized and asset characteristics are less homogeneous than in stock markets. This chapter offers an econometric assessment of market segmentation according to the effects of

[1] Cf. Liebl and Walders (2018) and Walders and Liebl (2017).

liquidity risk, which helps for example institutional investors or financial authorities to evaluate the risk of bond portfolios. It relies on a similar notion of parameter heterogeneity as in the first chapter and adapts the model of Vogt and Linton (2017) to the case of unbalanced panel data.

CHAPTER 3. The third chapter is based on joint work with Christian Hafner, which is published in the *Journal of Fixed Income*.[2] It assesses systematic heterogeneity in the risk pricing for fixed income securities and links in particular to the panel model from the first chapter. Among the many sources of risk explaining corporate bond spreads, the role of liquidity is the least well understood. This chapter investigates the impact of liquidity risk of unknown functional form on the yield spread over time. Heterogeneity is introduced via a latent group structure explaining differences in nonlinear liquidity effects across groups of bonds. A key feature of the model is that it can be estimated from highly unbalanced longitudinal data, allowing us to work with data at minimum levels of temporal aggregation. In an extensive empirical study we apply the suggested method to a large panel of trade data for US corporate bonds. Our procedure identifies nonlinear liquidity effects for a large fraction of the securities. The classification clearly distinguishes groups differing e.g. in bond characteristics such as spread levels, trading activity and also established measures such as credit ratings and the time to maturity. While most groups share similar dynamics of liquidity effects, their magnitudes as well as the interplay between different idiosyncratic liquidity proxies differ substantially across groups.

The unifying theme of the three chapters in this dissertation is the analysis of systematic heterogeneity in disaggregate complex data. All works address instabilities of the structure of statistical models over time or statistical units. The underlying ability to adjust statistical models automatically when the empirical phenomena change, will become increasingly important, not only in the analysis of financial markets. For example new digital consumer goods and new digital features of existing goods such as connected cars or smart homes will produce a much larger amount and a much more complex mix of data as it was available only twenty years ago. The resulting mixed, large scale data will demand new statistical concepts to generate insights from highly heterogeneous empirical phenomena. In particular the efficient online adjustment of statistical procedures will be a key to access the value of data from the new digital world.

[2] Cf. Hafner and Walders (2017).

Chapter 1

Parameter Regimes in Partial Functional Panel Regression

1.1 Introduction

The availability of mixed—functional and multivariate—data types and the need to analyze such data types appropriately, has trigged the development of new statistical models and procedures. In this work we consider the so-called partial functional linear model for scalar responses, which combines the classical functional linear regression model (see, e.g., Hall and Horowitz, 2007) with the classical multivariate regression model. This model was first proposed by Zhang et al. (2007) and Schipper et al. (2008)—two mixed effects modeling approaches. The first theoretical work is by Shin (2009), who uses a functional-principal-components-based estimation procedure and derives convergence rates for the case of independent cross-sectional data. Recently, the partial functional linear regression model was extended in several directions. Shin and Lee (2012) consider the case of prediction, Lu et al. (2014) and Tang and Cheng (2014) focus on quantile regression, Kong et al. (2016) consider the case of a high-dimensional multivariate model component, Peng et al. (2016) allow for varying coefficients in the multivariate model component, and Wang (2016) and Ding et al. (2017) are concerned with a functional single-index model component.

Motivated by our real data application, we deviate from this literature and contribute a new partial functional linear panel regression model with time-varying parameters allowing for $K < \infty$ latent parameter regimes, which can be estimated from the data. In the theoretical part of this work we show consistency of our estimators and of our unsupervised classification procedure identifying the K parameter regimes. In addition, we derive convergence rates of the regression slope estimators under a double asymptotic, for which we differentiate among different asymptotic

scenarios. These scenarios depend on the relative order of the panel dimensions n and T.
The consideration of time-varying parameters is quite novel in the literature on functional data analysis. To the best of our knowledge, the only other work concerned with this issue is Horváth and Reeder (2012), who focus on testing the hypothesis of a time constant parameter function in the case of a classical fully-functional regression model. Closely related to the partial functional linear model is the so-called Semi-Functional Partial Linear (SFPL) model proposed by Aneiros-Pérez and Vieu (2006), where the functional component consists of a nonparametric functional regression model instead of a functional linear regression model. The SFPL model is further investigated by Aneiros-Pérez and Vieu (2008), Lian (2011), Zhou and Chen (2012), and Aneiros-Pérez and Vieu (2013), among others. Readers with a general interest in functional data analysis are referred to the textbooks of Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) and Hsing and Eubank (2015).
The usefulness of our model and the applicability of our estimation procedure is demonstrated by means of a simulation study and a real data application. For the latter we consider the so-called "idiosyncratic volatility puzzle" using high frequency stock-level data from the S&P 500. Our model allows us to consider this puzzle at a much less aggregated time scale than considered so far in the literature. This leads to new insights into the temporal heterogeneity in the pricing of idiosyncratic risk in equity markets.
The remainder of this work is structured as follows. In sections 1.2 and 1.3 we introduce the model and present the estimation procedure. Section 1.4 contains our main assumptions and asymptotic results. Section 1.5 discusses the practical choice of the tuning parameters involved. The finite sample performance of the estimators is explored in Section 1.6. Section 1.7 offers an empirical study examining regime dependent pricing of idiosyncratic risk in the US stock market. Section 1.8 contains a short conclusion. All proofs can be found in the appendix of this chapter.

1.2 Model

We introduce a partial linear regression model for panel data, which allows us to model the time-varying effect of a square integrable random function $X_{it} \in L^2([0,1])$ on a scalar response $y_{it} \in \mathbb{R}$ in the presence of a random, finite dimensional explanatory variable $z_{it} \in \mathbb{R}^P$. Indexing the cross-section units $i = 1, \ldots, n$ and points in time $t = 1, \ldots, T$, our statistical model reads as

$$y_{it} = \rho_t + \int_0^1 \alpha_t(s) X_{it}(s) \mathrm{d}s + \beta_t^\top z_{it} + \epsilon_{it}, \tag{1.1}$$

where ρ_t is a time fixed effect, $\alpha_t \in L^2([0,1])$ is a time-varying deterministic functional parameter, $\beta_t \in \mathbb{R}^P$ is a time-varying deterministic parameter vector, and ϵ_{it} is a scalar error term with zero mean and finite but potentially time heteroscedastic variances (see also our assumptions in Section 1.4).[1]

The unknown function-valued parameters α_t, $1 \leq t \leq T$, are assumed to differ only across unknown time regimes $G_k \subset \{1, \ldots, T\}$. That is, each regime G_k is associated with a regime specific parameter function $A_k \in L^2([0,1])$, such that

$$\alpha_t(s) \equiv A_k(s) \quad \text{if} \quad t \in G_k. \tag{1.2}$$

The regimes $G_1, \ldots, G_K$ form a partition of the set of periods $\{1, \ldots, T\}$ and do not have to consist of subsequent periods t. The number of regimes K is fixed and does not depend on the number of points in time T. For our theoretical analysis in Section 1.4, we also allow the joint and the marginal distributions of X_{it}, z_{it} and ϵ_{it} to vary over the different regimes G_k.

We have in mind a situation where G_k is a collection of periods t, which belong to the k-th risk pricing regime as in our application. Here, the k-th regime is characterized by the function-valued slope parameter A_k, which is the marginal effect of the functional idiosyncratic risk measure X_{it} on the scalar stock price return y_{it}, $t \in G_k$, for stocks $1 \leq i \leq n$. The objective of our application is to differentiate between different pricing mechanisms A_k, controlling for potentially time heterogeneous effects of additional control variables z_{it}.

Model (1.1) nests several different specifications. It might be the case that $K = 1$ and hence $G_1 = \{1, \ldots, T\}$. In this situation the effect of the random function on the response is time invariant. The classical functional or the classical multivariate linear regression model are obtained if $\beta_t = 0$ or $\alpha_t = 0$ for all $t = 1, \ldots, T$.

1.3 Estimation

Our objective is to estimate the model parameters A_k, β_t, and the regimes $G_1, \ldots, G_K$ from a sample $\{(y_{it}, X_{it}, z_{it}) : 1 \leq i \leq n, 1 \leq t \leq T\}$. For this purpose, we suggest a three-step estimation procedure. The first step is a pre-estimation step where model (1.1) is fitted to the data *separately* for each $t = 1, \ldots, T$. This pre-estimation step reveals information about the regime memberships, which is used in the second step, where we apply our unsupervised classification procedure in order to estimate the regimes $G_1, \ldots, G_K$. The third step is the final estimation step, in which we improve the estimation of the functional parameter A_k by employing information

[1] We use the term *time fixed effect* in the sense that ρ_t is a latent time (t-) specific random variable, which is potentially correlated with the regressors.

about the regime membership gathered in step two. The general procedure is inspired by the work of Vogt and Linton (2017), but differs from it as we consider a functional data context which demands for a different estimation procedure. In the following we explain the three estimation steps in more detail:

Step 1. In this step, we pre-estimate the parameters α_t and compute the final estimates of β_t separately for each $t = 1, \ldots, T$. Estimation starts from removing the fixed effect ρ_t using a classical within-transformation. For this we denote the centered variables as $y_{it}^c = y_{it} - \bar{y}_t$, $X_{it}^c = X_{it} - \bar{X}_t$, $z_{it}^c = z_{it} - \bar{z}_t$, and $\epsilon_{it}^c = \epsilon_{it} - \bar{\epsilon}_t$, where $\bar{y}_t = n^{-1}\sum_{i=1}^n y_{it}$, $\bar{X}_t = n^{-1}\sum_{i=1}^n X_{it}$, $\bar{z}_t = n^{-1}\sum_{i=1}^n z_{it}$, and $\bar{\epsilon}_t = n^{-1}\sum_{i=1}^n \epsilon_{it}$. Then, the within-transformed version of model (1.1) is

$$y_{it}^c = \int_0^1 \alpha_t(u) X_{it}^c(u) \mathrm{d}u + \beta_t^\top z_{it}^c + \epsilon_{it}^c.$$

By adapting the methodology in Hall and Horowitz (2007), we estimate the slope parameter α_t using (t-wise) truncated series expansions of α_t and X_{it}^c, i.e.

$$\alpha_t(s) \approx \sum_{j=1}^{m_t} a_{j,t}\hat{\phi}_{j,t}(s) \quad \text{where} \quad a_{j,t} := \langle \alpha_t, \hat{\phi}_{j,t}\rangle,\ 1 \le j \le m_t$$

$$\text{and} \qquad X_{it}^c(s) = \sum_{j=1}^{n} \langle X_{it}^c, \hat{\phi}_{j,t}\rangle \hat{\phi}_{j,t}(s) \approx \sum_{j=1}^{m_t} \langle X_{it}^c, \hat{\phi}_{j,t}\rangle \hat{\phi}_{j,t}(s),$$

which can be used to approximate the functional $\int_0^1 \alpha_t(u) X_{it}^c(u)\mathrm{d}u$ in the regression equation by $\sum_{j=1}^{m_t}\langle X_{it}^c, \hat{\phi}_{j,t}\rangle a_{j,t}$. Here, $\langle\cdot,\cdot\rangle$ is the inner product in $L^2([0,1])$ and $\hat{\phi}_{j,t}$ denotes the eigenfunction corresponding to the j-th largest eigenvalue $\hat{\lambda}_{j,t}$ of the empirical covariance operator $\hat{\Gamma}_t$ of $\{X_{it} : 1 \le i \le n\}$. The operator $\hat{\Gamma}_t$ is defined as

$$(\hat{\Gamma}_t x)(u) := \int_0^1 \hat{K}_{X,t}(u,v) x(v)\mathrm{d}v \quad \text{for any} \quad x \in L^2([0,1])$$

$$\text{and} \quad \hat{K}_{X,t}(u,v) := \frac{1}{n}\sum_{i=1}^n X_{it}^c(u) X_{it}^c(v).$$

The eigenfunctions $\hat{\phi}_{j,t}$ and eigenvalues $\hat{\lambda}_{j,t}$ are defined as the solutions of the eigen-equations $\int_0^1 \hat{K}_{X,t}(u,v)\hat{\phi}_{j,t}(v)\mathrm{d}v = \hat{\lambda}_{j,t}\hat{\phi}_{j,t}(u)$, where $\langle \hat{\phi}_{j,t}, \hat{\phi}_{\ell,t}\rangle = 1$ for all $j = \ell$ and $\langle \hat{\phi}_{j,t}, \hat{\phi}_{\ell,t}\rangle = 0$ if $j \neq \ell$, where $j, \ell \in \{1, 2, \ldots, n\}$. This leads to the following estimators for the functional slope

parameter α_t and the slope parameter β_t:

$$\hat{\alpha}_t = \sum_{j=1}^{m_t} \hat{a}_{j,t}\hat{\phi}_{j,t} \quad \text{with} \quad \hat{a}_{j,t} = \hat{\lambda}_{j,t}^{-1} \frac{1}{n}\sum_{i=1}^{n} \langle X_{it}^c, \hat{\phi}_{j,t}\rangle (y_{it}^c - \hat{\beta}_t^\top z_{it}^c) \quad \text{and}$$

$$\hat{\beta}_t = \left[\hat{\mathbf{K}}_{z,t} - \hat{\mathbf{\Phi}}_t(\hat{\mathbf{K}}_{zX,t})\right]^{-1} \left[\hat{\mathbf{K}}_{zy,t} - \hat{\Phi}_t\left(\hat{K}_{yX,t}\right)\right],$$

where

$$\hat{\mathbf{K}}_{z,t} := \frac{1}{n}\sum_{i=1}^{n} z_{it}^c z_{it}^{c\top}, \;\; \hat{\mathbf{K}}_{zX,t}(s) := [\hat{K}_{z_1X,t}(s), \ldots, \hat{K}_{z_PX,t}(s)]^\top, \;\; \hat{\mathbf{K}}_{zy,t} := [\hat{K}_{z_1y,t}, \ldots, \hat{K}_{z_Py,t}]^\top,$$

$$\hat{K}_{yX,t}(s) := \frac{1}{n}\sum_{i=1}^{n} y_{it}^c X_{it}^c(s), \quad \hat{K}_{z_pX,t}(s) := \frac{1}{n}\sum_{i=1}^{n} z_{p,it}^c X_{it}^c(s), \quad \hat{K}_{z_py,t} := \frac{1}{n}\sum_{i=1}^{n} z_{p,it}^c y_{it}^c,$$

$$\hat{\Phi}_t(g) := [\hat{\Phi}_{1,t}(g), \ldots, \hat{\Phi}_{P,t}(g)]^\top, \quad \hat{\Phi}_{p,t}(g) := \sum_{j=1}^{m_t} \frac{\langle \hat{K}_{z_pX,t}, \hat{\phi}_{j,t}\rangle \langle \hat{\phi}_{j,t}, g\rangle}{\hat{\lambda}_{j,t}} \text{ for any } g \in L^2([0,1]),$$

and $\hat{\mathbf{\Phi}}_t(\hat{\mathbf{K}}_{zX,t}) := [\hat{\Phi}_{p,t}(\hat{K}_{z_qX,t})]_{1\le p\le P,\, 1\le q\le P}$;

see Shin (2009) for similar estimators in a cross-section context.

For our theoretical analysis, we let $m_t = m_{t,nT} \to \infty$ as $n, T \to \infty$. In practice, the cut-off parameter m_t can be chosen, for instance, by Cross Validation (CV) or by a suitable information criterion as introduced in Section 1.5.

Besides obtaining the final estimators $\hat{\beta}_t$ for β_t, this first estimation step is intended to pave the way for the classification procedure in Step 2. With such classification we aim to distinguish systematically large from systematically small differences between estimated functions $\hat{\alpha}_t$ and $\hat{\alpha}_s$ across different periods $t \neq s$. For this purpose one could compare the magnitude of the differences between the functions $\hat{\alpha}_t$ and $\hat{\alpha}_s$ to an appropriate threshold. However, the estimators $\hat{\alpha}_t, \hat{\alpha}_s$ are not well suited for deriving a practically useful threshold parameter. We, therefore, suggest the following transformed estimators, for which it is straightforward to derive a valid threshold parameter using distributional arguments (see Section 1.5):

$$\hat{\alpha}_t^{(\Delta)} := \sum_{j=1}^{\underline{m}} \frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{\epsilon,t}} \hat{a}_{j,t}\hat{\phi}_{j,t}, \tag{1.3}$$

where $\hat{\sigma}_{\epsilon,t}^2 := n^{-1}\sum_{i=1}^{n}\left(y_{it}^c - \langle \hat{\alpha}_t, X_{it}^c\rangle + \hat{\beta}_t^\top z_{it}^c\right)^2$ and $\underline{m} := \min_{1\le t\le T} m_t$.

Step 2. In this step, we use the scaled estimators $\hat{\alpha}_t^{(\Delta)}$ from (1.3) to classify periods $t = 1\ldots, T$ into regimes $G_1, \ldots, G_K$. Our classification algorithm aims to detect systematic differences in

the empirical distances $\hat{\Delta}_{ts} := ||\hat{\alpha}_t^{(\Delta)} - \hat{\alpha}_s^{(\Delta)}||_2^2$, where $||.||_2^2$ denotes the squared L^2 norm defined as $||x||_2^2 = \langle x, x\rangle$ for any $x \in L^2([0,1])$.

The algorithm detects regimes by iteratively searching for large differences $\hat{\Delta}_{ts}$. If $\hat{\Delta}_{ts}$ exceeds the value of a threshold parameter $\tau_{nT} > 0$, it classifies periods t and s in different regimes. The procedure is initialized by setting $S^{(0)} := \{1, \dots, T\}$ and then iterates over $k = 0, 1, 2, \dots$ the following procedure:

while $\left|S^{(k)}\right| > 0$ **do**
  select any $t \in S^{(k)}$, $\hat{G}_{k+1} \leftarrow \emptyset$, $S^{(k+1)} \leftarrow \emptyset$
  for $s \in S^{(k)}$ **do**
    if $\hat{\Delta}_{ts} \leq \tau_{nT}$ **then**
      $\hat{G}_{k+1} \leftarrow \hat{G}_{k+1} \cup \{s\}$
    else $S^{(k+1)} \leftarrow S^{(k+1)} \cup \{s\}$
    end if
  end for
end while

where we use $|\cdot|$ to denote the cardinality of a set. The algorithm stops as soon as all periods t are classified into regimes and the total number $\hat{K}$ of estimated regimes $\hat{G}_1, \dots, \hat{G}_{\hat{K}}$ serves as a natural estimator for the true K. Our theoretical results show that this procedure consistently estimates the true regimes G_k and the true number K. However, in order to improve the classification in finite samples, we suggest to set an upper bound $K_{\max}$ on $\hat{K}$, such that $\hat{K} \leq K_{\max}$. The practical choice of $K_{\max}$ is described in Section 1.5. In the case where $K_{\max}$ is binding, the algorithm is stopped after $K_{\max} - 1$ iterations and all remaining periods t are assigned to a final regime $\hat{G}_{K_{\max}}$.

Step 3. In this step, we build upon the regime structure determined in Step 2 in order to estimate A_k, $k = 1, \dots, \hat{K}$. For a regime k and any $t \in \hat{G}_k$, let X_{it}^{cc} denote the regime specific centered functional regressor defined as $X_{it}^{cc} := X_{it} - |\hat{G}_k|^{-1} \sum_{s \in \hat{G}_k} \bar{X}_s$. Further we define the corresponding k-specific empirical covariance operator $\tilde{\Gamma}_k$ by

$$(\tilde{\Gamma}_k x)(u) := \int_0^1 \tilde{K}_{X,k}(u,v)x(v)\mathrm{d}v \quad \text{for all} \quad x \in L^2([0,1])$$

$$\text{where} \quad \tilde{K}_{X,k}(u,v) := \frac{1}{n|\hat{G}_k|} \sum_{i=1}^{n} \sum_{t \in \hat{G}_k} X_{it}^{cc}(u) X_{it}^{cc}(v).$$

We obtain our final estimator $\tilde{A}_k$ for A_k, in analogy to the pre-estimator $\hat{\alpha}_t$, as

$$\tilde{A}_k = \sum_{j=1}^{\tilde{m}_k} \tilde{a}_{j,k}\tilde{\phi}_{j,k}, \quad \text{with} \quad \tilde{a}_{j,k} = \tilde{\lambda}_{j,k}^{-1}\frac{1}{n|\hat{G}_k|}\sum_{i=1}^{n}\sum_{t\in\hat{G}_k}\langle\tilde{\phi}_{j,k}, X_{it}^{cc}\rangle(y_{it}^c - \hat{\beta}_t^\top z_{it}^c).$$

Here $(\tilde{\lambda}_{j,k}, \tilde{\phi}_{j,k})_{1\le j\le n|\hat{G}_k|}$ denote the eigenvalue-eigenfunction pairs of the empirical covariance operator $\tilde{\Gamma}_k$, where $\tilde{\lambda}_{j,k}$ is the j-th largest eigenvalue. Again, for our theoretical analysis, we let $\tilde{m}_k = m_{k,nT} \to \infty$ as $n, T \to \infty$. In practice the cut-off parameter $\tilde{m}_k$ can be chosen, for instance, by CV or by a suitable information criterion as introduced in Section 1.5.

1.4 Asymptotic Theory

In the asymptotic analysis of our estimators we need to address two problems: first, there is a classification error contaminating the estimation of A_k. Second, the estimation of t-specific parameters β_t cannot be separated from the estimation of the regime specific parameter A_k. In the following we list our theoretical assumptions, allowing us to deal with these aspects in a large sample framework.

A1

1. For every $1 \le k \le K$, the random variables $\{(X_{it}, z_{it}, \epsilon_{it}) : 1 \le i \le n,\ t \in G_k\}$ are strictly stationary and further independent over the index i for any $t \in G_k$. Beyond that, the errors ϵ_{it} are centered and also independent over the index $1 \le t \le T$.
2. For every $1 \le k \le K$ and $1 \le i \le n$, the random variables $\{X_{it} : t \in G_k\}$ are L^4-m-approximable in the sense of Definition 2.1 in Hörmann and Kokoszka (2010).
3. For every $1 \le k \le K$ and $1 \le i \le n$, the random variables $\{z_{it} : t \in G_k\}$ are m-dependent.
4. Suppose that $||E[X_{it}^4]||_2 < \infty$, $E[z_{it}^4] < \infty$, $E[\epsilon_{it}^4] < \infty$ for any $1 \le i \le n$ and $1 \le t \le T$.
5. The error ϵ_{it} is independent of the covariates X_{js} and z_{js} for any $1 \le i, j \le n$ and $1 \le t, s \le T$.

A2 Suppose there exist constants $0 < C_\lambda, C'_\lambda, C_\theta, C_a, C_{zX}, C_\beta < \infty$, such that the following holds for every $1 \le k \le K$:

1. $C_\lambda^{-1} j^{-\mu} \le \lambda_{j,k} \le C_\lambda j^{-\mu}$ and $\lambda_{j,k} - \lambda_{j+1,k} \ge C'_\lambda j^{-(\mu+1)}$, $j \ge 1$ for the eigenvalues $\lambda_{1,k} > \lambda_{2,k} > \dots$ of the covariance operator Γ_k of X_{it}, $t \in G_k$ and a $\mu > 1$,

2. $E\left[\langle X_{it} - E[X_{it}], \phi_{j,k}\rangle^4\right] \leq C_\theta \lambda_{j,k}^2$ for the eigenfunction $\phi_{j,k}$ of Γ_k corresponding to the j-th largest eigenvalue $\lambda_{j,k}$, $j \geq 1$,
3. $|\langle A_k, \phi_{j,k}\rangle| \leq C_a j^{-\nu}$, $j \geq 1$,
4. $|\langle K_{z_p X,k}, \phi_{j,k}\rangle| \leq C_{zX} j^{-(\mu+\nu)}$, $j \geq 1$ for any $1 \leq p \leq P$, where $K_{z_p X,k} := E[(X_{it} - E[X_{it}])(z_{p,it} - E(z_{p,it}))]$ and
5. $\sup_{1\leq t\leq T} \beta_{p,t} \leq C_\beta$, for any $1 \leq p \leq P$, with $\beta_{p,t}$ being the p-th coordinate in β_t.

A3 Let $n \to \infty$ and $T \to \infty$ jointly, such that $T \propto n^\delta$ for some $0 < \delta < 1$ and $|G_k| \propto T$.

A4 Suppose that $\nu > 3\max\{r_1, r_2\}$, where $r_1 := 1 + \frac{1}{2}\mu$ and $r_2 := \frac{1+\mu(1+\delta)/3}{2(1-\delta)}$.

A5 Suppose that $m_t = m_{t,nT}$ and $\tilde{m}_k = \tilde{m}_{k,nT}$ with $m_t \propto n^{\frac{1}{\mu+2\nu}}$ and $\tilde{m}_k \propto (n|G_k|)^{\frac{1}{\mu+2\nu}}$ for any $1 \leq t \leq T$ and $1 \leq k \leq K$.

A6 Consider the random vector $\mathbf{s}_{it} := [s_{1,it}, \ldots, s_{P,it}]^\top$, defined according to

$$s_{p,it} := (z_{p,it} - E[z_{p,it}]) - \int_0^1 (X_{it}(u) - E[X_{it}](u)) \left(\sum_{j=1}^\infty \frac{\langle K_{z_p X,k}, \phi_{j,k}\rangle}{\lambda_{j,k}} \phi_{j,k}(u) \right) \mathrm{d}u, \ 1 \leq p \leq P.$$

Suppose that for any $1 \leq k \leq K$ the random variables $\{\mathbf{s}_{it} : \ 1 \leq i \leq n, \ t \in G_k\}$ are strictly stationary and further independent over the index i for any $t \in G_k$. Also, suppose they are strictly stationary, ergodic and m-dependent over the index t for any $1 \leq i \leq n$. In addition, assume that $E[\mathbf{s}_{it}|\mathbf{X}_k] = \mathbf{0}$, where $\mathbf{X}_k := \{X_{it} : \ 1 \leq i \leq n, \ t \in G_k\}$ and that the matrix $E[\mathbf{s}_{it}\mathbf{s}_{it}^\top]$ is positive semi-definite.

A7

1. There exists some $C_\Delta > 0$ such that for any $1 \leq k \leq K$ and any $t \in G_k$

$$\left|\left|\alpha_t^{(\Delta)} - \alpha_s^{(\Delta)}\right|\right|_2^2 =: \Delta_{ts} \begin{cases} \geq C_\Delta & \text{if } s \notin G_k \\ = 0 & \text{if } s \in G_k, \end{cases}$$

where $\alpha_r^{(\Delta)} := \sigma_{\epsilon,l}^{-1} \sum_{j=1}^\infty \lambda_{j,l}^{1/2} \langle \alpha_r, \phi_{j,l}\rangle \phi_{j,l}$ and $\sigma_{\epsilon,l}^2 := E[\epsilon_{ir}^2]$ for $r \in G_l$.

2. The threshold parameter $\tau_{nT} \to 0$ satisfies $\mathbb{P}\left(\max_{t,s\in G_k} \hat{\Delta}_{ts} \leq \tau_{nT}\right) \to 1$ as $n, T \to \infty$ for all $1 \leq k \leq K$.

Beyond the above assumptions we also suppose that the sign of the estimated eigenfunctions from Step 1 and Step 3 of our estimation procedure coincide with their population counterparts in the sense that $\int_0^1 \hat{\phi}_{j,t}(u)\phi_{j,k}(u)\mathrm{d}u \geq 0$, $1 \leq j \leq m_t$ and $\int_0^1 \tilde{\phi}_{j,k}(u)\phi_{j,k}(u)\mathrm{d}u \geq 0$, $1 \leq j \leq \tilde{m}_k$.

Assumptions A1-A6 correspond to existing standard assumptions in the literature (see Hall and Horowitz, 2007 and Shin, 2009), adapted to our panel data version of the partial functional linear regression model. Assumption A7 is a slightly modified version of Assumption C_τ in Vogt and Linton (2017).

Our theoretical results establish the consistency of our classification procedure and the convergence rates for the proposed regression slope estimators. We provide convergence rates of the period-wise estimators $\hat{\beta}_t$ and $\hat{\alpha}_t$ from Step 1 of our estimation procedure in Theorem 1.4.1. Lemma 1.4.1 establishes uniform consistency of these estimators as well as the adjusted slope function estimator $\hat{\alpha}_t^{(\Delta)}$ over $t = 1, \ldots, T$. This is an important prerequisite for the consistency of our classification procedure, which is established in Theorem 1.4.2. Finally, Theorem 1.4.3 establishes the convergence rate of our estimator $\tilde{A}_k$ from Step 3 of the estimation procedure.

Theorem 1.4.1 *Given Assumptions A1–A6 hold, it follows for all* $1 \leq t \leq T$ *that*

$$\begin{aligned} \left|\left|\hat{\beta}_t - \beta_t\right|\right|^2 &= O_p\left(n^{-1}\right) \\ \textit{and} \quad ||\hat{\alpha}_t - \alpha_t||_2^2 &= O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right), \end{aligned}$$

where $||.||$ *denotes the Euclidean norm and* $||.||_2$ *the* L^2 *norm.*

Theorem 1.4.1 is related to Theorems 3.1 and 3.2 in Shin (2009), though our proof deviates from that in Shin (2009) at important instances. The above rates for $\hat{\alpha}_t$ correspond to the rates in the cross-section context of Hall and Horowitz (2007). These pointwise rates provide a benchmark for the asymptotic properties of $\tilde{A}_k$, however the theorem is per se not sufficient for the consistency of our classification algorithm. For this, we need the following uniform consistency results:

Lemma 1.4.1 *Given Assumptions A1–A6 hold, it follows that*

$$\begin{aligned} \max_{1\leq t\leq T} \left|\left|\hat{\beta}_t - \beta_t\right|\right|^2 &= o_p(1) \\ \max_{1\leq t\leq T} ||\hat{\alpha}_t - \alpha_t||_2^2 &= o_p(1), \\ \textit{and} \quad \max_{1\leq t\leq T} \left|\left|\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}\right|\right|_2^2 &= o_p(1). \end{aligned}$$

The following theorem establishes consistency of our classification procedure and is based on our results in Lemma 1.4.1:

Theorem 1.4.2 *Given Assumptions A1–A7 hold, it follows that*

$$\mathbb{P}\left(\{\hat{G}_1, \ldots, \hat{G}_{\hat{K}}\} \neq \{G_1, \ldots, G_K\}\right) = o(1).$$

The statement of Theorem 1.4.2 is twofold. First, it says that the number of regimes K is asymptotically correctly determined. Second, it says that the estimators $\hat{G}_k$, $1 \leq k \leq \hat{K}$ consistently estimate their population counterparts. This notion of classification consistency is sufficient to obtain the following asymptotic result for the corresponding estimators $\tilde{A}_k$, $1 \leq k \leq \hat{K}$ from Step 3 of the estimation procedure:

Theorem 1.4.3 *Given Assumptions A1–A7 hold, it follows for all* $1 \leq k \leq \hat{K}$ *that*

$$\left|\left|\tilde{A}_k - A_k\right|\right|_2^2 = \begin{cases} O_p\left(n^{-1}\right) & \text{if} \quad \delta \geq \frac{1+\mu}{2\nu-1} \\ O_p\left((nT)^{\frac{1-2\nu}{\mu+2\nu}}\right) & \text{if} \quad \delta \leq \frac{1+\mu}{2\nu-1} \end{cases}.$$

Theorem 1.4.3 quantifies the extent to which the estimation error $||\hat{\beta}_t - \beta_t||$ contaminates the estimation of A_k. In the first case ($\delta \geq (1+\mu)/(2\nu-1)$), n diverges relatively slowly in comparison to T and, therefore, the contamination due to estimating β_t is not negligible. This results in the relatively slow convergence rate of $n^{-1/2}$, where the attribute "slow" has to be seen in relation to our panel context with $n \to \infty$ and $T \to \infty$. In the second case ($\delta \leq (1+\mu)/(2\nu-1)$), n diverges sufficiently fast such that the contamination due to estimating β_t becomes asymptotically negligible, which results in the faster convergence rate of $(nT)^{(1-2\nu)/(\mu+2\nu)}$. The latter rate coincides with the minimax optimal convergence result in Hall and Horowitz (2007).

1.5 Practical Choice of the Tuning Parameters

While we illuminated the large-sample properties in the previous section, we now turn to the further specification of our estimation procedure in finite samples.

Inspired by the thresholding procedure in Vogt and Linton (2017), we suggest choosing the threshold parameter τ_{nT} based on an approximate law for $\hat{\Delta}_{ts} = ||\hat{\alpha}_t^{(\Delta)} - \hat{\alpha}_s^{(\Delta)}||_2^2$ under the hypothesis that t and s belong to the same regime G_k. As argued in Section 1.A in the appendix of this chapter, the scaling of the estimators $\hat{\alpha}_t$ and $\hat{\alpha}_s$ as suggested in (1.3) leads, for large n, to

$$\frac{n}{2}\hat{\Delta}_{ts} = \frac{n}{2}||\hat{\alpha}_t^{(\Delta)} - \hat{\alpha}_s^{(\Delta)}||_2^2 \sim \chi^2_{\underline{m}} \quad \text{approximately}.$$

Hence we recommend setting the threshold τ_{nT} to be $2/n$ times the p_τ-quantile of a $\chi^2_{\underline{m}}$ distribution, where p_τ is close to one, for instance, $p_\tau = 0.99$ or $p_\tau = 0.999$. By scaling this quantile with $2/n$, the threshold converges to zero as n tends to infinity (see Section 1.A in the appendix for more details).

For selecting the truncation parameters m_t and $\tilde{m}_k$ the literature offers two general strategies.[2]

[2]Cf. Section 2.2 in Reiss et al. (2016) and references therein.

The first strategy is to choose the truncation parameters in order to find an optimal prediction. A cross validation procedure is shown, e.g., in Shin (2009). The second one is to choose the cut-off levels according to the covariance structure of the functional regressor. This is for large sample sizes particularly convenient from a computational point of view. We thus suggest choosing m_t and $\tilde{m}_k$ according to the eigenvalue ratio criterion suggested in Ahn and Horenstein (2013). This choice obtains according to

$$m_t = \arg\max_{1\leq l<n} \hat{\lambda}_{l,t}/\hat{\lambda}_{l+1,t}, \quad 1 \leq t \leq T$$
$$\text{and} \quad \tilde{m}_k = \arg\max_{1\leq l<n|\hat{G}_k|} \tilde{\lambda}_{l,k}/\tilde{\lambda}_{l+1,k}, \quad 1 \leq k \leq \hat{K}.$$

For selecting $K_{\max}$ we employ a standard estimate for the number of clusters from classical multivariate cluster analysis as introduced by Caliński and Harabasz (1974). This translates to our context as follows. On an equidistant grid $0 = s_1 < s_2 < \cdots < s_L = 1$ in $[0, 1]$ we calculate the L-vectors $v_t := [\hat{\alpha}_t^{(\Delta)}(s_l)]_{l=1,\ldots,L}$ for $1 \leq t \leq T$. Based on these quantities we employ the maximizer

$$K_{\max} := \arg\max_{1\leq k\leq(T-1)} \frac{tr\left(\sum_{j=1}^{k} |C_j|(v_t - \bar{v})(v_t - \bar{v})^\top\right)/(k-1)}{tr\left(\sum_{j=1}^{k}\sum_{t\in C_j}(v_t - c_j)(v_t - c_j)^\top\right)/(T-k)}$$

as an upper bound for $\hat{K}$. Here $C_j \subset \{1, \ldots, T\}$ is the j-th cluster formed from a k-means algorithm with c_j being the corresponding centroid. We further denote $\bar{v} := T^{-1}\sum_{t=1}^{T} v_t$ and use $tr(\cdot)$ for the trace operator. Choosing $K_{\max}$ equal to the optimal number of clusters in the multivariate analogue of the functional classification problem leads to a comparably conservative choice of $\hat{K} \leq K_{\max}$. This guarantees a parsimonious parameterization of our model in finite samples and improves the interpretability of the estimates.
We assess how this configuration of our estimation procedure performs in different finite-sample environments by means of a simulation study, which is described in the next section.

1.6 Simulations

In the following simulation study we consider two different data generating processes (Scenarios 1 and 2). In both scenarios there are $K = 2$ parameter regimes and we set $\alpha_t = A_1$ if $t \in G_1 =$

$\{1, \ldots, T/2\}$ and $\alpha_t = A_2$ if $t \in G_2 = \{T/2 + 1, \ldots, T\}$, where

$$A_1(u) = \begin{cases} \sqrt{2}\sin(\pi u/2) - u^3/2 + \sqrt{18}\sin(3\pi u/2) & \text{in Scenario 1} \\ 8u - 4u^2 - 5u^3 + 2\sin(8u) & \text{in Scenario 2} \end{cases}$$

$$A_2(u) = -2u + 8u^2 + 5u^3 + 2\sin(8u) \qquad \text{in Scenarios 1 and 2.}$$

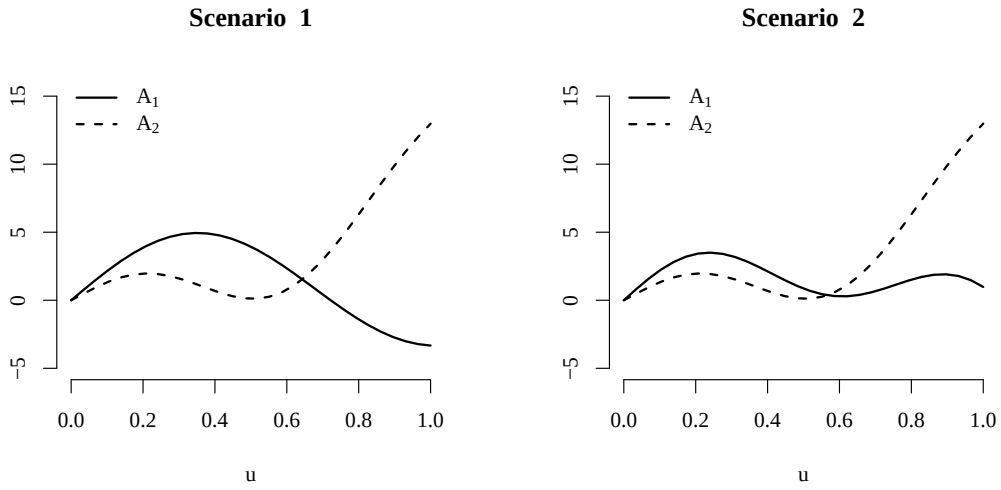

Figure 1.1: **Regime specific Parameter Functions A_1 and A_2 of the two different Scenarios.**

The graphs of the parameter functions are shown in Figure 1.1. Note that the distance between the regime specific slope functions A_1 and A_2 is smaller in Scenario 2 than in Scenario 1, which makes Scenario 2 the more challenging one.

For both scenarios we set $\beta_t = 5\sin(t/\pi)$ and $\rho_t = 5\cos(t/\pi)$. We simulate the regressor z_{it} and the error term ϵ_{it} according to $z_{it} \sim \mathcal{N}(0, 1)$ and $\epsilon_{it} \sim \mathcal{N}(0, 1)$. The trajectories X_{it} are obtained as $X_{it}(u) = \sum_{j=1}^{20} \theta_{it,j}\phi_j(u)$ with independent scores $\theta_{it,j} \sim \mathcal{N}\left(0, [(j - 1/2)\pi]^{-2}\right)$ and eigenfunctions $\phi_j(u) = \sqrt{2}\sin((j - 1/2)\pi u)$. Regarding the choice of the tuning parameters we proceed as described in Section 1.5. For selecting the threshold τ_{nT} we set $p_\tau = 0.99$.

In order to measure the precision of the classification procedure we calculate the classification error as the number of incorrectly classified periods t divided by T. We consider the following three different (n, T)-combinations: (i) $(n, T) = (50, 50)$, (ii) $(n, T) = (100, 50)$ and (iii) $(n, T) = (150, 80)$. For each specification we generate 1000 Monte Carlo samples. The results are reported in Table 1.1. The consistency of all parameter estimators as well as the accuracy of the classification procedure are well reflected in our simulation results. Parameter estimates improve with increasing n, given fixed T. The classification error is on a fairly low level, even for

$(\mathbf{n}, \mathbf{T}) = (\mathbf{50}, \mathbf{50})$	**Scenario 1**					**Scenario 2**				
	$q_{0.25}$	$q_{0.5}$	avg.	$q_{0.75}$	sd.	$q_{0.25}$	$q_{0.5}$	avg.	$q_{0.75}$	sd.
$T^{-1}\sum_{t=1}^{T}(\hat{\beta}_t - \beta_t)^2$	0.02	0.02	0.02	0.03	0.00	0.02	0.02	0.02	0.03	0.00
Classification Error	0.00	0.02	0.03	0.04	0.04	0.00	0.02	0.03	0.04	0.06
$\|\|\tilde{A}_1 - A_1\|\|_2^2/\|\|A_1\|\|_2^2$	0.01	0.01	0.02	0.03	0.03	0.05	0.07	0.10	0.11	0.10
$\|\|\tilde{A}_2 - A_2\|\|_2^2/\|\|A_2\|\|_2^2$	0.02	0.03	0.05	0.06	0.05	0.02	0.03	0.04	0.05	0.04

$(\mathbf{n}, \mathbf{T}) = (\mathbf{100}, \mathbf{50})$	**Scenario 1**					**Scenario 2**				
	$q_{0.25}$	$q_{0.5}$	avg.	$q_{0.75}$	sd.	$q_{0.25}$	$q_{0.5}$	avg.	$q_{0.75}$	sd.
$T^{-1}\sum_{t=1}^{T}(\hat{\beta}_t - \beta_t)^2$	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.00
Classification Error	0.00	0.00	0.02	0.04	0.04	0.00	0.00	0.02	0.02	0.04
$\|\|\tilde{A}_1 - A_1\|\|_2^2/\|\|A_1\|\|_2^2$	0.00	0.00	0.01	0.01	0.01	0.04	0.05	0.05	0.06	0.02
$\|\|\tilde{A}_2 - A_2\|\|_2^2/\|\|A_2\|\|_2^2$	0.02	0.02	0.04	0.04	0.04	0.01	0.02	0.03	0.03	0.02

$(\mathbf{n}, \mathbf{T}) = (\mathbf{150}, \mathbf{80})$	**Scenario 1**					**Scenario 2**				
	$q_{0.25}$	$q_{0.5}$	avg.	$q_{0.75}$	sd.	$q_{0.25}$	$q_{0.5}$	avg.	$q_{0.75}$	sd.
$T^{-1}\sum_{t=1}^{T}(\hat{\beta}_t - \beta_t)^2$	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.00
Classification Error	0.00	0.01	0.03	0.02	0.04	0.00	0.01	0.02	0.02	0.04
$\|\|\tilde{A}_1 - A_1\|\|_2^2/\|\|A_1\|\|_2^2$	0.00	0.00	0.00	0.01	0.00	0.04	0.04	0.04	0.05	0.01
$\|\|\tilde{A}_2 - A_2\|\|_2^2/\|\|A_2\|\|_2^2$	0.01	0.02	0.03	0.03	0.04	0.01	0.02	0.02	0.02	0.02

Table 1.1: **Simulation Results.** *The quantities $q_{0.25}$, $q_{0.5}$, $q_{0.75}$, "avg.", and "sd." denote the 25%, 50% and 75% quantiles, the arithmetic mean, and the standard deviation of the empirical distribution over Monte Carlo samples.*

the most challenging situations, in which the distance between the population parameters A_1 and A_2 is smaller and T is of the same magnitude as n.

1.7 Regime Dependent Pricing of Idiosyncratic Risk

Emerging from the influential work of Ang et al. (2006) a considerable number of studies confirm a negative cross-sectional correlation between idiosyncratic volatility and stock returns.[3] This finding is referred to as the "idiosyncratic volatility puzzle", since asset pricing theory suggests an opposite outcome. Either investors' portfolios are well diversified in equilibrium or investors are underdiversified. In the first case, idiosyncratic risk is diversified and the only risk to be priced is systematic. In the second case, idiosyncratic risk matters and investors with standard risk-return preferences asked for a premium to compensate for bearing this risk. Starting from theory it would thus be most reasonable to expect either no relation or a positive relation between idiosyncratic volatility and stock returns. As demonstrated in Hou and Loh (2016) the idiosyncratic volatility

[3] Cf. Fu (2009), Hou and Loh (2016) and references therein.

puzzle has, to a substantial extent, remained unsolved.

In the literature, the idiosyncratic volatility puzzle is typically examined using *aggregate monthly* measures of idiosyncratic volatility and returns. In contrast to that, we consider the relation between the daytime returns $y_{it} \in \mathbb{R}$ of assets $i = 1, \ldots, n$ and *disaggregate daily* idiosyncratic volatility measures $X_{it} \in L^2([0,1])$. Such disaggregate volatility curve X_{it} is intended to proxy an idiosyncratic component in asset price variation over continuous (standardized) intraday time $[0,1]$. Formally it is obtained as an empirical measure of the random spot volatility curve of *idiosyncratic* intraday stock price returns. The continuous time asset pricing framework, in which spot volatility is rigorously defined can be found, for instance, in Barndorff-Nielsen and Shephard (2002).[4] In the spirit of Ang et al. (2006), who estimate monthly varying risk premiums, we allow for daily varying marginal premiums on idiosyncratic volatility. We implement this, using the time-varying parameters in our novel panel regression model

$$y_{it} = \rho_t + \int_0^1 \alpha_t(s) X_{it}(s) \mathrm{d}s + \beta_t^\top z_{it} + \epsilon_{it}. \tag{1.4}$$

Here $\rho_t \in \mathbb{R}$ is a daily fixed effect and $\alpha_t \in L^2([0,1])$ denotes the time-varying parameter function describing the marginal premium on the idiosyncratic volatility curve $X_{it} \in L^2([0,1])$ at day t. The time-varying parameter vector $\beta_t \in \mathbb{R}^P$ describes the effect of additional control variables $z_{it} \in \mathbb{R}^P$. The term ϵ_{it} is a scalar error with zero mean and finite but potentially time heteroscedastic variances. We postulate that there are only $K < T$ different *risk pricing regimes* $G_1, \ldots, G_K$ collecting identical parameter functions α_t. As above, the common slope function of regime k is denoted by A_k. If a coefficient function A_k is clearly negative over most of its domain, the corresponding regime appears to be non-conform with traditional asset pricing theory and thus constitutes a temporary idiosyncratic volatility puzzle. Hence, the advantage of our model in (1.4) is its capability to segment the set of trading days into puzzling, non-puzzling and ambiguous pricing regimes in a data-driven way.

Following Fu (2009), we define the dependent variable as the daytime log-return

$$y_{it} := \log(P_{it}(1)/P_{it}(0)),$$

where $P_{it}(0)$ and $P_{it}(1)$ denote the opening price and closing price of asset i at day t. As control variable $z_{it} \in \mathbb{R}$ we use the daily average bid-ask spread which serves as a proxy for liquidity risk—an important pricing-relevant factor as discussed for example in Hou and Loh (2016).[5]

[4] See also Chapter 2 of this dissertation.

[5] See also Chapter 3 of this dissertation for a discussion of liquidity risk in the corporate bond market.

Data. We consider intraday price data for $n = 377$ stocks listed on the S&P 500. Our sample consists of $T = 136$ trading days between June 3, 2016, and December 15, 2016. Intraday stock prices are sampled every $\Delta = 10$ minutes during the trading hours of the S&P 500. More concretely, for asset i at day t we consider the last recorded transaction prices, $P_{it}(s_j)$, within 10-minute intervals, of which the j-th interval is denoted $[s_{j-1}, s_j]$ with $s_0 = 0$, $s_J = 1$, where $1 \leq j \leq J = 39$. For the construction of the idiosyncratic volatility curves $X_{it}(.)$, which is described below, we make use of three Fama-French factors. The Fama-French factors were downloaded from Kenneth French's homepage, while all other data were gathered from Bloomberg.[6]

Preprocessing. For constructing the idiosyncratic volatility curves $X_{it}(.)$ we use the method proposed in Müller et al. (2011) with a straightforward adaption to our context for estimating *idiosyncratic* volatility curves instead of *total* volatility curves. Müller et al. (2011) propose an estimator of the total volatility curve of asset i at day t, smoothing the points $(s_j, \tilde{Y}_{it,j})$, $j = 1, \ldots, J$ based on an algorithm, which allows to obtain functional principal components from discrete noisy data. Here $\tilde{Y}_{it,j} := \log(\Delta^{-1} Y_{it}(s_j)^2) + q_0$ are scaled and logarithmized versions of the squared intraday returns $Y_{it}(s_j)^2$ with $Y_{it}(s_j) := \log(P_{it}(s_j)/P_{it}(s_{j-1}))$. The points $\tilde{Y}_{it,j}$, $1 \leq j \leq J$ can be interpreted as noisy evaluations of the random spot volatility curve of the underlying continuous time return process at the corresponding points s_j, $1 \leq j \leq J$. The constant $q_0 = 1.27$ is necessary for re-centering the involved error term, whereas technical details can be found in Müller et al. (2011). We mainly proceed along the lines of their approach, however, instead of using the total intraday returns $Y_{it}(s_j)$, $j = 1, \ldots, J$, we employ the *idiosyncratic* intraday return components $Y^*_{it}(s_j)$, $j = 1, \ldots, J$. This leads to a measure of the *idiosyncratic* volatility curve $X_{it}(.)$ rather than a measure of the total volatility curve. For computing the idiosyncratic intraday returns $Y^*_{it}(s_j)$, we follow the usual approach and correct the total intraday returns $Y_{it}(s_j)$ for their systematic market component by regressing them on three Fama-French factors (cf. Fama and French, 1995). We do so by estimating the functional Fama-French regression model

$$Y_{it}(s_j) = b_{0,it}(s_j) + b_{1,it} \cdot M_t(s_j) + b_{2,it} \cdot S_t + b_{3,it} \cdot H_t + u_{it}(s_j), \quad j = 1, \ldots, J, \tag{1.5}$$

which was proposed by Kokoszka et al. (2014). The term $M_t(s_j)$ is the intraday S&P 500 market return, S_t denotes the "small minus large" factor and H_t the "high minus low" factor. S_t describes the difference in returns between portfolios of small and large stocks and H_t describes the difference in returns between portfolios of high and low book-to-market value stocks. For estimating the model parameters in (1.5) we use the least-squares estimators proposed by Kokoszka et al. (2014). The idiosyncratic intraday returns are finally obtained as $Y^*_{it}(s_j) = \hat{b}_{0,it}(s_j) + \hat{u}_{it}(s_j)$, where $\hat{b}_{0,it}(s)$ denotes the fitted functional intercept parameter of the (i,t)-th regression and $\hat{u}_{it}(s)$

[6] We thank Kenneth French for making this data publicly available on his homepage.

are the corresponding function-valued regression residuals.
Table 1.2 provides summary statistics for our sample. Figure 1.2 shows the idiosyncratic volatility curves X_{it}, along with their raw scatter points, for the Apple stock at two randomly selected trading days.

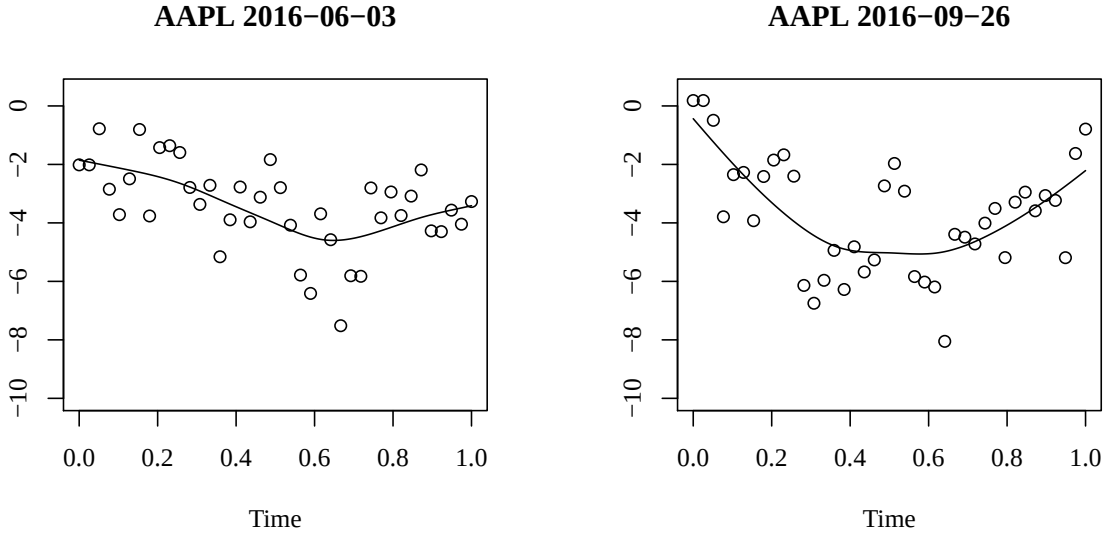

Figure 1.2: **Exemplary Volatility Curves.** *Idiosyncratic volatility curves X_{it} and raw scatter points for the Apple Inc. stock (AAPL) at two randomly selected trading days.*

Remark. Applying the method of Müller et al. (2011) in order to estimate the idiosyncratic volatility curves X_{it} from the idiosyncratic intraday returns $Y^*_{it}(s_j)$, $j = 1, \ldots, J$, leads to "volatility curves", which are indeed logarithmic volatility curves (cf. Eq. (9) in Müller et al., 2011). The log-transformation is monotonic and is thus sign-preserving what concerns the estimation of α_t. This is of particular importance when assessing the idiosyncratic volatility puzzle. Working with log-transformed volatility objects is generally advisable, since the raw volatility measures are often heavily skewed (cf. Herskovic et al., 2016).

	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	avg.	sd.
y (in %)	-1.74	-0.56	0.01	0.60	1.79	0.02	1.16
$\int_0^1 X(u)\mathrm{d}u$	-4.42	-3.72	-3.20	-2.63	-1.64	-3.14	0.85
$\|\|X\|\|_2^2$	3.46	7.67	11.02	14.67	20.50	11.39	5.19
z (in %)	0.02	0.02	0.03	0.05	0.09	0.04	0.03

Table 1.2: **Summary Statistics.** *Quantiles, means, and standard deviations of the considered variables.*

The estimation of model (1.4) proceeds as described in Section 1.3. Using our classification algorithm, we find a number of $\hat{K} = 2$ regimes, where each regime contains about the same number of days (see right panel in Figure 1.3). Interestingly, regime memberships do not show a striking temporal pattern, a view which is supported by the empirical transition probabilities, which are slightly above 0.5 as reported in Table 1.3. The left panel in Figure 1.3 shows the estimated regime

k/k	1	2
1	0.43	0.57
2	0.58	0.42

Table 1.3: **Empirical Relative Transition Frequencies.** *Table reports the empirical relative transition frequencies, indicating the empirical probability for a trading day (column) to be in the same/another regime as the previous day (row). The index k refers to the k-th estimated regime.*

specific slope functions $\tilde{A}_1$ and $\tilde{A}_2$. In order to examine how idiosyncratic volatility is priced in the daytime returns, we consider *aggregate marginal effects* defined according to $\int_0^1 \tilde{A}_k(u)\mathrm{d}u$, $k = 1, 2$. For the first regime this marginal effect is clearly negative, for the second one clearly positive. Our classification thus separates trading days revealing an idiosyncratic volatility puzzle from days which are conform with asset pricing theory. Both parameter functions, however, indicate that the intensity of the pricing varies over trading time within a day.

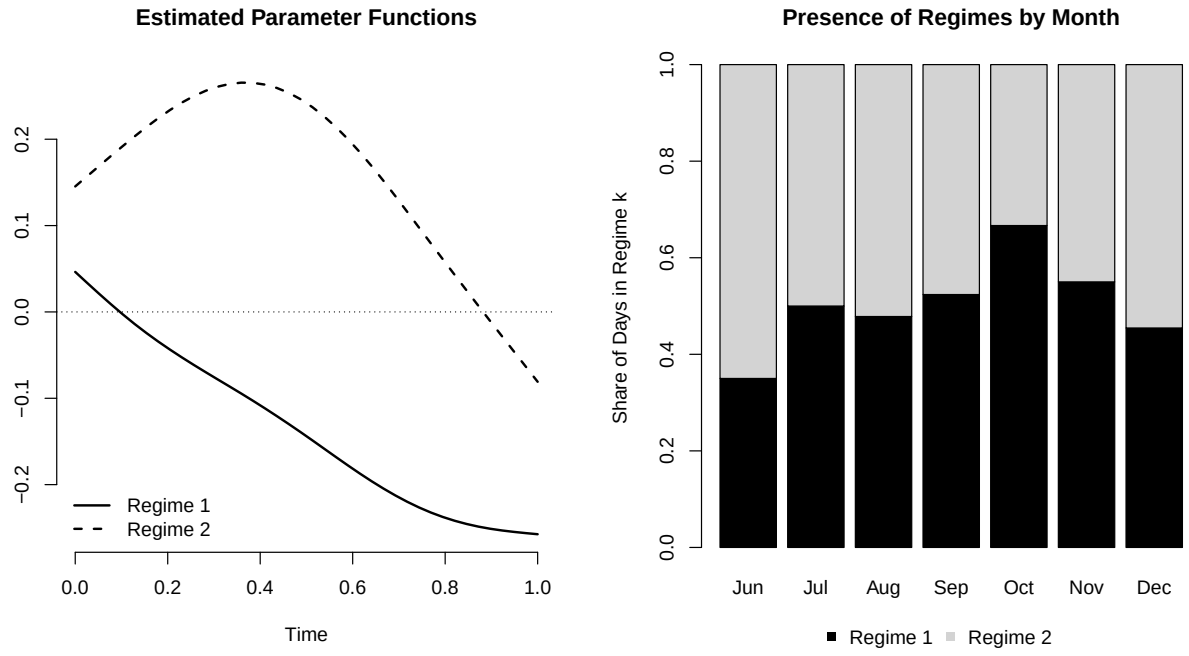

Figure 1.3: **Estimation Results.** *Estimated regime specific slope functions $\tilde{A}_1$ and $\tilde{A}_2$ (left panel) and marginal effect of idiosyncratic volatility (right panel).*

In summary, our results indicate a rather complex pattern of puzzling and non-puzzling days. The high resolution of the time scale our novel model operates on, allows us to uncover a much more heterogeneous pricing of idiosyncratic risk than one can infer from monthly data for example (see Ang et al., 2006 for instance). Hence, aggregating the data could thus misleadingly convolute the puzzling and non-puzzling pricing mechanisms. This in turn might contribute to the failure of current explanations of the idiosyncratic volatility puzzle (cf. Hou and Loh, 2016).

1.8 Conclusion

In this chapter we present a novel regression framework, which allows us to examine regime specific effects of a random function on a scalar response in the presence of a multivariate regressor and time fixed effects. The suggested estimation procedure is designed for a panel data context. We prove consistency of the estimators including rates of convergence and address the practical choice of the tuning parameters involved. In summary, our framework offers a very flexible and data-driven way of assessing heterogeneity in large panels. Our model could be extended in multiple directions for further research. For instance, establishing a connection to the work of Su et al. (2016) would allow us to identify latent group structures in the cross section in addition to identifying latent time-regime structures.

The statistical model is motivated by our real data application, where we explore a phenomenon referred to as the idiosyncratic volatility puzzle. In an empirical study we search for the presence of such a puzzle in a large panel of US stock prices. Our method allows us to separate puzzling days from non-puzzling days. The results suggest a much more heterogeneous pricing of idiosyncratic risk than indicated by many existing analyses in the literature.

Appendix

In Part 1.A of this appendix we provide formal proofs for Theorems 1.4.1, 1.4.2, and 1.4.3 as well as Lemma 1.4.1. Further, we briefly discuss the properties of the threshold τ_{nT} as suggested in Section 1.5. In Part 1.B we provide additional results from our simulation study.

Throughout this appendix we use the symbols C and c to denote positive constants. Their precise meaning is allowed to vary from term to term.

1.A Technical Appendix

In this part we use the following notation for norms in addition to the ones introduced in the main body of the chapter. Given a mapping $F_1 : L^2([0,1]) \to \mathbb{R}$, we use as norm of F_1 the operator norm $||F_1||_{H'} := \sup_{||f_1||_2=1} |F_1(f_1)|$. Further, for an integral operator $F_2 : L^2([0,1]) \to L^2([0,1])$ with kernel $f_2 \in L^2([0,1] \times [0,1])$, denote its Hilbert-Schmidt norm as $||F_2||_{\mathcal{S}} := ||f_2||_2$, where in this case $|| \cdot ||_2$ is the L^2 norm in $L^2([0,1] \times [0,1])$.
For the sake of readability we will proof the lemma and theorems for $P = 1$, while the generalization to $P > 1$ is straightforward and does not add any additional insights. In this spirit we ease our notation by dropping boldface notation and the dependence on coordinate labels p.

Now, turning to a formal argumentation, we begin collecting a number of basic results readily available in the functional data literature. Provided Assumption 1 holds, the random variables $\{(z_{it}, X_{it}, \epsilon_{it}) : \ 1 \leq i \leq n\}$ are iid with finite fourth moments for every $1 \leq t \leq T$. Moment calculations as well as the results in Hörmann and Kokoszka (2010) imply for any $1 \leq t \leq T$ as

$n \to \infty$ that

$$E\left[\left|\left|\hat{K}_{zX,t} - K_{zX,k}\right|\right|_2^2\right] = O(n^{-1}) \tag{1.6}$$

$$E\left[\left|\hat{K}_{z,t} - K_{z,k}\right|^2\right] = O(n^{-1}) \tag{1.7}$$

$$E\left[\left|\left|\hat{K}_{X,t} - K_{X,k}\right|\right|_2^2\right] = O(n^{-1}), \tag{1.8}$$

where the index k is such that $t \in G_k$, which we use in what follows without further reference. In Equation (1.8) $K_{X,k}$ denotes the covariance function in the k-th regime, i.e. $K_{X,k}(u,v) := E[(X_{it}(u) - E[X_{it}](u))(X_{it}(v) - E[X_{it}](v))]$ and in analogy $K_{z,k} := E[(z_{it} - E[z_{it}])^2]$. Further, it obviously holds that

$$\begin{aligned}
&E\left[|\bar{z}_t - E[z_{it}]|^2\right] = O(n^{-1}) \\
&E\left[\left|\left|\bar{X}_t - E[X_{it}]\right|\right|_2^2\right] = O(n^{-1}), \\
&E\left[||\hat{K}_{X\epsilon,t}||_2^2\right] = O(n^{-1}) \\
&E\left[|\hat{K}_{z\epsilon,t}|^2\right] = O(n^{-1})
\end{aligned}$$

where

$$\hat{K}_{X\epsilon,t} := n^{-1}\sum_{i=1}^{n} X_{it}^c \epsilon_{it}^c$$

$$\text{and} \quad \hat{K}_{z\epsilon,t} := n^{-1}\sum_{i=1}^{n} z_{it}^c \epsilon_{it}^c.$$

Denote the Hilbert-Schmidt norm of the distance between t-wise empirical covariance operator and population covariance operator as $\mathcal{D}_t := ||\hat{\Gamma}_t - \Gamma_k||_{\mathcal{S}}$. Note that for any $1 \leq j \leq n$, $|\hat{\lambda}_{j,t} - \lambda_{j,k}| \leq \mathcal{D}_t$ almost surely (see Theorem 1 in Hall and Hosseini–Nasab, 2006 and references therein). Since $E[\mathcal{D}_t^q] = O(n^{-q/2})$ for $q = 1, 2, \ldots$ (provided sufficiently high moments exist) it holds that

$$E\left[\left|\hat{\lambda}_{j,t} - \lambda_{j,k}\right|^q\right] = O(n^{-q/2}) \;\; q = 1, 2 \ldots \tag{1.9}$$

for any $1 \leq j \leq m_t$ (cf. Equation A.11 in Kneip et al., 2016).

As a final observation, note that combining the results in Shin (2009) and Hall and Horowitz

(2007) allows to conclude that for any $1 \leq t \leq T$

$$||\hat{\Phi}_t - \Phi_k||^2_{H'} = \left|\left|\sum_{j=1}^{m} \frac{\langle \hat{K}_{zX,t}, \hat{\phi}_{j,t}\rangle}{\hat{\lambda}_{j,t}} \hat{\phi}_{j,t} - \sum_{j=1}^{\infty} \frac{\langle K_{zX,k}, \phi_{j,k}\rangle}{\lambda_{j,k}} \phi_{j,k}\right|\right|^2_2$$
$$= O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right), \tag{1.10}$$

where we denote $m_t = m$ for simplicity, which we continue to do without further reference. The mapping $\Phi_k : L^2([0,1]) \to \mathbb{R}$ is the population counterpart of $\hat{\Phi}_t$ and was implicitly used already in Assumption 6. It is formally defined according to

$$\Phi_k(g) := \sum_{j=1}^{\infty} \frac{\langle K_{zX,k}, \phi_{j,k}\rangle}{\lambda_{j,k}} \langle \phi_{j,k}, g\rangle \tag{1.11}$$

for any $g \in L^2([0,1])$.

Proof of Theorem 1.4.1

Consider any $1 \leq t \leq T$, with t in some regime k, i.e. $t \in G_k$. Note that the estimator $\hat{\beta}_t$ can be written as

$$\hat{\beta}_t = \hat{B}_t^{-1}[\hat{K}_{zy,t} - \hat{\Phi}_t(\hat{K}_{yX,t})]$$

with $\hat{B}_t := [\hat{K}_{z,t} - \hat{\Phi}_t(\hat{K}_{zX,t})]$. Regarding the inverse in $\hat{\beta}_t$ note that it follows from (1.6), (1.7), and (1.10) in analogy to Shin (2009) that

$$\hat{B}_t := [\hat{K}_{z,t} - \hat{\Phi}_t(\hat{K}_{zX,t})] \xrightarrow{\mathbb{P}} [K_{z,k} - \Phi_k(K_{zX,k})] =: B_k > 0$$

as $n \to \infty$, which certainly implies $\hat{B}_t^{-1} = B_k^{-1} + o_p(1)$ by the continuous mapping theorem, whereas $B_k = E[\mathbf{s}_{it}^2] > 0$ follows from Assumption 6. To see this also consider the decomposition shown in (1.25). As in Shin (2009), we assess the difference

$$\hat{\beta}_t - \beta_t = \hat{B}_t^{-1}\left[n^{-1}\sum_{i=1}^{n}\left(z_{it}^c - \hat{\Phi}_t(X_{it}^c)\right)(\langle X_{it}^c, \alpha_t\rangle + \epsilon_{it}^c)\right]$$

by splitting the term $n^{-1}\sum_{i=1}^{n}\left(z_{it}^c - \hat{\Phi}_t(X_{it}^c)\right)(\langle X_{it}^c, \alpha_t\rangle + \epsilon_{it}^c)$ according to

$$\left| n^{-1} \sum_{i=1}^{n} \left(z_{it}^c - \hat{\Phi}_t(X_{it}^c) \right) \left(\langle X_{it}^c, \alpha_t \rangle + \epsilon_{it}^c \right) \right| \leq |R_{0,1,t}| + |R_{0,2,t}| + |R_{0,3,t}|,$$

where, in analogy to her work,

$$\begin{aligned} R_{0,1,t} &:= n^{-1} \sum_{i=1}^{n} (z_{it}^c - \Phi_k(X_{it}^c)) \epsilon_{it}^c = O_p(n^{-1/2}) \\ R_{0,2,t} &:= n^{-1} \sum_{i=1}^{n} (\Phi_k(X_{it}^c) - \hat{\Phi}_t(X_{it}^c)) \epsilon_{it}^c = O_p(n^{-1/2}). \end{aligned}$$

due to the exogeneity of the covariates and the assumed iid nature of the error term (cf. Assumption 1). However, the remaining term we approach in a different manner:

$$\begin{aligned} |R_{0,3,t}| &:= \left| n^{-1} \sum_{i=1}^{n} (z_{it}^c - \hat{\Phi}_t(X_{it}^c)) \langle X_{it}^c, \alpha_t \rangle \right| \\ &\leq \left| \langle \hat{K}_{zX,t} - K_{zX,k}, \alpha_t \rangle \right| + \left| \langle K_{zX,k}, \alpha_t \rangle - n^{-1} \sum_{i=1}^{n} \hat{\Phi}_t(X_{it}^c) \langle X_{it}^c, \alpha_t \rangle \right| \\ &\leq R_{1,1,t} + R_{1,2,t} \end{aligned}$$

where for $R_{1,1,t}$

$$\begin{aligned} R_{1,1,t} &:= \left| \langle \hat{K}_{zX,t} - K_{zX,k}, \alpha_t \rangle \right| \\ &\leq ||\alpha_t||_2 \cdot ||\hat{K}_{zX,t} - K_{zX,k}||_2 \\ &= O_p(n^{-1/2}) \end{aligned}$$

as a consequence of (1.6). The second term, $R_{1,2,t}$, in $R_{0,3,t}$ is defined as

$$\begin{aligned} R_{1,2,t} &:= \left| \langle K_{zX,k}, \alpha_t \rangle - n^{-1} \sum_{i=1}^{n} \hat{\Phi}_t(X_{it}^c) \langle X_{it}^c, \alpha_t \rangle \right| \\ &\leq R_{2,1} + R_{2,2,t}, \end{aligned}$$

with

$$R_{2,1} := \left| \sum_{j=m+1}^{\infty} \langle K_{zX,k}, \phi_{j,k} \rangle a^*_{j,t} \right|$$
$$R_{2,2,t} := \left| \sum_{j=1}^{m} \langle K_{zX,k}, \phi_{j,k} \rangle a^*_{j,t} - \sum_{j=1}^{m} \langle \hat{K}_{zX,t}, \hat{\phi}_{j,t} \rangle \langle \hat{\phi}_{j,t}, \alpha_t \rangle \right|,$$

where we used $a^*_{j,t} := \langle \alpha_t, \phi_{j,k} \rangle$ due to Assumptions 2, 4 and 5. For the first term observe $R_{2,1} = O\left(n^{\frac{1-\mu-2\nu}{\mu+2\nu}}\right) = O(n^{-1/2})$. The second one can be split in three parts

$$R_{2,2,t} \leq R_{3,1,t} + R_{3,2,t} + R_{3,3,t}$$

with

$$R_{3,1,t} := ||\hat{K}_{zX,t} - K_{zX,k}||_2 \sum_{j=1}^{m} \left(||\hat{\phi}_{j,t} - \phi_{j,k}||_2 \cdot ||\alpha_t||_2 + |a^*_{j,t}| \right),$$
$$R_{3,2,t} := ||\alpha_t||_2 \sum_{j=1}^{m} |\langle K_{zX,k}, \phi_{j,k} \rangle| \cdot ||\hat{\phi}_{j,t} - \phi_{j,k}||_2$$

and

$$R_{3,3,t} := ||K_{zX,k}||_2 \cdot ||\alpha_t||_2 \sum_{j=1}^{m} ||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 + ||K_{zX,k}||_2 \sum_{j=1}^{m} ||\hat{\phi}_{j,t} - \phi_{j,k}||_2 \cdot |a^*_{j,t}|.$$

An assessment of the asymptotic properties of $R_{3,1,t}, R_{3,2,t}$ and $R_{3,3,t}$ requires to examine the asymptotic properties of $||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2$ explicitly. Bounds can, for example, be obtained from Theorem 1 in Hall and Hosseini–Nasab (2006) as

$$||\hat{\phi}_{t,j} - \phi_{j,k}||_2^q \leq \left[\frac{8^{1/2} \mathcal{D}_t}{\min_{1 \leq l \leq j} \{\lambda_{j,k} - \lambda_{j+1,k}\})} \right]^q \quad \text{almost surely} \tag{1.12}$$

which holds for $1 \leq j \leq m$, $q = 1, 2, \ldots$ and any size n of the cross section (see also Equation (5.2) in Hall and Horowitz, 2007). In the context of theory for functional linear regression, Hall and Horowitz (2007) develop *asymptotic* bounds on $||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2$, $1 \leq j \leq m$, which are valid on events which occur with probability tending to one as $n \to \infty$. These bounds are particularly helpful, when addressing (weighted) sums over estimation errors as they appear e.g. in $R_{3,1,t}$–$R_{3,3,t}$. We will make use of these bounds, slightly adapting the arguments in Hall and Horowitz (2007), in order to formulate the result more explicitly. For this purpose we consider the three events

1. $\mathcal{F}_{1,n,t} := \left\{ Cn^{\frac{2(1+\mu)}{\mu+2\nu}} \mathcal{D}_t^2 \leq 1/8 \right\}$

2. $\mathcal{F}_{2,n,t} := \left\{ |\hat{\lambda}_{j,t} - \lambda_{l,k}|^{-2} \leq 2|\lambda_{j,k} - \lambda_{l,k}|^{-2} \leq Cn^{\frac{2(1+\mu)}{\mu+2\nu}},\ 1 \leq j \leq m, j \neq l \in \mathbb{N} \right\}$.

3. $\mathcal{F}_{3,n,t} := \mathcal{F}_{1,n,t} \cap \mathcal{F}_{2,n,t}$

of which the second coincides with their work and the first one is a straightforward derivative of their arguments. Denoting the complement of a set A as A^c, note that $\mathbb{P}(\mathcal{F}_{1,n,t}^c) = o(1)$ as well as $\mathbb{P}(\mathcal{F}_{2,n,t}^c) = o(1)$ due Assumptions 4–5 and root-n consistency of the empirical covariance operator and its corresponding eigenvalues as well as assuming the constants in $\mathcal{F}_{1,n,t}$ and $\mathcal{F}_{2,n,t}$ to be appropriate. Since $\mathbb{P}(\mathcal{F}_{3,n,t}^c) \leq \mathbb{P}(\mathcal{F}_{1,n,t}^c) + \mathbb{P}(\mathcal{F}_{2,n,t}^c)$, we conclude $\mathbb{P}(\mathcal{F}_{3,n,t}^c) = o(1)$. We also show that this property holds uniformly over $1 \leq t \leq T$ as $(n, T) \to \infty$ in the proof of Lemma 1.4.1 below. Equation (5.21) in Hall and Horowitz (2007), reads in our notation as

$$||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 \leq 8 \left(1 - 4Cn^{\frac{2(1+\mu)}{\mu+2\nu}} \mathcal{D}_t^2\right)^{-1} R_{j,t}^{(\phi)}, \tag{1.13}$$

$$\text{where } R_{j,t}^{(\phi)} := \sum_{l:l\neq j} (\lambda_{j,k} - \lambda_{l,k})^{-2} \left[\int_0^1 \int_0^1 (\hat{K}_{X,t}(u,v) - K_{X,k}(u,v)) \phi_{j,k}(u) \phi_{l,k}(v) \mathrm{d}u \mathrm{d}v \right]^2.$$

The inequality in (1.13) is valid on $\mathcal{F}_{2,n,t}$, whereas the constant C on the right hand side is the constant in $\mathcal{F}_{1,n,t}$. On this event $\mathcal{F}_{1,n,t}$ it further holds that

$$\left(1 - 4Cn^{\frac{2(1+\mu)}{\mu+2\nu}} \mathcal{D}_t^2\right)^{-1} \leq 2$$

which implies, that on $\mathcal{F}_{3,n,t}$, it holds that

$$||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 \leq 16 R_{j,t}^{(\phi)}. \tag{1.14}$$

Note that Equation (5.22) in Hall and Horowitz (2007) states that

$$E\left[R_{j,t}^{(\phi)}\right] = O\left(j^2 n^{-1}\right) \tag{1.15}$$

uniformly in $1 \leq j \leq m$ (see also the corresponding proof of Equation (5.22) in Section 5.3 in Hall and Horowitz, 2007). Note that (1.14) obviously implies that on $\mathcal{F}_{3,n,t}$,

$$||\hat{\phi}_{j,t} - \phi_{j,k}||_2 \leq 4\left(R_{j,t}^{(\phi)}\right)^{\frac{1}{2}} \tag{1.16}$$

of which the right hand side has the property $E\left[\left(R_{j,t}^{(\phi)}\right)^{1/2}\right] \leq E\left[R_{j,t}^{(\phi)}\right]^{1/2} = O\left(jn^{-1/2}\right)$ uni-

formly over $1 \leq j \leq m$, what follows from Jensen's inequality and (1.15).
These observations imply that

$$\mathbb{P}\left(nm^{-3}\sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2^2 > c\right) \leq \mathbb{P}\left(16nm^{-3}\sum_{j=1}^{m}R_{j,t}^{(\phi)} > c\right) + \mathbb{P}\left(\mathcal{F}_{3,n,t}^{c}\right)$$
$$\leq \frac{nm^{-3}\sum_{j=1}^{m}E\left[R_{j,t}^{(\phi)}\right]}{c/16} + o(1) \tag{1.17}$$

by the Markov inequality. The numerator on the right hand side of (1.17) is bounded above as a consequence of (1.15) and Assumptions 4 & 5, and thus $\sum_{j=1}^{m}||\phi_{j,t}-\phi_{j,k}||_2^2 = O_p\left(n^{-1}m^3\right)$. From this and Assumptions 4 & 5, of which the former is slightly stronger than in Hall and Horowitz (2007) and Shin (2009), we conclude for the first summand in $R_{3,3,t}$,

$$\begin{aligned}||K_{zX,k}||_2 \cdot ||\alpha_t||_2 \sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2^2 &= O_p(n^{-1}m^3)\\ &= O_p\left(n^{\frac{3-\mu-2\nu}{\mu+2\nu}}\right)\\ &= O_p(n^{-1/2})\end{aligned}$$

because $\nu > 3-\mu/2$. Note that from our observations for (1.16), we can further conclude

$$\sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2 = O_p(n^{-1/2}m^2)$$

using similar arguments as before. We have $n^{-1/2}m^2 = n^{\frac{2-\mu/2-\nu}{\mu+2\nu}} = o(1)$ by Assumption 4, which allows to conclude in combination with (1.6) and Assumption 2, that $R_{3,1,t} = O_p(n^{-1/2})$.
Using similar arguments as for (1.17), allows us to conclude for the second term in $R_{3,3,t}$:

$$\mathbb{P}\left(n^{1/2}\sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2 \cdot |a_{j,t}^{*}| > c\right) \leq \mathbb{P}\left(4n^{1/2}\sum_{j=1}^{m}\left(R_{j,t}^{(\phi)}\right)^{1/2}C_a j^{-\nu} > c\right) + \mathbb{P}\left(\mathcal{F}_{3,n,t}^{c}\right)$$
$$\leq \frac{n^{1/2}\sum_{j=1}^{m}E\left[R_{j,t}^{(\phi)}\right]^{1/2}j^{-\nu}}{c/(4C_a)} + o(1),$$

where the numerator on the right hand side of the last inequality is bounded above thanks to Assumptions 4–5 as well as our observation in (1.16). An analogue argument shows $R_{3,2,t} = O_p(n^{-1/2})$ (see also points 3 and 4 in Assumption 2 to see this).
Combining arguments implies $\hat{\beta}_t - \beta_t = O_p(n^{-1/2})$ for every $1 \leq t \leq T$, which concludes the

proof of the first result in Theorem 1.4.1. Turning to $\hat{\alpha}_t$ note that

$$||\hat{\alpha}_t - \alpha_t||_2^2 \leq 3\sum_{j=1}^{m}(\hat{a}_{j,t} - a_{j,t}^*)^2 + 3m\sum_{j=1}^{m}(a_{j,t}^*)^2||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 + 3\sum_{j=m+1}^{\infty}(a_{j,t}^*)^2.$$

The results in Hall and Horowitz (2007) and Shin (2009) immediately translate to $m\sum_{j=1}^{m}(a_{j,t}^*)^2||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2$ and $\sum_{j=m+1}^{\infty} a_{j,t}^*$ which are both $O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right)$. The remaining term can be split according to

$$\sum_{j=1}^{m}(\hat{a}_{j,t} - a_{j,t}^*)^2 \leq 2\sum_{j=1}^{m}(\hat{\lambda}_{j,t}^{-1}\langle \hat{K}_{yX,t}^{\#} - \hat{\beta}_t\hat{K}_{zX,t}^{\#}, \hat{\phi}_{j,t}\rangle - a_{j,t}^*)^2 + 2\sum_{j=1}^{m}(\hat{\lambda}_{j,t}^{-1}\langle r_{y,t}r_{x,t} - \hat{\beta}_t r_{z,t}r_{x,t}, \hat{\phi}_{j,t}\rangle)^2 \tag{1.18}$$

with $\hat{K}_{yX,t}^{\#} := n^{-1}\sum_{i=1}^{n}(y_{it} - E[y_{it}])(X_{it} - E[X_{it}])$, $\hat{K}_{zX,t}^{\#} := n^{-1}\sum_{i=1}^{n}(z_{it} - E[z_{it}])(X_{it} - E[X_{it}])$, $r_{x,t} := E[X_{it}] - \bar{X}_t$, $r_{y,t} := E[y_{it}] - \bar{y}_t$ and $r_{z,t} := E[z_{it}] - \bar{z}_t$. Note that $||r_{x,t}||_2$, $|r_{y,t}|$ and $|r_{z,t}|$ all correspond to errors from parametric estimation problems and are thus of order $n^{-1/2}$. Bounds on $\hat{\lambda}_{j,t} - \lambda_{j,k}$ as well as $||\hat{\phi}_{j,t} - \phi_{j,k}||_2$ are asymptotically equivalent for data centered around their arithmetic mean and data centered around their population expectation. Together with the above arguments it follows that the first term in (1.18) is asymptotically equivalent to the corresponding term in Shin (2009), implying $\sum_{j=1}^{m}(\hat{\lambda}_{j,t}^{-1}\langle \hat{K}_{yX,t}^{\#} - \hat{\beta}_t\hat{K}_{zX,t}^{\#}, \hat{\phi}_{j,t}\rangle - a_{j,t}^*)^2 = O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right)$. Now, define the event

$$\mathcal{F}_{4,n,t} := \{|\hat{\lambda}_{j,t} - \lambda_{j,k}| < \lambda_{j,k}/2 : \ 1 \leq j \leq m\}$$

for which we conclude $\mathbb{P}(\mathcal{F}_{4,n,t}^c) = o(1)$ for any $1 \leq t \leq T$ as $n \to \infty$ as a consequence of (1.9). On this event the second term in (1.18) can be bounded according to

$$\begin{aligned}\sum_{j=1}^{m}(\hat{\lambda}_{j,t}^{-1}\langle r_{y,t}r_{X,t} - \hat{\beta}_t r_{z,t}r_{X,t}, \hat{\phi}_{j,t}\rangle)^2 &\leq 8\sum_{j=1}^{m}\lambda_{j,k}^{-2}r_{y,t}^2||r_{X,t}||_2^2 + 8\sum_{j=1}^{m}\lambda_{j,k}^{-2}\hat{\beta}_t^2 r_{z,t}^2||r_{X,t}||_2^2 \\ &= O_p\left(n^{\frac{1+2\mu-2\mu-4\nu}{\mu+2\nu}}\right) = o_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right).\end{aligned}$$

Finally combining arguments yields $||\hat{\alpha}_t - \alpha_t||_2^2 = O_p(n^{\frac{1-2\nu}{\mu+2\nu}})$ for any $1 \leq t \leq T$ as $n \to \infty$, which concludes the proof of the second part of Theorem 1.4.1. ■

Proof of Lemma 1.4.1

In what follows we show that the quantities $\hat{\alpha}_t^{(\Delta)}$ are consistent for $\alpha_t^{(\Delta)}$ in the L^2 norm, uniformly over $1 \leq t \leq T$. The remaining claims in the Lemma are required for this result to hold and are validated en route.

We begin introducing additional notation and listing a number of basic observations, which are a consequence of the iid sampling scheme in the cross-section as well as stationarity of the regressors and the error over time within regimes. Note that since the random variables $\{(X_{it}, z_{it}, \epsilon_{it})\, t \in G_k,\ 1 \leq i \leq n\}$ are stationary, expectations of the below statistics calculated from these random variables do not vary over index t for a given regime k. In order to reduce the complexity of our notation, however, we do not make this invariance explicit in every step. For the following properties we also use the results in Hall and Horowitz (2007) and Hörmann and Kokoszka (2010).

- Based on the above convention for our notation, we conclude, using the results in Hörmann and Kokoszka (2010), our first observation:

$$\mathbb{P}\left(\max_{1\leq t\leq T} \mathcal{D}_t^2 > c\right) \leq \sum_{k=1}^{K} \sum_{t\in G_k} \mathbb{P}\left(\mathcal{D}_t^2 > c\right) \leq K \max_{1\leq k\leq K} |G_k| \frac{E\left[\mathcal{D}_t^2\right]}{c} = O(n^{\delta-1}) = o(1),$$

  since $|G_k| \propto T \propto n^\delta$ according to Assumption 3, which we will use in what follows without reference.

- Further, empirical variances of z_{it} and ϵ_{it} behave according to

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq t\leq T} |\hat{K}_{z,t} - K_{z,k}|^2 > c\right) \\
&\leq K \max_{1\leq k\leq K} |G_k| \frac{n^{-1} E\left[|(z_{it} - E[z_{it}])^2 - K_{z,k}|^2\right]}{c} + K \max_{1\leq k\leq K} |G_k| \frac{E\left[(\bar{z}_t - E[z_{it}])^4\right]}{c} \\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1) \qquad (1.19)
\end{aligned}$$

  and similarly

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq t\leq T} \left| n^{-1} \sum_{i=1}^{n} (\epsilon_{it} - \bar{\epsilon}_t)^2 - \sigma_\epsilon^2 \right| > c\right) &\leq K \max_{1\leq k\leq K} |G_k| \frac{n^{-1} E\left[(\epsilon_{it}^2 - \sigma_\epsilon^2)^2\right]}{c} + K \max_{1\leq k\leq K} |G_k| \frac{E\left[\bar{\epsilon}_t^4\right]}{c} \\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1).
\end{aligned}$$

- In analogy to before introduce $\hat{K}^{\#}_{z\epsilon,t} := n^{-1} \sum_{i=1}^{n} (z_{it} - E[z_{it}])(\epsilon_{it} - E[\epsilon_{it}])$ and $\hat{K}^{\#}_{X\epsilon,t}(u) := n^{-1} \sum_{i=1}^{n} (X_{it}(u) - E[X_{it}](u))\epsilon_{it}$ as well as $r_{\epsilon,t} := \bar{\epsilon}_t$. It follows from simple moment

calculations for the cross sectional empirical covariances between regressors and error that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq t\leq T}||\hat{K}_{X\epsilon,t}||_2^2 > c\right) \leq \sum_{t=1}^{T}\mathbb{P}\left(||\hat{K}^{\#}_{X\epsilon,t}||_2^2 > c/4\right) + \sum_{t=1}^{T}\mathbb{P}\left(||r_{x,t}||_2^2 r_{\epsilon,t}^2 > c/4\right) \\
&\leq K\max_{1\leq k\leq K}|G_k|\frac{n^{-1}\sigma_{\epsilon,k}^2 E\left[||X_{it}-E[X_{it}]||_2^2\right]}{c} + K\max_{1\leq k\leq K}|G_k|\frac{E[(\bar{\epsilon}_t)^2]E\left[||\bar{X}_t - E[X_{it}]||_2^2\right]}{c} \\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1) \qquad (1.20)
\end{aligned}$$

Similar arguments can be used to show

$$\mathbb{P}\left(\max_{1\leq t\leq T}|\hat{K}_{z\epsilon,t}|^2 > c\right) = O(n^{\delta-1}) + O(n^{\delta-2}) = o(1). \qquad (1.21)$$

- Uniform consistency of the empirical covariance $\hat{K}_{zX,t}(u)$ can be shown with similar arguments according to

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq t\leq T}||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c\right) \\
\leq &\sum_{t=1}^{T}\mathbb{P}\left(||\hat{K}^{\#}_{zX,t} - K_{zX,k}||_2^2 > c\right) + \sum_{t=1}^{T}\mathbb{P}\left(||r_{x,t}||_2^2 r_{z,t}^2 > c\right) \\
\leq &K\max_{1\leq k\leq K}|G_k|\frac{n^{-1}E\left[||(z_{it}-E[z_{it}])(X_{it}-E[X_{it}]) - K_{zX,k}||_2^2\right]}{c} \\
&+ K\max_{1\leq k\leq K}|G_k|\frac{E\left[r_{z,t}^2\right]E\left[||r_{x,t}||_2^2\right]}{c} \\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1). \qquad (1.22)
\end{aligned}$$

- Beyond the above observations, the following part of the proof requires the term $\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\phi_{j,k} - \hat{\phi}_{j,t}||_2^2$ to vanish in probability, uniformly over $1 \leq t \leq T$.

  To see this, note that $E[R_{j,t}^{(\phi)}]$ as in (1.15) does not vary over the index $t \in G_k$ within a regime k, but potentially across regimes $k = 1, \ldots, K$. This due to the stationarity of the

functional regressor within regimes as postulated in Assumption 1. We thus conclude:

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\leq t\leq T}\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\phi_{j,k}-\hat{\phi}_{j,t}||_2^2>c\right)\\
\leq&\sum_{t=1}^{T}\mathbb{P}\left(\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\phi_{j,k}-\hat{\phi}_{j,t}||_2^2>c\right)\\
\leq&\sum_{t=1}^{T}\left(\mathbb{P}\left(16\cdot 4C_{\lambda}^{-2}\sum_{j=1}^{m}j^{2\mu}R_{j,t}^{(\phi)}>c\right)+\mathbb{P}\left(\mathcal{F}_{1,n,t}^{c}\right)+\mathbb{P}\left(\mathcal{F}_{2,n,t}^{c}\right)+\mathbb{P}\left(\mathcal{F}_{4,n,t}^{c}\right)\right)\\
\leq&K\max_{1\leq k\leq K}|G_k|\frac{\sum_{j=1}^{m}E\left[R_{j,t}^{(\phi)}\right]j^{2\mu}}{c\cdot C_{\lambda}^{2}/(16\cdot 4)}+\sum_{t=1}^{T}\left(\mathbb{P}\left(\mathcal{F}_{1,n,t}^{c}\right)+\mathbb{P}\left(\mathcal{F}_{2,n,t}^{c}\right)+\mathbb{P}\left(\mathcal{F}_{4,n,t}^{c}\right)\right)\\
=&O\left(Tn^{-1}m^{3+2\mu}\right)+\sum_{t=1}^{T}\left(\mathbb{P}\left(\mathcal{F}_{1,n,t}^{c}\right)+\mathbb{P}\left(\mathcal{F}_{2,n,t}^{c}\right)+\mathbb{P}\left(\mathcal{F}_{4,n,t}^{c}\right)\right),
\end{aligned}
\tag{1.23}
$$

where C_λ is the constant from point 1 in Assumption 2. To obtain the second inequality, we used once more that $\hat{\lambda}_{j,t}\geq\lambda_{j,k}/2$ for $1\leq j\leq m$ on $\mathcal{F}_{4,n,t}$. The sequence in (1.23) is a null sequence because on the one hand

$$
Tn^{-1}m^{3+2\mu}=O\left(n^{\frac{3+(1+\delta)\mu-2(1-\delta)\nu}{\mu+2\nu}}\right)=o(1)
$$

thanks to Assumption 4 and on the other hand since

$$
\sum_{t=1}^{T}\mathbb{P}\left(\mathcal{F}_{l,n,t}^{c}\right)\leq K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{F}_{l,n,t}^{c}\right)=o\left(1\right),\ l=1,2,4
$$

as we argue next. First we observe

$$
\begin{aligned}
K\max_{1\leq k\leq K}|G_k|\mathbb{P}(\mathcal{F}_{1,n,t}^{c})=&K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(Cn^{\frac{2(1+\mu)}{\mu+2\nu}}\mathcal{D}_t^2>1/8\right)\\
&\leq K\max_{1\leq k\leq K}|G_k|8Cn^{\frac{2(1+\mu)}{\mu+2\nu}}E[\mathcal{D}_t^2]\\
&=O\left(n^{\frac{2+(1+\delta)\mu-2(1-\delta)\nu)}{\mu+2\nu}}\right)=o(1)
\end{aligned}
$$

for any $C > 0$ (cf. Assumption 4). Second, we argue that

$$\begin{aligned}
K \max_{1\leq k\leq K} |G_k|\mathbb{P}(\mathcal{F}^c_{2,n,t}) &= K \max_{1\leq k\leq K} |G_k|\mathbb{P}(\exists 1 \leq j \leq m,\ j \neq l:\ |\hat{\lambda}_{j,t} - \lambda_{l,k}|^{-2} > 4|\lambda_{j,k} - \lambda_{l,k}|^{-2}) \\
&= K \max_{1\leq k\leq K} |G_k|\mathbb{P}(\exists 1 \leq j \leq m,\ j \neq l:\ |\hat{\lambda}_{j,t} - \lambda_{l,k}| < \frac{1}{2}|\lambda_{j,k} - \lambda_{l,k}|) \\
&\leq K \max_{1\leq k\leq K} |G_k|\mathbb{P}(\exists 1 \leq j \leq m,\ j \neq l:\ |\hat{\lambda}_{j,t} - \lambda_{j,k}| > \frac{1}{2}|\lambda_{j,k} - \lambda_{l,k}|) \\
&\leq K \max_{1\leq k\leq K} |G_k|\mathbb{P}(\mathcal{D}_t > \frac{1}{2}\min\{\lambda_{j,k} - \lambda_{j+1,k}, \lambda_{j-1,k} - \lambda_{j,k}\}) \\
&\leq K \max_{1\leq k\leq K} |G_k|\frac{E[\mathcal{D}_t^2]}{\min\{\lambda_{j,k} - \lambda_{j+1,k}, \lambda_{j-1,k} - \lambda_{j,k}\}^2/4} \\
&= O\left(Tn^{-1}m^{2(1+\mu)}\right) \\
&= o(1)
\end{aligned}$$

by the fact that $\mathcal{D}_t \geq \sup_j |\hat{\lambda}_{j,t} - \lambda_{j,k}|$ almost surely as well as Assumptions 2–5. In lines of our arguments from the proof of Theorem 1.4.1, we conclude $K \max_{1\leq k\leq K} |G_k|\mathbb{P}(\mathcal{F}^c_{3,n,t}) = o(1)$. Beyond that it holds

$$\begin{aligned}
K \max_{1\leq k\leq K} |G_k|\mathbb{P}(\mathcal{F}^c_{4,n,t}) &\leq K \max_{1\leq k\leq K} |G_k|\mathbb{P}\left(\sup_{1\leq j\leq m} |\hat{\lambda}_{j,t} - \lambda_{j,k}| > \frac{1}{2}\lambda_{m,k}\right) \\
&\leq K \max_{1\leq k\leq K} |G_k|\mathbb{P}\left(\mathcal{D}_t > \frac{1}{2}\lambda_{m,k}\right) \\
&\leq K \max_{1\leq k\leq K} |G_k|\frac{4E\left[\mathcal{D}_t^2\right]}{\lambda_{m,k}^2} \\
&= O\left(n^{\delta}n^{\frac{\mu-2\nu}{\mu+2\nu}}\right) = o(1)
\end{aligned}$$

again thanks to Assumption 4. Note that our result in (1.23) implies in particular that

$$\mathbb{P}\left(\max_{1\leq t\leq T}\sum_{j=1}^{m} ||\phi_{j,k} - \hat{\phi}_{j,t}||_2^2 > c\right) = o(1),$$

which will be used without further reference in what follows.

- As a last observation, we note that $\max_{1\leq t\leq T} ||\alpha_t||_2 = \max_{1\leq k\leq K} ||A_k||_2$ is a constant and does not vary in t and neither in k.

Now, turning to our concrete arguments for $\hat{\alpha}_t^{(\Delta)}$, we note that the scaling which distinguishes $\hat{\alpha}_t^{(\Delta)}$ from $\hat{\alpha}_t$ is composed by $\hat{\sigma}_{\epsilon,t}$ and the empirical eigenvalues $\hat{\lambda}_{j,t}$,$1 \leq j \leq m$. While the latter

can be treated in a comparably simple way, $\hat{\sigma}_{\epsilon,t}$ requires closer attention. We thus begin focusing on this object and its constituents. For this purpose, define the event $\mathcal{S}_{n,t}$ for later use according to

$$\mathcal{S}_{n,t} := \left\{ \left| \hat{\sigma}^2_{\epsilon,t} - \sigma^2_{\epsilon,k} \right| \leq \frac{1}{2} \sigma^2_{\epsilon,k} \right\}.$$

We show in a moment that $\sum_{t=1}^{T} \mathbb{P}(\mathcal{S}^c_{n,t}) = o(1)$. However this requires some preparation since $\hat{\sigma}^2_{\epsilon,t}$ includes estimation errors from $\hat{\beta}_t$ and $\hat{\alpha}_t$. We thus start arguing that (i) $\mathbb{P}(\max_{1\leq t\leq T} |\hat{\beta}_t - \beta_t| > c) = o(1)$ and (ii) $\mathbb{P}(\max_{1\leq t\leq T} ||\hat{\alpha}_t - \alpha_t||_2 > c) = o(1)$, as claimed in the lemma. Turning to the first point, note that the estimator $\hat{\beta}_t$ makes multiple use of the operator $\hat{\Phi}_t$, which can, starting from the Riesz-Frechet representation Theorem (cf. Shin, 2009), be handled according to

$$\left|\left| \hat{\Phi}_t - \Phi_k \right|\right|^2_{H'} = 3R_{4,1,t} + 3R_{4,2,t} + 3R_{4,3}.$$

The last summand is defined as $R_{4,3} := \left|\left| \sum_{j=m+1}^{\infty} \frac{\langle K_{zX,k}, \phi_{j,k} \rangle}{\lambda_{j,k}} \phi_{j,k} \right|\right|^2_2$, which is independent of t and $o(1)$ because the truncation parameter diverges at infinity and hence $R_{4,3}$ is arbitrarily small for n large enough. The remaining summands are defined and handled as follows. For the first one we observe that

$$\begin{aligned}
R_{4,1,t} :=& \left|\left| \sum_{j=1}^{m} \left(\frac{\langle \hat{K}_{zX,t}, \hat{\phi}_{j,t} \rangle}{\hat{\lambda}_{j,t}} - \frac{\langle K_{zX,k}, \phi_{j,k} \rangle}{\lambda_{j,k}} \right) \hat{\phi}_{j,t} \right|\right|^2_2 \\
\leq& 2 \sum_{j=1}^{m} (\hat{\lambda}_{j,t} \lambda_{j,k})^{-2} \left[\langle \lambda_{j,k} \hat{K}_{zX,t} - \hat{\lambda}_{j,t} K_{zX,k}, \phi_{j,k} \rangle + \langle \lambda_{j,k} \hat{K}_{zX,t}, (\hat{\phi}_{j,t} - \phi_{j,k}) \rangle \right]^2 \\
\leq& 4 \sum_{j=1}^{m} (\hat{\lambda}_{j,t} \lambda_{j,k})^{-2} \left[\langle K_{zX,k}, \phi_{j,k} \rangle^2 (\lambda_{j,k} - \hat{\lambda}_{j,t})^2 + \langle \hat{K}_{zX,t} - K_{zX,k}, \phi_{j,k} \rangle^2 \lambda^2_{j,k} \right] \\
&+ 2 \sum_{j=1}^{m} (\hat{\lambda}_{j,t})^{-2} \langle \hat{K}_{zX,t}, (\hat{\phi}_{j,t} - \phi_{j,k}) \rangle^2 \\
\leq& \underbrace{4 \sum_{j=1}^{m} (\hat{\lambda}_{j,t} \lambda_{j,k})^{-2} \langle K_{zX,k}, \phi_{j,k} \rangle^2 (\lambda_{j,k} - \hat{\lambda}_{j,t})^2}_{=:R_{5,1,t}} + \underbrace{4 \sum_{j=1}^{m} ||\hat{K}_{zX,t} - K_{zX,k}||^2_2 \hat{\lambda}^{-2}_{j,t}}_{=:R_{5,2,t}} \\
&+ \underbrace{2 \sum_{j=1}^{m} \hat{\lambda}^{-2}_{j,t} ||\hat{K}_{zX,t}||^2_2 ||\hat{\phi}_{j,t} - \phi_{j,k}||^2_2}_{=:R_{5,3,t}}.
\end{aligned}$$

For the three summands $R_{5,1,t}, R_{5,2,t}, R_{5,3,t}$ we use our above observation as well Assumptions 1–5 to conclude the following:

Ad $R_{5,1,t}$:

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} (\hat{\lambda}_{j,t}\lambda_{j,k})^{-2} \langle K_{zX,k}, \phi_{j,k}\rangle^2 (\lambda_{j,k} - \hat{\lambda}_{j,t})^2 > c\right) \\
\leq & K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(4\sum_{j=1}^{m} \lambda_{j,k}^{-4} \langle K_{zX,k}, \phi_{j,k}\rangle^2 (\lambda_{j,k} - \hat{\lambda}_{j,t})^2 > c\right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right) \\
\leq & K \max_{1\leq k\leq K} |G_k| \frac{E[\mathcal{D}_t^2] \sum_{j=1}^{m} \lambda_{j,k}^{-4} \langle K_{zX,k}, \phi_{j,k}\rangle^2}{c/4} + o(1) \\
= & O(n^{\delta-1}) + o(1) = o(1).
\end{aligned}$$

Ad $R_{5,2,t}$:

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} ||\hat{K}_{zX,t} - K_{zX,k}||_2^2 \hat{\lambda}_{j,t}^{-2} > c\right) \\
&\leq K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(4\sum_{j=1}^{m} ||\hat{K}_{zX,t} - K_{zX,k}||_2^2 \lambda_{j,k}^{-2} > c\right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}(\mathcal{F}_{4,n,t}^c) \\
&\leq K \max_{1\leq k\leq K} |G_k| \frac{\sum_{j=1}^{m} \lambda_{j,k}^{-2} E\left[||\hat{K}_{zX,t} - K_{zX,k}||_2^2\right]}{c/4} + o(1) \\
&= O\left(n^{\frac{1+(1+\delta)\mu-2(1-\delta)\nu}{\mu+2\nu}}\right) + o(1) \\
&= o(1).
\end{aligned}$$

Ad $R_{5,3,t}$:

$$\begin{aligned}
&\sum_{t=1}^{T}\mathbb{P}\left(\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\hat{K}_{zX,t}||_2^2||\hat{\phi}_{j,t}-\phi_{j,k}||_2^2>c\right)\\
\leq& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2||K_{zX,k}||_2^2\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\hat{\phi}_{j,t}-\phi_{j,k}||_2^2>c^{1/2}\right)\\
&+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2||\hat{K}_{zX,t}-K_{zX,k}||_2^2>c^{1/4}\right)\\
&+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\hat{\phi}_{j,t}-\phi_{j,k}||_2^2>c^{1/4}\right)\\
=&o(1).
\end{aligned}$$

Since $\mathbb{P}\left(\max_{1\leq t\leq T}R_{4,1,t}>c\right)\leq\sum_{l=1}^{3}\sum_{t=1}^{T}\mathbb{P}\left(R_{5,l,t}>c/3\right)$, it follows that $\mathbb{P}\left(\max_{1\leq t\leq T}R_{4,1,t}>c\right)=o(1)$. For $R_{4,2,t}$, defined as

$$R_{4,2,t}:=\left|\left|\sum_{j=1}^{m}\frac{\langle K_{zX,k},\phi_{j,k}\rangle}{\lambda_{j,k}}(\phi_{j,k}-\hat{\phi}_{j,t})\right|\right|_2^2,$$

we note this expression can be most easily handled using the almost sure bound in (1.12) according to

$$\begin{aligned}
&K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\left|\left|\sum_{j=1}^{m}\frac{\langle K_{zX,k},\phi_{j,k}\rangle}{\lambda_{j,k}}(\phi_{j,k}-\hat{\phi}_{j,t})\right|\right|_2^2>c\right)\\
\leq& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(m\sum_{j=1}^{m}\frac{\langle K_{zX,k},\phi_{j,k}\rangle^2}{\lambda_{j,k}^2}\left|\left|\phi_{j,k}-\hat{\phi}_{j,t}\right|\right|_2^2>c\right)\\
&K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(m\mathcal{D}_t^2\sum_{j=1}^{m}\frac{\langle K_{zX,k},\phi_{j,k}\rangle^2}{\lambda_{j,k}^2}\frac{j^{2(1+\mu)}}{C'_\lambda}>c\right)\\
\leq& K\max_{1\leq k\leq K}|G_k|\frac{mE[\mathcal{D}_t^2]\sum_{j=1}^{m}\langle K_{zX,k},\phi_{j,k}\rangle^2 j^{2(1+\mu)}/\lambda_{j,k}^2}{c\cdot C'_\lambda}\\
=&O\left(mn^{\delta-1}\right)=o(1).
\end{aligned}$$

thanks to Assumption 2–5. In particular, these results imply

$$\mathbb{P}\left(\max_{1\leq t\leq T}||\hat{\Phi}_t-\Phi_k||_{H'}^2>c\right)=o(1). \tag{1.24}$$

To proceed we work again on the differences $(\hat{\beta}_t - \beta_t) = \hat{B}_t^{-1}\left(R_{0,1,t} + R_{0,2,t} + R_{0,3,t}\right)$ with $R_{0,1,t}$, $R_{0,2,t}$ and $R_{0,3,t}$ as in the proof of Theorem 1.4.1. Addressing the inverse in these differences, define the t-wise event $Q_{n,t} := \{|\hat{B}_t - B_k| \leq \frac{1}{2}B_k\}$. For this event, note that $\sum_{t=1}^{T}\mathbb{P}\left(Q_{n,t}^c\right) \leq R_{6,1} + R_{6,2}$, where $R_{6,1} := \sum_{t=1}^{T}\mathbb{P}\left(|\hat{K}_{z,t} - K_{z,k}|^2 > c\right) = o(1)$ as shown in (1.19). For $R_{6,2}$ we use the arguments in Shin (2009) to obtain

$$R_{6,2} := \sum_{t=1}^{T}\mathbb{P}\left(||\hat{\Phi}_t - \Phi_k||_{H'}^2||K_{zX,k}||_2^2 + (||\hat{\Phi}_t - \Phi_k||_{H'} + ||\Phi_k||_{H'})^2||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c\right) \tag{1.25}$$

$$\begin{aligned}
&\leq \underbrace{K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||\hat{\Phi}_t - \Phi_k||_{H'}^2||K_{zX,k}||_2^2 > c\right)}_{=:R_{7,1}} \\
&\quad + \underbrace{K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2||\hat{\Phi}_t - \Phi_k||_{H'}^2||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c^{1/2}\right)}_{=:R_{7,2}} \\
&\quad + \underbrace{K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2||\Phi_k||_{H'}^2||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c^{1/2}\right)}_{=:R_{7,3}}.
\end{aligned}$$

As shown before $R_{7,1}, R_{7,3} = o(1)$. Further

$$\begin{aligned}
R_{7,2} &\leq K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||\hat{\Phi}_t - \Phi_k||_{H'}^2||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c^{1/2}\right) \\
&\leq K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||\hat{\Phi}_t - \Phi_k||_{H'}^2 > c^{1/4}\right) + K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c^{1/4}\right) \\
&= o(1).
\end{aligned}$$

For uniform consistency of $\hat{\beta}_t$ it remains to show that

- $\mathbb{P}(\max_{1\leq t\leq T}|R_{0,1,t}| > c) = o(1)$,
- $\mathbb{P}(\max_{1\leq t\leq T}|R_{0,2,t}| > c) = o(1)$ and
- $\mathbb{P}(\max_{1\leq t\leq T}|R_{0,3,t}| > c) = o(1)$.

For this we argue

$$\begin{aligned}\mathbb{P}\left(\max_{1\le t\le T}|R_{0,1,t}|>c\right)&=\mathbb{P}\left(\max_{1\le t\le T}\left|n^{-1}\sum_{i=1}^{n}(z_{it}^{c}-\Phi_k(X_{it}^{c}))\epsilon_{it}^{c}\right|>c\right)\\&\le K\max_{1\le k\le K}|G_k|\mathbb{P}\left(\left|\hat{K}_{z\epsilon,t}\right|^2>c^2/4\right)+K\max_{1\le k\le K}|G_k|\mathbb{P}\left(||\Phi_k||_{H'}^2||\hat{K}_{\epsilon X,t}||_2^2>c^2/4\right)\\&=o(1)\end{aligned}$$

due to (1.20) and (1.21). Further note for $R_{0,2,t}$

$$\begin{aligned}\mathbb{P}\left(\max_{1\le t\le T}|R_{0,2,t}|>c\right)&=\mathbb{P}\left(\max_{1\le t\le T}\left|\hat{\Phi}_t(\hat{K}_{\epsilon X,t})-\Phi_k(\hat{K}_{\epsilon X,t})\right|>c\right)\\&\le K\max_{1\le k\le K}|G_k|\mathbb{P}\left(||\hat{\Phi}_t-\Phi_k||_{H'}||\hat{K}_{\epsilon X,t}||_2>c\right)\\&\le K\max_{1\le k\le K}|G_k|\mathbb{P}\left(||\hat{K}_{\epsilon X,t}||_2^2>c\right)\\&\quad+K\max_{1\le k\le K}|G_k|\mathbb{P}\left(||\hat{\Phi}_t-\Phi_k||_{H'}^2>c\right)\\&=o(1)\end{aligned}$$

as a consequence of (1.20) and (1.24). For the remaining terms, we argue along the same lines as in the proof of Theorem 1.4.1:

$$\mathbb{P}\left(\max_{1\le t\le T}|R_{0,3,t}|>c\right)\le\mathbb{P}\left(\max_{1\le t\le T}|R_{1,1,t}|>c\right)+\mathbb{P}\left(\max_{1\le t\le T}|R_{1,2,t}|>c\right)$$

$$\begin{aligned}\mathbb{P}\left(\max_{1\le t\le T}|R_{1,1,t}|>c\right)&\le\mathbb{P}\left(\max_{1\le t\le T}||\alpha_t||_2\cdot||\hat{K}_{zX,t}-K_{zX}||_2>c\right)\\&=o(1)\end{aligned}$$

because of (1.22). The remaining term was shown to be bounded according to $R_{1,2}\le R_{2,1,t}+R_{2,2,t}$, where the two summands are defined above. While $R_{2,1}=O(n^{-1/2})$ deterministically and independently of t, note for the second summand $R_{2,2,t}\le R_{3,1,t}+R_{3,2,t}+R_{3,3,t}$ as before and

further:

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\leq t\leq T}|R_{3,1,t}|>c\right)\\
\leq& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\sum_{j=1}^{m}||\hat{K}_{zX,t}-K_{zX,k}||_2(||\hat{\phi}_{j,t}-\phi_{j,k}||_2\cdot||A_k||_2+|a^*_{j,t}|)>c\right)\\
\leq& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2||\hat{K}_{zX,t}-K_{zX,k}||_2\cdot||A_k||_2\sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2>c\right)\\
&+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2||\hat{K}_{zX,t}-K_{zX,k}||_2\sum_{j=1}^{m}|a^*_{j,t}|>c\right)\\
\leq& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2||\hat{K}_{zX,t}-K_{zX,k}||_2\cdot||A_k||_2 4\sum_{j=1}^{m}\left(R^{(\phi)}_{j,t}\right)^{1/2}>c\right)\\
&+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{F}^c_{3,n,t}\right)+o(1)\\
\leq& K\max_{1\leq k\leq K}|G_k|\frac{||A_k||_2^2\cdot E\left[||\hat{K}_{zX,t}-K_{zX,k}||_2^2\right]^{\frac{1}{2}}m^{\frac{1}{2}}\left(\sum_{j=1}^{m}E\left[R^{(\phi)}_{j,t}\right]\right)^{\frac{1}{2}}}{c}+o(1)+o(1)\\
=& O(n^{\delta-1}m^2)+o(1)=o(1),
\end{aligned}
$$

due to Assumptions 2–5 and our above observations. Further for $R_{3,2,t}$ similar arguments yield:

$$
\begin{aligned}
\mathbb{P}\left(\max_{1\leq t\leq T}|R_{3,2,t}|>c\right)\leq& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||A_k||_2\sum_{j=1}^{m}|\langle K_{zX,k},\phi_{j,k}\rangle|\cdot||\hat{\phi}_{j,t}-\phi_{j,k}||_2>c\right)\\
\leq& K\max_{1\leq k\leq K}|G_k|\frac{||A_k||_2^2\cdot 16\cdot C_{zX}^2 m\sum_{j=1}^{m}j^{-2(\mu+\nu)}E\left[R^{(\phi)}_{j,t}\right]}{c}+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{F}^c_{3,n,t}\right)\\
=& O(mn^{\delta-1})+o(1)=o(1).
\end{aligned}
$$

Similarly, we argue for $R_{3,3,t}$,

$$
\begin{aligned}
\mathbb{P}\left(\max_{1\leq t\leq T}|R_{3,3,t}|>c\right)\leq& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||K_{zX,k}||_2||A_k||_2\sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2^2>c\right)\\
&+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||K_{zX,k}||_2\sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2|a^*_{j,t}|>c\right)\\
=& o(1)
\end{aligned}
$$

where the first term is a null sequence as implied by (1.23). The second term is of the order

$O(n^{\delta-1}m) = o(1)$ which follows from analogous arguments as used for $R_{3,2,t}$.
Combining our above arguments, we conclude $\mathbb{P}(\max_{1\leq t\leq T}(\hat{\beta}_t - \beta_t)^2 > c) = o(1)$ as claimed in the lemma.

Now, turning to the estimation error in $\hat{\alpha}_t$ we employ upper bounds

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq t\leq T}||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) &\leq K \max_{1\leq k\leq K}|G_k|\mathbb{P}\left(||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) \\ &\leq R_{8,1} + R_{8,2} + R_{8,3} + R_{8,4}.\end{aligned}$$

While the four summands on the right and side are defined below, the term $\sum_{j=m+1}^{\infty} {a_{j,t}^*}^2$ does not appear in the upper bound, as it is a null sequence and hence arbitrarily small for sufficiently large n (cf. Assumptions 2,4 and 5). The terms $R_{8,1} - R_{8,4}$ are as follows:

Ad $R_{8,1}$:

$$\begin{aligned}R_{8,1} &:= K \max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}\left(n^{-1}\sum_{i=1}^{n}\langle X_{it}^c, \hat{\phi}_{j,t}\rangle\epsilon_{it}^c\right)^2 > c\right) \\ &\leq K \max_{1\leq k\leq K}|G_k|\frac{4\sum_{j=1}^{m}\lambda_{j,k}^{-2}E\left[||\hat{K}_{X\epsilon,t}||_2^2\right]}{c} + K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right) \\ &= O\left(n^{\frac{1+(1+\delta)\mu-2(1-\delta)\nu}{\mu+2\nu}}\right) + o(1) = o(1)\end{aligned}$$

due to Assumptions 2–5.

Ad $R_{8,2}$:

$$\begin{aligned}
R_{8,2} :=& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-2} \left(n^{-1} \sum_{i=1}^{n} \langle X_{it}^c, \hat{\phi}_{j,t} \rangle z_{it}^c \right)^2 (\hat{\beta}_t - \beta_t)^2 > c \right) \\
\leq & K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(4 \sum_{j=1}^{m} \lambda_{j,k}^{-2} \langle \hat{K}_{zX,t}, \hat{\phi}_{j,t} \rangle^2 (\hat{\beta}_t - \beta_t)^2 > c \right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{F}_{4,n,t}^c \right) \\
\leq & K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} \lambda_{j,k}^{-2} \langle K_{zX,k}, \phi_{j,k} \rangle^2 (\hat{\beta}_t - \beta_t)^2 > c \right) + 2K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left((\hat{\beta}_t - \beta_t)^2 > c \right) \\
& + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(||K_{zX,k}||_2^2 \sum_{j=1}^{m} \lambda_{j,k}^{-2} ||\phi_{j,k} - \hat{\phi}_{j,t}||_2^2 > c \right) \\
& + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(||K_{zX,k} - \hat{K}_{zX,t}||_2^2 > c \right) \\
& + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{F}_{4,n,t}^c \right) \\
=& o(1)
\end{aligned}$$

which follows from our above observations.

Ad $R_{8,3}$:

With $a_{j,t} := \langle \alpha_t, \hat{\phi}_{j,t} \rangle = \langle A_k, \hat{\phi}_{j,t} \rangle$, we obtain

$$\begin{aligned}
R_{8,3} &:= K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} (a_{j,t}^* - a_{j,t}) \hat{\phi}_{j,t} \right|\right|_2^2 > c \right) \\
&\leq K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(||A_k||_2^2 \sum_{j=1}^{m} \left|\left| \phi_{j,k} - \hat{\phi}_{j,t} \right|\right|_2^2 > c \right) \\
&= o(1)
\end{aligned}$$

as a consequence of (1.23).

Ad $R_{8,4}$:

$$\begin{aligned}
R_{8,4} &:= K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} a^*_{j,t}(\hat{\phi}_{j,t} - \phi_{j,k}) \right|\right|_2^2 > c \right) \\
&\leq K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(m \sum_{j=1}^{m} {a^*_{j,t}}^2 \left|\left| \hat{\phi}_{j,t} - \phi_{j,k} \right|\right|_2^2 > c \right) \\
&= O\left(mn^{\delta-1}\right) = o(1),
\end{aligned}$$

which follows from the arguments used for $R_{4,2,t}$, because $|\langle K_{zX,k}, \phi_{j,k}\rangle|/\lambda_{j,k}$ and $|a^*_{j,t}|$ are of the same order in j. Combining arguments yields $\mathbb{P}\left(\max_{1\leq t\leq T} ||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) = o(1)$ proving the second claim of the Lemma.

This would already justify classification on the distances $||\hat{\alpha}_t - \hat{\alpha}_s||_2^2$. However, as scaled versions of the estimators are employed the behavior of the scaling, which itself is random, needs to be explored. Contributing to this, now turn to the event $\mathcal{S}_{n,t}$, for which

$$\begin{aligned}
K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{S}^c_{n,t}\right) \leq & K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left({\epsilon^c_{it}}^2 - \sigma^2_{\epsilon,k} + 2\epsilon^c_{it}\tilde{r}_{it} + \tilde{r}^2_{it} \right) \right| > \frac{1}{2}\sigma^2_{\epsilon,k} \right) \\
\leq & K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left({\epsilon^c_{it}}^2 - \sigma^2_{\epsilon,k} + 2\epsilon^c_{it}\tilde{r}_{it} + \tilde{r}^2_{it} \right) \right| > \frac{1}{2} \min_{1\leq k\leq K} \sigma^2_{\epsilon,k} \right) \\
\leq & R_{9,1} + R_{9,2} + R_{9,3}
\end{aligned}$$

where $\tilde{r}_{it} := z^c_{it}(\beta_t - \hat{\beta}_t) + \langle X^c_{it}, \alpha_t - \hat{\alpha}_t\rangle$, $\min_{1\leq k\leq K} \sigma^2_{\epsilon,k}$ a constant, and $R_{9,1} - R_{9,3}$ are as follows.

Ad $R_{9,1}$:

$$\begin{aligned}
R_{9,1} &:= K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left({\epsilon^c_{it}}^2 - \sigma^2_{\epsilon,k} \right) \right| > c \right) \\
&\leq K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left({\epsilon_{it}}^2 - \sigma^2_{\epsilon,k} \right) \right| > c \right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left(n^{-1} \sum_{i=1}^{n} \epsilon_{it} \right)^2 > c \right) \\
&\leq K \max_{1\leq k\leq K} |G_k| \frac{n^{-1} E\left[\left({\epsilon_{it}}^2 - \sigma^2_{\epsilon,k} \right)^2 \right]}{c} + K \max_{1\leq k\leq K} |G_k| \frac{n^{-1} E\left[\epsilon^2_{it}\right]}{c} \\
&= o(1).
\end{aligned}$$

Ad $R_{9,2}$:

$$\begin{aligned} R_{9,2} :=& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} (\epsilon_{it}^c \tilde{r}_{it}) \right| > c \right) \\ &\leq R_{10,1} + R_{10,2} \end{aligned}$$

with $R_{10,1} - R_{10,2}$ as follows:

$$\begin{aligned} R_{10,1} :=& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| (\beta_t - \hat{\beta}_t) \hat{K}_{z\epsilon,t} \right| > c \right) \\ &\leq K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\hat{K}_{z\epsilon,t}^2 > c \right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left((\beta_t - \hat{\beta}_t)^2 > c \right) = o(1) \end{aligned}$$

by (1.21) and the above results. Further

$$\begin{aligned} R_{10,2} :=& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| \langle \hat{K}_{X\epsilon,t}, \alpha_t - \hat{\alpha}_t \rangle \right| > c \right) \\ &\leq K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(||\hat{K}_{X\epsilon,t}||_2^2 > c \right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(||\alpha_t - \hat{\alpha}_t||_2^2 > c \right) = o(1) \end{aligned}$$

by (1.20) and the above results on $\hat{\alpha}_t$.

Ad $R_{9,3}$:

$$\begin{aligned} R_{9,3} :=& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \tilde{r}_{it}^2 \right| > c \right) \\ &\leq \underbrace{K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(|\hat{K}_{z,t}| \cdot (\hat{\beta}_t - \beta_t)^2 > c \right)}_{=:R_{11,1}} \\ &\quad + \underbrace{K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} ||X_{it}^c||_2^2 ||\hat{\alpha}_t - \alpha_t||_2^2 > c \right)}_{=:R_{11,2}} \end{aligned}$$

with $R_{11,1}$ and $R_{11,2}$ to be treated as follows.

$$\begin{aligned}
R_{11,1} :=& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(|K_{z,k}| \cdot (\hat{\beta}_t - \beta_t)^2 > c\right) \\
\leq& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(|\hat{K}_{z,t} - K_{z,k}|(\hat{\beta}_t - \beta_t)^2 > c\right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(|K_{z,k}| \cdot (\hat{\beta}_t - \beta_t)^2 > c\right) \\
\leq& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(|\hat{K}_{z,t} - K_{z,k}| > c\right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left((\hat{\beta}_t - \beta_t)^2 > c\right) \\
&+ K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(|K_{z,k}|(\hat{\beta}_t - \beta_t)^2 > c\right) = o(1)
\end{aligned}$$

by (1.19) and the above results. Further it holds that

$$\begin{aligned}
R_{11,2} :=& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} ||X_{it}^c||_2^2 ||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) \\
\leq& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \left|||X_{it}^c||_2^2 - E\left[||X_{it}^c||_2^2\right]\right| \cdot ||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) \\
&+ K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(E\left[||X_{it}^c||_2^2\right] ||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) \\
\leq& K \max_{1\leq k\leq K} |G_k| \frac{n^{-1} E\left[\left(||X_{it}^c||_2^2 - E\left[||X_{it}^c||_2^2\right]\right)^2\right]}{c} + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) \\
&+ K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(E\left[||X_{it}^c||_2^2\right] ||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) \\
=& O(n^{\delta-1}) + o(1) + o(1) = o(1)
\end{aligned}$$

in light of our above findings. Combining results yields $K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{S}_{n,t}^c\right) = o(1)$.

Now, finally turning to $\hat{\alpha}_t^{(\Delta)}$, for sufficiently large n

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq t\leq T} \left|\left|\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}\right|\right|_2^2 > c\right) &\leq K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left|\left|\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}\right|\right|_2^2 > c\right) \\
&\leq R_{12,1} + R_{12,2}
\end{aligned}$$

with

$$R_{12,1} := K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} (\hat{a}_{j,t} - a_{j,t})^2 \frac{\hat{\lambda}_{j,t}}{\hat{\sigma}_{\epsilon,t}^2} > c\right)$$

$$\text{and} \quad R_{12,2} := K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left|\left|\sum_{j=1}^{m} \left(\frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{\epsilon,t}} \hat{\phi}_{j,t} a_{j,t} - \frac{\lambda_{j,k}^{1/2}}{\sigma_{\epsilon,k}} \phi_{j,k} a_{j,t}^*\right)\right|\right|_2^2 > c\right).$$

$R_{12,1}$ can be decomposed according to

$$R_{12,1} \leq R_{13,1} + R_{13,2} + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{S}_{n,t}^c\right)$$

where

$$R_{13,1} := K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sigma_{\epsilon,k}^{-2} \sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-1} \langle \hat{K}_{zX,t}, \hat{\phi}_{j,t}\rangle^2 (\beta_t - \hat{\beta}_t)^2 > c\right)$$

$$\text{and} \quad R_{13,2} := K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sigma_{\epsilon,k}^{-2} \sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-1} \langle \hat{K}_{X\epsilon,t}, \hat{\phi}_{j,t}\rangle^2 > c\right)$$

because $\hat{\sigma}_{\epsilon,t}^{-2} \leq 2\sigma_{\epsilon,k}^{-2}$ on $\mathcal{S}_{n,t}$. Noting that $\sigma_{\epsilon,k}^{-2}$ is obviously bounded above by a constant, these terms in turn behave as follows:

$$\begin{aligned}
R_{13,1} \leq & K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} \lambda_{j,k}^{-1} \langle K_{zX,k}, \phi_j\rangle^2 (\beta_t - \hat{\beta}_t)^2 > c\right) + 2K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left((\beta_t - \hat{\beta}_t)^2 > c\right) \\
& + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} \lambda_{j,k}^{-1} ||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c\right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} \lambda_{j,k}^{-1} ||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 > c\right) \\
& + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right) \\
& = o(1)
\end{aligned}$$

which follows from our above arguments. Further we conclude

$$\begin{aligned}
R_{13,2} &\leq K \max_{1\leq k\leq K} |G_k| \frac{\sum_{j=1}^{m} \lambda_{j,k}^{-1} E\left[||\hat{K}_{X\epsilon,t}||_2^2\right]}{c} + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right) \\
&= O(n^{\delta-1} m^{1+\mu}) = o(1)
\end{aligned}$$

as consequence of Assumptions 2–5. Now turning to $R_{12,2}$ note that

$$R_{12,2} \leq R_{14,1} + R_{14,2}$$

where

$$R_{14,1} := K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left|\left|\sum_{j=1}^{m} \left(\frac{\hat{\lambda}_{j,t}^{1/2} \sigma_{\epsilon,k}}{\sigma_{\epsilon,k} \hat{\sigma}_{\epsilon,t}} \hat{\phi}_{j,t} - \frac{\lambda_{j,k}^{1/2} \hat{\sigma}_{\epsilon,t}}{\sigma_{\epsilon,k} \hat{\sigma}_{\epsilon,t}} \phi_{j,k}\right) a_{j,t}^*\right|\right|_2^2 > c\right)$$

and

$$R_{14,2} = K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} (a_{j,t}^* - a_{j,t}) \frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{\epsilon,t}} \hat{\phi}_{j,k} \right|\right|_2^2 > c \right).$$

Note for $R_{14,1}$:

$$R_{14,1} \leq R_{15,1} + R_{15,2} + R_{15,3} + K \max_{1\leq k\leq K} |G_k| \mathbb{P}(\mathcal{S}_{n,t}^c)$$

with

$$\begin{aligned}
R_{15,1} &:= K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} (a_{j,t}^*)^2 \hat{\lambda}_{j,t} (\sigma_{\epsilon,k} - \hat{\sigma}_{\epsilon,t})^2 > c \right) \\
R_{15,2} &:= K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(m \sum_{j=1}^{m} (a_{j,t}^*)^2 \hat{\lambda}_{j,t} \hat{\sigma}_{\epsilon,t}^2 ||\phi_{j,k} - \hat{\phi}_{j,t}||_2^2 > c \right) \\
R_{15,3} &:= K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} (a_{j,t}^*)^2 \hat{\sigma}_{\epsilon,t}^2 \left(\hat{\lambda}_{j,t}^{1/2} - \lambda_{j,k}^{1/2} \right)^2 > c \right).
\end{aligned}$$

In order to assess the asymptotic behavior of these terms, we note that by the mean value theorem

- it holds on $\mathcal{S}_{n,t}$ that $|\hat{\sigma}_{\epsilon,t} - \sigma_{\epsilon,k}| \leq \frac{\sqrt{2}}{\sigma_{\epsilon,k}} |\hat{\sigma}_{\epsilon,t}^2 - \sigma_{\epsilon,k}^2|$ and
- it holds on $\mathcal{F}_{4,n,t}$ that $|\hat{\lambda}_{j,t}^{1/2} - \lambda_{j,k}^{1/2}| \leq \left(\frac{2}{\lambda_{j,k}} \right)^{\frac{1}{2}} |\hat{\lambda}_{j,t} - \lambda_{j,k}|$.

Adding these observations to the above allows us to conclude the following for $R_{15,1} - R_{15,3}$:

Ad $R_{15,1}$:

$$\begin{aligned}
R_{15,1} =& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} (a_{j,t}^*)^2 \hat{\lambda}_{j,t} |\sigma_{\epsilon,k} - \hat{\sigma}_{\epsilon,t}|^2 > c \right) \\
\leq& K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\frac{2}{\sigma_{\epsilon,k}^2} \sum_{j=1}^{m} (a_{j,t}^*)^2 \lambda_{j,k} |\hat{\sigma}_{\epsilon,t}^2 - \sigma_{\epsilon,k}^2|^2 > c \right) \\
&+ K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{F}_{4,n,t}^c \right) + K \max_{1\leq k\leq K} |G_k| \mathbb{P}\left(\mathcal{S}_{n,t}^c \right) \\
=& o(1).
\end{aligned}$$

Ad $R_{15,2}$:

$$\begin{aligned}R_{15,2} \leq & K \max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2m\sigma_{\epsilon,k}^2\sum_{j=1}^{m}(a_{j,t}^*)^2\hat{\lambda}_{j,t}||\phi_{j,k}-\hat{\phi}_{j,t}||_2^2>c\right)+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{S}_{n,t}^c\right)\\ =&O(mn^{\delta-1})+o(1)=o(1)\end{aligned}$$

which follows from similar arguments as the ones used for $R_{4,2,t}$.

Ad $R_{15,3}$:

$$\begin{aligned}R_{15,3} \leq & K \max_{1\leq k\leq K}|G_k|\mathbb{P}\left(2\sigma_{\epsilon,k}^2\sum_{j=1}^{m}(a_{j,t}^*)^2\lambda_{j,k}^{-1}|\hat{\lambda}_{j,t}-\lambda_{j,k}|^2>c\right)+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{S}_{n,t}^c\right)\\ &+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right)\\ \leq & K\max_{1\leq k\leq K}|G_k|\frac{2\sigma_{\epsilon,k}^2\sum_{j=1}^{m}(a_{j,t}^*)^2\lambda_{j,k}^{-1}E[\mathcal{D}_t^2]}{c}+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{S}_{n,t}^c\right)+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right)\\ =&o(1).\end{aligned}$$

It remains to show that $R_{14,2}=o(1)$. For this purpose note

$$\begin{aligned}R_{14,2} =& K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\sum_{j=1}^{m}(a_{j,t}^*-a_{j,t})^2\frac{\hat{\lambda}_{j,t}}{\hat{\sigma}_{\epsilon,t}^2}>c\right)\\ \leq & K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(4||\alpha_t||_2^2\sum_{j=1}^{m}||\hat{\phi}_{j,t}-\phi_{j,k}||_2^2\frac{\lambda_{j,t}}{\sigma_{\epsilon,k}^2}>c\right)+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right)\\ &+K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\mathcal{S}_{n,t}^c\right)\\ =&o(1)+o(1)=o(1),\end{aligned}$$

which follows from (1.23) and our above arguments.

Combining arguments implies the last statement in Lemma 1.4.1. ■

Proof of Theorem 1.4.2

Using the results presented in the previous lemma it is possible to argue in analogy to the proof of Theorem 1 in Vogt and Linton (2017) to validate the classification consistency claimed in our

Theorem 1.4.2. For this purpose consider the set $S^{(j)} = \{1, \dots, T\} \setminus \bigcup_{l<j} \hat{G}_l$ at an iteration step $1 \leq j \leq \hat{K} - 1$ of the algorithm described in Section 1.3. For a $t \in S^{(j)}$ denote the set of indexes corresponding to the ordered distances $\hat{\Delta}_{t(1)} \leq \cdots \leq \hat{\Delta}_{t(|S^{(j)}|)}$ as $\{(1), \dots, (|S^{(j)}|)\}$. In analogy, the index set corresponding to the ordered population distances $\Delta_{t[1]} \leq \cdots \leq \Delta_{t[|S^{(j)}|]}$ is denoted as $\{[1], \dots, [|S^{(j)}|]\}$, where Δ_{ts} is as in Assumption 7. Now, define the index $\hat{\kappa}$ according to $\hat{\Delta}_{t(\hat{\kappa})} < \tau_{nT} < \hat{\Delta}_{t(\hat{\kappa}+1)}$. Its population counterpart, κ, obtains as $0 = \Delta_{t[\kappa]} < \tau_{nT} < \Delta_{t[\kappa+1]}$. It holds that

$$\begin{aligned} \mathbb{P}\left(\{(1), \dots, (\hat{\kappa})\} \neq \{([1], \dots, [\kappa]\}\right) &\leq \mathbb{P}\left(\{(1), \dots, (\kappa)\} \neq \{[1], \dots, [\kappa]\}\right) + \mathbb{P}(\hat{\kappa} \neq \kappa) \quad (1.26) \\ &= o(1) + o(1). \end{aligned}$$

In order to prove that the first probability on the right hand side of (1.26) is a null sequence, suppose that $t \in G_k$, with $1 \leq k \leq K$. As indicated, there are $\kappa \geq 1$ indexes in $S^{(j)}$ being elements of G_k. For the corresponding distances it holds that $\Delta_{t[1]} = \cdots = \Delta_{t[\kappa]} = 0$ by definition. The remaining distances are bounded away from zero by $0 < C_\Delta \leq \Delta_{t[\kappa+1]} \leq \cdots \leq \Delta_{t[|S^{(j)}|]}$ due to Assumption 7.

As stated in Lemma 1.4.1, $\max_{1 \leq t \leq T} ||\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}||_2^2 = o_p(1)$ implying that $\max_{1 \leq s \leq T} |\hat{\Delta}_{ts} - \Delta_{ts}| = o_p(1)$, which holds for any reference period t. Combining arguments allows to conclude $\max_{1 \leq s \leq \kappa} \hat{\Delta}_{t(s)} = o_p(1)$ and $\min_{\kappa < s \leq |S^{(j)}|} \hat{\Delta}_{t(s)} \geq C_\Delta + o_p(1)$ as well as $\max_{1 \leq s \leq \kappa} \hat{\Delta}_{t[s]} = o_p(1)$ and $\min_{\kappa < s \leq |S^{(j)}|} \hat{\Delta}_{t[s]} \geq C_\Delta + o_p(1)$. This implies that the first probability on the right hand side of (1.26) tends to zero. Further note that the specification of the threshold in Assumption 7 immediately implies $\mathbb{P}\left(\hat{\Delta}_{t[\kappa]} < \tau_{nT}\right) \to 1$ and $\mathbb{P}\left(\hat{\Delta}_{t[\kappa+1]} > \tau_{nT}\right) \to 1$ as $n \to \infty$ in light of the preceding arguments. As a consequence of this $\mathbb{P}\left(\hat{\Delta}_{t[\kappa]} < \tau_{nT} < \hat{\Delta}_{t[\kappa+1]}\right) \to 1$ as $n \to \infty$, implying that the second probability on the right hand side of (1.26) is a null sequence. ■

Remark 1

For the calculation of the convergence rate of our estimator $\tilde{A}_k$, the classification error is negligible as a consequence of Theorem 1.4.2. To see this note that an analogous argument as in Vogt and Linton (2017) holds in our context: let $s(n, T)$ be an arbitrary deterministic sequence such that

$s(n,T) \to 0$ as $n, T \to \infty$. Now, note that for any constant $C > 0$

$$\begin{aligned}
&\mathbb{P}\left((s(n,T))^{-1}||\tilde{A}_k - A_k||_2^2 > C\right) \\
&\leq \mathbb{P}\left(\left\{(s(n,T))^{-1}||\tilde{A}_k - A_k||_2^2 > C\right\} \cap \left\{\hat{G}_k = G_k\right\}\right) + \mathbb{P}\left(\left\{\hat{G}_k \neq G_k\right\}\right) \\
&= \mathbb{P}\left((s(n,T))^{-1}||\tilde{A}_k^* - A_k||_2^2 > C\right) + o(1),
\end{aligned}$$

where the quantity $\tilde{A}_k^*$ denotes the estimator $\tilde{A}_k$ calculated from $\{(y_{it}, X_{it}, z_{it}) : \ 1 \leq i \leq n, \ t \in G_k\}$, i.e. from correctly classified periods. Note in particular that the time series dependence formulated in Assumption 1 does not affect this argument.

In light of this remark, the proof of Theorem 1.4.3 starts from the ideal *oracle* estimators $\tilde{A}_k^*$ rather than their contaminated counterparts.

Remark 2

For the proof of Theorem 1.4.3, we work with classification-error-free oracle variants of the estimators $\tilde{\phi}_{j,k}, \tilde{\lambda}_{j,k}, \tilde{K}_{X,k}, \tilde{K}_{zX,k}, \tilde{K}_{z,k}$ and $\tilde{\Gamma}_k$. Such estimators, calculated from $\{(z_{it}, X_{it}) : \ 1 \leq i \leq n, \ t \in G_k\}$, are denoted $\tilde{\phi}_{j,k}^*, \tilde{\lambda}_{j,k}^*, \tilde{K}_{X,k}^*, \tilde{K}_{zX,k}^*, \tilde{K}_{z,k}^*$ and $\tilde{\Gamma}_k^*$. In analogy to before, we further denote the Hilbert Schmidt norm of the difference between $\tilde{\Gamma}_k^*$ and the population counterpart Γ_k as $\widetilde{\mathcal{D}}_k^* := ||\tilde{\Gamma}_k^* - \Gamma_k||_{\mathcal{S}}$. Beyond these quantities, the estimator $\tilde{A}_k^*$ makes implicitly use of the operator $\tilde{\Phi}_k^*$ which estimates, in analogy to $\hat{\Phi}_t$, the operator Φ_k as in (1.11). $\tilde{\Phi}_k^*$ is defined according to

$$\tilde{\Phi}_k^*(g) := \sum_{j=1}^{\tilde{m}} \frac{\langle \tilde{K}_{zX,k}^*, \tilde{\phi}_{j,k}^* \rangle}{\tilde{\lambda}_{j,k}^*} \langle \tilde{\phi}_{j,k}^*, g \rangle$$

for any $g \in L^2([0,1])$, where $\tilde{m} \equiv \tilde{m}_k$ for simplicity of notation.
Assessing the asymptotic properties of the classification-error-free estimators, note that due to Assumption 1 for every regime G_k, the random variables $\{X_{it} : \ 1 \leq i \leq n, \ t \in G_k\}$ are L_m^4-approximable. Thus, for suitably large constants, the following inequalities from Hörmann and

Kokoszka (2010) hold:[7]

$$E\left[\left(\widetilde{\mathcal{D}}_k^*\right)^2\right] \leq C(n|G_k|)^{-1} \tag{1.27}$$

$$E\left[\left|\left|\tilde{K}_{X,k}^* - K_{X,k}\right|\right|_2^2\right] \leq C(n|G_k|)^{-1} \tag{1.28}$$

$$E\left[\left|\tilde{\lambda}_{j,k}^* - \lambda_{j,k}\right|^2\right] \leq E\left[\left(\widetilde{\mathcal{D}}_k^*\right)^2\right] \leq C(n|G_k|)^{-1} \tag{1.29}$$

for $1 \leq j \leq \tilde{m}$. Further note that the dependence of the random variables $\{(z_{it}, X_{it}) : \ 1 \leq i \leq n, \ t \in G_k\}$ is sufficiently weak, such that

$$E\left[\left|\left|\tilde{K}_{zX,k}^* - K_{zX,k}\right|\right|_2^2\right] = O((n|G_k|)^{-1})$$

and further

$$E\left[\left|\tilde{K}_{z,k}^* - K_{z,k}\right|^2\right] = O((n|G_k|)^{-1}),$$

which can be shown by straightforward moment calculations. In addition to that, bounds on $\left|\left|\tilde{\phi}_{j,k}^* - \phi_{j,k}\right|\right|_2^2$ can be obtained in analogy to the almost sure bound in (1.12) and the asymptotic bound as in (1.13)–(1.15). We make the latter precise defining the analogues to $\mathcal{F}_{1,n,t}$–$\mathcal{F}_{3,n,t}$ as

1. $\widetilde{\mathcal{F}}_{1,n,T,k} := \left\{C(n|G_k|)^{\frac{2(1+\mu)}{\mu+2\nu}} \left(\widetilde{\mathcal{D}}_k^*\right)^2 \leq 1/8\right\}$
2. $\widetilde{\mathcal{F}}_{2,n,T,k} := \left\{|\tilde{\lambda}_{j,k}^* - \lambda_{l,k}|^{-2} \leq 2|\lambda_{j,k} - \lambda_{l,k}|^{-2} \leq C(n|G_k|)^{\frac{2(1+\mu)}{\mu+2\nu}}, \ 1 \leq j \leq \tilde{m}, j \neq l \in \mathbb{N}\right\}$.
3. $\widetilde{\mathcal{F}}_{3,n,T,k} := \widetilde{\mathcal{F}}_{1,n,T,k} \cap \widetilde{\mathcal{F}}_{2,n,T,k}$

for which we note $\mathbb{P}(\widetilde{\mathcal{F}}_{3,n,T,k}^c) \leq \mathbb{P}(\widetilde{\mathcal{F}}_{1,n,T,k}^c) + \mathbb{P}(\widetilde{\mathcal{F}}_{2,n,T,k}^c) = o(1) + o(1)$ as $(n,T) \to \infty$ from similar arguments as before. Also, as in our arguments for the t-wise estimators it holds on $\widetilde{\mathcal{F}}_{2,n,T,k}$ that

$$||\tilde{\phi}_{j,k}^* - \phi_{j,k}||_2^2 \leq 8\left(1 - 4C(n|G_k|)^{\frac{2(1+\mu)}{\mu+2\nu}} \left(\widetilde{\mathcal{D}}_k^*\right)^2\right)^{-1} \tilde{R}_{j,k}^{(\phi)}, \tag{1.30}$$

where $\tilde{R}_{j,k}^{(\phi)} := \sum_{l:l\neq j} (\lambda_{j,k} - \lambda_{l,k})^{-2} \left[\int_0^1 \int_0^1 (\tilde{K}_{X,k}^*(u,v) - K_{X,k}(u,v))\phi_{j,k}(u)\phi_{l,k}(v)\mathrm{d}u\mathrm{d}v\right]^2$,

[7] Cf. Theorem 3.2 and the consequent discussion in Hörmann and Kokoszka (2010).

from which we conclude, that on $\widetilde{\mathcal{F}}_{3,n,T,k}$, it holds that

$$||\tilde{\phi}^*_{j,k} - \phi_{j,k}||_2^2 \leq 16\tilde{R}^{(\phi)}_{j,k}. \tag{1.31}$$

The results in Hall and Horowitz (2007) also allow to conclude $E\left[\tilde{R}^{(\phi)}_{j,k}\right] = O\left(j^2(n|G_k|)^{-1}\right)$ uniformly in $1 \leq j \leq \tilde{m}$ for weakly dependent random variables $\{X_{it} : \ 1 \leq i \leq n, \ t \in G_k\}$.

As a further important observation we note that

$$\left|\left|\tilde{\Phi}^*_k - \Phi_k\right|\right|^2_{H'} = O_p\left((n|G_k|)^{\frac{1-2\nu}{\mu+2\nu}}\right)$$

given Assumptions 1-6 hold. This can be seen from a regression

$$z_{it} - E[z_{it}] = \langle \zeta, X_{it} - E[X_{it}] \rangle + s_{it} \tag{1.32}$$

in the $k-th$ regime, where $1 \leq t \leq |G_k|$, $1 \leq i \leq n$ and s_{it} as in Assumption 6. Since the functional parameter ζ is formulated as being time invariant, it can be estimated as in Hall and Horowitz (2007) from pooled data $(X_{j(i,t)}, z_{j(i,t)})$, where $1 \leq j(i,t) := (i-1)|G_k| + t \leq n|G_k|$. As noted by Shin (2009), the resulting estimator, say $\hat{\zeta}$, links to the operator $\tilde{\Phi}^*_k$ according to

$$||\hat{\zeta} - \zeta||_2^2 = \left|\left|\tilde{\Phi}^*_k - \Phi_k\right|\right|^2_{H'}. \tag{1.33}$$

The argumentation in Hall and Horowitz (2007) (cf. their Theorem 1 and corresponding proof) transfers mutatis mutandis to a setup with weakly dependent regressors (L^4_m dependence) and weakly dependent errors (m-dependence) as is the case in our auxiliary regression (1.32). This can be shown using the fundamental results formulated in Hörmann and Kokoszka (2010). As $\hat{\zeta}$ is calculated from a sample of size $n|G_k|$, the results in Hall and Horowitz (2007) together with (1.33) thus imply

$$\left|\left|\tilde{\Phi}^*_k - \Phi_k\right|\right|^2_{H'} = O_p\left((n|G_k|)^{\frac{1-2\nu}{\mu+2\nu}}\right)$$

as claimed before.

Proof of Theorem 1.4.3

Note that on $\bigcap_{t\in G_k} Q_{n,t}$ it holds that $\hat{B}_t^{-1} \leq 2{B_k}^{-1}$ for any $t \in G_k$, and so

$$\begin{aligned}
&\mathbb{P}\left(n\left(|G_k|\right)^{-1}\left|\sum_{t\in G_k}(\hat{\beta}_t-\beta_t)^2\right|>c\right)\\
\leq&\mathbb{P}\left(4B_k^{-2}n\left(|G_k|\right)^{-1}\left|\sum_{t\in G_k}\sum_{l=1}^{3}R_{0,l,t}^2\right|>c\right)+\mathbb{P}\left(\bigcup_{t\in G_k}Q_{n,t}^c\right)\\
\leq&\mathbb{P}\left(4B_k^{-2}n\left(|G_k|\right)^{-1}\sum_{t\in G_k}\left(\sum_{l=1}^{2}R_{0,l,t}^2+R_{1,1,t}^2+R_{2,1}^2+\sum_{j=1}^{3}R_{3,j,t}^2\right)>c\right)\\
&+|G_k|\mathbb{P}\left(Q_{n,t}^c\right).
\end{aligned}$$

In the proof of Lemma 1.4.1 it was shown that $\mathbb{P}\left(Q_{n,t}^c\right) = o(|G_k|^{-1})$. Regarding the remaining term, note that due to the exogeneity of the regressors and stationary distributions within the regimes the following holds: for any $c_j > 0$, $j = 1, 2, 3$, there exist constants $C_j = C_j(c_j)$, $j = 1, 2, 3$ such that

$$\begin{aligned}
\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k}R_{0,1,t}^2>c_1\right)&\leq\frac{nE[R_{0,1,t}^2]}{c_1}\leq C_1\\
\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k}R_{0,2,t}^2>c_2\right)&\leq\frac{nE[R_{0,2,t}^2]}{c_2}\leq C_2\\
\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k}R_{1,1,t}^2>c_3\right)&\leq\frac{nCE[||\hat{K}_{zX,t}-K_{zX}||_2^2]}{c_3}\leq C_3.
\end{aligned}$$

Further we observe that $R_{2,1} = o(n^{-1/2})$. Using that $\mathbb{P}(\bigcup_{t\in G_k}\mathcal{F}_{3,n,t}^c) \leq \sum_{t=G_k}\mathbb{P}(\mathcal{F}_{3,n,t}^c) = o(1)$ as shown in the proof of Lemma 1.4.1, we argue that for any constant c_4, there exists a constant

C_4, such that

$$\begin{aligned}
&\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k} R_{3,1,t}^2 > c_4\right)\\
\leq&\frac{n||A_k||E\left[||\hat{K}_{zX,t}-K_{zX,k}||_2^4\right]^{\frac{1}{2}} E\left[\mathcal{D}_t^4\right]^{\frac{1}{2}}(C_\lambda')^{-2}\left(\sum_{j=1}^m j^{1+\mu}\right)^2}{c\cdot c_4}\\
&+\frac{nE\left[||\hat{K}_{zX,t}-K_{zX,k}||_2^2\right]\left(\sum_{j=1}^m |a_{j,t}^*|\right)^2}{c\cdot c_4}\\
\leq& C_4,
\end{aligned}$$

due to the stationarity of (X_{it}, z_{it}), $t \in G_k$ and (1.12). Beyond that we argue for $R_{3,2,t}$ that for any $c_5 > 0$ it follows from similar arguments that there exists a $C_5 > 0$ such that

$$\begin{aligned}
&\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k} R_{3,2,t}^2 > c_5\right)\\
\leq&\frac{n||A_k||\cdot E\left[\mathcal{D}_t^2\right](C_\lambda')^{-2}C_{zX}^2\left(\sum_{j=1}^m j^{1-\nu}\right)^2}{c\cdot c_5}\\
\leq& C_5.
\end{aligned}$$

Finally, note that for any $c_6 > 0$, there exists a $C_6 > 0$ such that

$$\begin{aligned}
&\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k} R_{3,3,t}^2 > c_6\right)\\
\leq&\frac{2n||K_{zX,k}||_2^2||A_k||_2^2(C_\lambda')^{-2}E\left[\mathcal{D}_t^4\right]\left(\sum_{j=1}^m j^{2(1+\mu)}\right)^2}{c_6}\\
&+\frac{2n||K_{zX,k}||_2^2(C_\lambda')^{-2}(C_a)^2E\left[\mathcal{D}_t^2\right]\left(\sum_{j=1}^m j^{1+\mu-\nu}\right)^2}{c_6}\\
\leq& C_6
\end{aligned}$$

thanks to Assumptions 2–5, stationarity and once more the bound in (1.12). Combining arguments allows us to conclude that $|G_k|^{-1}\sum_{t\in G_k}(\hat{\beta}_t-\beta_t)^2 = O_p(n^{-1})$ as $n, T \to \infty$. We use this finding in a moment to obtain the convergence rate for $\tilde{A}_k$. To assess the underlying problem, we use the following notation:

- $X_{it}^{cc,*} := X_{it} - \bar{\bar{X}}_k^*$ with $\bar{\bar{X}}_k^* := \frac{1}{n|G_k|}\sum_{t\in G_k}\sum_{i=1}^n X_{it}$,

- $\epsilon_{it}^{cc,*} := \epsilon_{it} - \bar{\bar{\epsilon}}_k^*$ with $\bar{\bar{\epsilon}}_k^* := \frac{1}{n|G_k|} \sum_{t \in G_k} \sum_{i=1}^{n} \epsilon_{it}$.

The classification-error-free oracle estimator for the regime specific parameter function reads as $\tilde{A}_k^* := \sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}^* \tilde{\phi}_{j,k}^*$. The basis coefficients indexed $1 \leq j \leq \tilde{m}$ obtain as

$$\begin{aligned} \tilde{a}_{j,k}^* &:= (\tilde{\lambda}_{j,k}^*)^{-1} \frac{1}{n|G_k|} \sum_{t \in G_k} \sum_{i=1}^{n} \langle X_{it}^{cc,*}, \tilde{\phi}_{j,k}^* \rangle (y_{it}^c - z_{it}^c \hat{\beta}_t) \\ &= \tilde{a}_{j,k}^{(1)} + \tilde{a}_{j,k}^{(2)}, \end{aligned}$$

where

$$\tilde{a}_{j,k}^{(1)} := (\tilde{\lambda}_{j,k}^*)^{-1} \frac{1}{n|G_k|} \sum_{t \in G_k} \sum_{i=1}^{n} \langle X_{it}^{cc,*}, \tilde{\phi}_{j,k}^* \rangle \left(\langle X_{it}^{cc,*}, A_k \rangle + \epsilon_{it}^{cc,*} \right)$$

and

$$\tilde{a}_{j,k}^{(2)} := (\tilde{\lambda}_{j,k}^*)^{-1} \frac{1}{n|G_k|} \sum_{t \in G_k} \sum_{i=1}^{n} \langle X_{it}^{cc,*}, \tilde{\phi}_{j,k}^* \rangle \left(z_{it}^c (\beta_t - \hat{\beta}_t) + \langle \bar{\bar{X}}_k^* - \bar{X}_t, A_k \rangle + \bar{\bar{\epsilon}}_k^* - \bar{\epsilon}_t \right).$$

The upper bound

$$||\tilde{A}_k^* - A_k||_2^2 = \left|\left| \sum_{j=1}^{\tilde{m}} \left(\tilde{a}_{j,k}^{(1)} + \tilde{a}_{j,k}^{(2)} \right) \tilde{\phi}_{j,k}^* - A_k \right|\right|_2^2 \tag{1.34}$$

$$\leq 2 \left|\left| \sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}^{(1)} \tilde{\phi}_{j,k}^* - A_k \right|\right|_2^2 + 2 \sum_{j=1}^{\tilde{m}} \left(\tilde{a}_{j,k}^{(2)} \right)^2 \tag{1.35}$$

can be obtained using the Cauchy Schwarz inequality. The first term is the estimator from Hall and Horowitz (2007) in the case of $n|G_k|$ pooled observations and an L_m^4 approximable regressor function. Along the lines of our second remark and Assumptions 1-5, it holds that $\left|\left| \sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}^{(1)} \tilde{\phi}_{j,k}^* - A_k \right|\right|_2^2 = O_p\left(n^{\frac{(1+\delta)(1-2\nu)}{\mu+2\nu}} \right)$. The remaining term in (1.35) we split according to

$$\sum_{j=1}^{\tilde{m}} \left(\tilde{a}_{j,k}^{(2)} \right)^2 \leq 3 \cdot (R_{16,1} + R_{16,2} + R_{16,3}).$$

where the terms $R_{16,1}$, $R_{16,2}$ and $R_{16,3}$ are as follows:

Ad $R_{16,1}$:

$$\begin{aligned}
R_{16,1} :=& \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(\frac{1}{|G_k|} \sum_{t \in G_k} \langle n^{-1} \sum_{i=1}^{n} X_{it}^{cc,*}, \tilde{\phi}_{j,k}^* \rangle \left(\langle \bar{\bar{X}}_k^* - \bar{X}_t, A_k \rangle \right) \right)^2 \\
\leq & 4 \sum_{j=1}^{\tilde{m}} (\lambda_{j,k})^{-2} \left(\frac{1}{|G_k|} \sum_{t \in G_k} ||X_t - \bar{\bar{X}}_k^*||_2^2 ||A_k||_2 \right)^2 \\
= & O_p(\tilde{m}^{1+2\mu} n^{-2})
\end{aligned}$$

on an event $\widetilde{\mathcal{F}}_{4,n,T,k} := \{|\tilde{\lambda}_{j,k}^* - \lambda_{j,k}| \leq \frac{1}{2}\lambda_{j,k} : \quad 1 \leq j \leq \tilde{m}\}$. For this event in turn, note that $\mathbb{P}(\widetilde{\mathcal{F}}_{4,n,T,k}) \to 1$ which follows from analogous arguments, which lead to $\mathbb{P}(\mathcal{F}_{4,n,t}) \to 1$ above.

Ad $R_{16,2}$:

$$\begin{aligned}
R_{16,2} :=& \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(\frac{1}{n|G_k|} \sum_{t \in G_k} \sum_{i=1}^{n} \langle X_{it}^{cc,*}, \tilde{\phi}_{j,k}^* \rangle \left(\bar{\bar{\epsilon}}_k^* - \bar{\epsilon}_t \right) \right)^2 \\
= & \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(\frac{1}{|G_k|} \sum_{t \in G_k} \langle \bar{X}_t - \bar{\bar{X}}_k^*, \tilde{\phi}_{j,k}^* \rangle \bar{\epsilon}_t \right)^2 \\
\leq & 4 \sum_{j=1}^{\tilde{m}} (\lambda_{j,k})^{-2} \left(\frac{1}{|G_k|} \sum_{t \in G_k} ||\bar{X}_t - \bar{\bar{X}}_k^*||_2 \bar{\epsilon}_t \right)^2 \\
= & O_p(\tilde{m}^{1+2\mu} n^{-2})
\end{aligned}$$

on $\widetilde{\mathcal{F}}_{4,n,T,k}$.

Ad $R_{16,3}$:

$$\begin{aligned}
R_{16,3} :=& \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(\frac{1}{n|G_k|} \sum_{t \in G_k} \sum_{i=1}^{n} \langle X_{it}^{cc,*}, \tilde{\phi}_{j,k}^* \rangle z_{it}^c \left(\beta_t - \hat{\beta}_t \right) \right)^2 \\
= & \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(\frac{1}{|G_k|} \sum_{t \in G_k} \langle \hat{K}_{zX,t}, \tilde{\phi}_{j,k}^* \rangle \left(\beta_t - \hat{\beta}_t \right) \right)^2 \\
\leq & \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(\frac{1}{|G_k|} \sum_{t \in G_k} \langle \hat{K}_{zX,t}, \tilde{\phi}_{j,k}^* \rangle^2 \right) \left(\frac{1}{|G_k|} \sum_{t \in G_k} \left(\beta_t - \hat{\beta}_t \right)^2 \right),
\end{aligned}$$

of which it is known from before that $\frac{1}{|G_k|}\sum_{t\in G_k}\left(\beta_t-\hat{\beta}_t\right)^2=O_p(n^{-1})$ and

$$\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^*)^{-2}|G_k|^{-1}\sum_{t\in G_k}\langle\hat{K}_{zX,t},\tilde{\phi}_{j,k}^*\rangle^2$$
$$\leq\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^*)^{-2}|G_k|^{-1}\sum_{t\in G_k}3\left(\langle K_{zX,k},\phi_{j,k}\rangle^2+\langle\hat{K}_{zX,t}-K_{zX,k},\tilde{\phi}_{j,k}^*\rangle^2+\langle K_{zX,k},\phi_{j,k}-\tilde{\phi}_{j,k}^*\rangle^2\right).$$

We further conclude that on $\widetilde{\mathcal{F}}_{4,n,T,k}$

$$\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^*)^{-2}\langle K_{zX,k},\phi_{j,k}\rangle^2\leq 4\sum_{j=1}^{\tilde{m}}\lambda_{j,k}^{-2}\langle K_{zX,k},\phi_{j,k}\rangle^2\propto\sum_{j=1}^{\tilde{m}}j^{2\mu-2(\mu+\nu)}=O(1)$$

as well as

$$\begin{aligned}|G_k|^{-1}\sum_{t\in G_k}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^*)^{-2}\langle\hat{K}_{zX,t}-K_{zX,k},\tilde{\phi}_{j,k}^*\rangle^2&\leq 2|G_k|^{-1}\sum_{t\in G_k}||\hat{K}_{zX,t}-K_{zX,k}||_2^2\sum_{j=1}^{\tilde{m}}\lambda_{j,k}^{-2}\\&=O_p\left(n^{-1}n^{\frac{(1+\delta)(1+2\mu)}{\mu+2\nu}}\right)\\&=O_p\left(n^{\frac{(1+\delta)(1+2\mu)-\mu-2\nu}{\mu+2\nu}}\right)=o_p(1).\end{aligned}$$

Further, we use similar arguments as before (see the proof of Theorem 1.4.1) to obtain

$$\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^*)^{-2}\langle K_{zX,k},\tilde{\phi}_{j,k}^*-\phi_{j,k}\rangle^2\leq 4||K_{zX,k}||_2^2\sum_{j=1}^{\tilde{m}}||\tilde{\phi}_{j,k}^*-\phi_{j,k}||_2^2\lambda_{j,k}^{-2}=o_p(1)$$

on $\widetilde{\mathcal{F}}_{3,n,T,k}\cap\widetilde{\mathcal{F}}_{4,n,T,k}$, which implies $\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^*)^{-2}\langle K_{zX,k},\tilde{\phi}_{j,k}^*-\phi_{j,k}\rangle^2=O_p(1)$. Combining our above statements yields $\sum_{j=1}^{\tilde{m}}\left(\tilde{a}_j^{(2)}\right)^2=O_p(n^{-1})$. Further, if $\nu>\frac{1+\mu+\delta}{2\delta}$, or equivalently $\delta>(1+\mu)/(2\nu-1)$, then $(nT)^{\frac{1-2\nu}{\mu+2\nu}}=o(n^{-1})$ and in case $\nu<\frac{1+\mu+\delta}{2\delta}$, $n^{-1}=o\left((nT)^{\frac{1-2\nu}{\mu+2\nu}}\right)$. Together with our Remark 1 on the classification error the result in the theorem follows. ■

Threshold Choice

In order to illustrate the properties of the threshold τ_{nT} as suggested in Section 1.5, suppose for a moment that the truncation error in regime k is negligible (i.e., $\lambda_{j,k}\approx 0, j\geq\underline{m}+1$) and that the eigenvalue-eigenfunction pairs $(\lambda_{j,k},\phi_{j,k})_{j\geq 1}$ as well as the error variance $\sigma_{\epsilon,k}^2$ of

regime k were known. In this case our estimation procedure yields variance adjusted estimators $\hat{\alpha}_t^{(\Delta^*)} := \sum_{j=1}^{\underline{m}} \sigma_{\epsilon,k}^{-1} \lambda_{j,k}^{1/2} \hat{a}_{j,t} \phi_{j,k}$ and $\hat{\alpha}_s^{(\Delta^*)} := \sum_{j=1}^{\underline{m}} \sigma_{\epsilon,k}^{-1} \lambda_{j,k}^{1/2} \hat{a}_{j,s} \phi_{j,k}$ where the appropriately scaled difference of their j-th components $(n/2)^{1/2} \sigma_{\epsilon,k}^{-1} \lambda_{j,k}^{1/2} (\hat{a}_{j,t} - \hat{a}_{j,s})$ is approximately standard normal (for large n and small temporal correlations), such that for all $t, s \in G_k$

$$\frac{n}{2}\Delta_{ts}^* := \frac{n}{2}||\hat{\alpha}_t^{(\Delta^*)} - \hat{\alpha}_s^{(\Delta^*)}||_2^2 = \sum_{j=1}^{\underline{m}} \left(\left(\frac{n}{2}\right)^{1/2} \sigma_{\epsilon,k}^{-1} \lambda_{j,k}^{1/2} (\hat{a}_{j,t} - \hat{a}_{j,s}) \right)^2 =: Q_{ts}^{\underline{m}}$$

$$\Rightarrow \Delta_{ts}^* = \frac{2}{n} Q_{ts}^{\underline{m}}, \quad \text{where (for large } n) \quad Q_{ts}^{\underline{m}} \sim \chi_{\underline{m}}^2 \quad \text{if } t \neq s \text{ and } \; Q_{ts}^{\underline{m}} \approx 0 \text{ if } t = s.$$

For accurate estimates and a small truncation error, we expect that $||\hat{\alpha}_t^{(\Delta)} - \hat{\alpha}_s^{(\Delta)}||_2^2 \approx ||\hat{\alpha}_t^{(\Delta^*)} - \hat{\alpha}_s^{(\Delta^*)}||_2^2$ and hence that $\hat{\Delta}_{ts} \approx \Delta_{ts}^*$. Note that neglecting the truncation error is often justified in practice, where a small number of eigencomponents is typically sufficient to explain virtually the total variance (see, for instance, Aue et al., 2015 who use an essentially equivalent practical approach and successfully approximate an infinite dimensional functional time-series using a finite dimensional VAR-model).

To achieve a consistent classification, it is necessary that the threshold parameter $\tau_{nT} \to 0$ as $n, T \to \infty$ since the distances Δ_{ts}^* are null sequences. However, τ_{nT} converges so fast that τ_{nT} remains slightly larger than the maximum within-regime distance $\max_{s \in G_k} \hat{\Delta}_{ts}$. That is, we need to require that $\mathbb{P}\big(\max_{s \in G_k} \hat{\Delta}_{ts} \leq \tau_{nT} \big) \to 1$ or equivalently that $\mathbb{P}\big(\max_{s \in G_k} \hat{\Delta}_{ts} \geq \tau_{nT} \big) \to 0$ for any $t \in G_k$. For finite samples this means requiring that $\mathbb{P}\big(\max_{s \in G_k} \hat{\Delta}_{ts} \geq \tau_{nT} \big) \leq \varepsilon$ for some small $\varepsilon > 0$. Next we use the approximation $\hat{\Delta}_{ts} \approx \Delta_{ts}^*$. Observe that for a given $t \in G_k$,

$$\mathbb{P}\left(\max_{s \in G_k} \Delta_{ts}^* \geq \tau_{nT} \right) = \mathbb{P}\left(\bigcup_{s \in G_k} \{ \Delta_{ts}^* \geq \tau_{nT} \} \right) \leq |G_k| \mathbb{P}\left(Q_{ts}^{\underline{m}} \geq \frac{n}{2} \tau_{nT} \right),$$

where the latter inequality follows from Boole's inequality. From this upper bound we can learn about τ_{nT} according to

$$|G_k| \mathbb{P}\left(Q_{ts}^{\underline{m}} \geq \frac{n}{2} \tau_{nT} \right) = \varepsilon \quad \Leftrightarrow \quad \tau_{nT} = \frac{2}{n} F_{\underline{m}}^{-1}\left(1 - \frac{\varepsilon}{|G_k|} \right),$$

where $F_{\underline{m}}^{-1}$ denotes the quantile function of the $\chi_{\underline{m}}^2$-distribution. As we consider a context where $|G_k|$ is large ($|G_k| \propto T$ in Assumption A3), we expect the value of $\varepsilon / |G_k|$ to be very close to zero. This motivates setting $\tau_{nT} = (2/n) F_{\underline{m}}^{-1}(p_\tau)$, for some p_τ very close to one as mentioned in Section 1.5. Note that according to Theorem A in Inglot (2010) and our assumptions in Section

1.4

$$\tau_{nT} = \frac{2}{n} F_{\underline{m}}^{-1}\left(1 - \frac{\varepsilon}{|G_k|}\right) \leq \frac{2\underline{m}}{n} + \frac{4}{n}\left(\log\left(\frac{|G_k|}{\varepsilon}\right) + \sqrt{\underline{m}\log\left(\frac{|G_k|}{\varepsilon}\right)}\right) \to 0$$

as $n, T \to \infty$, which points at the large sample validity of the proposed threshold.

1.B Additional Simulation Results

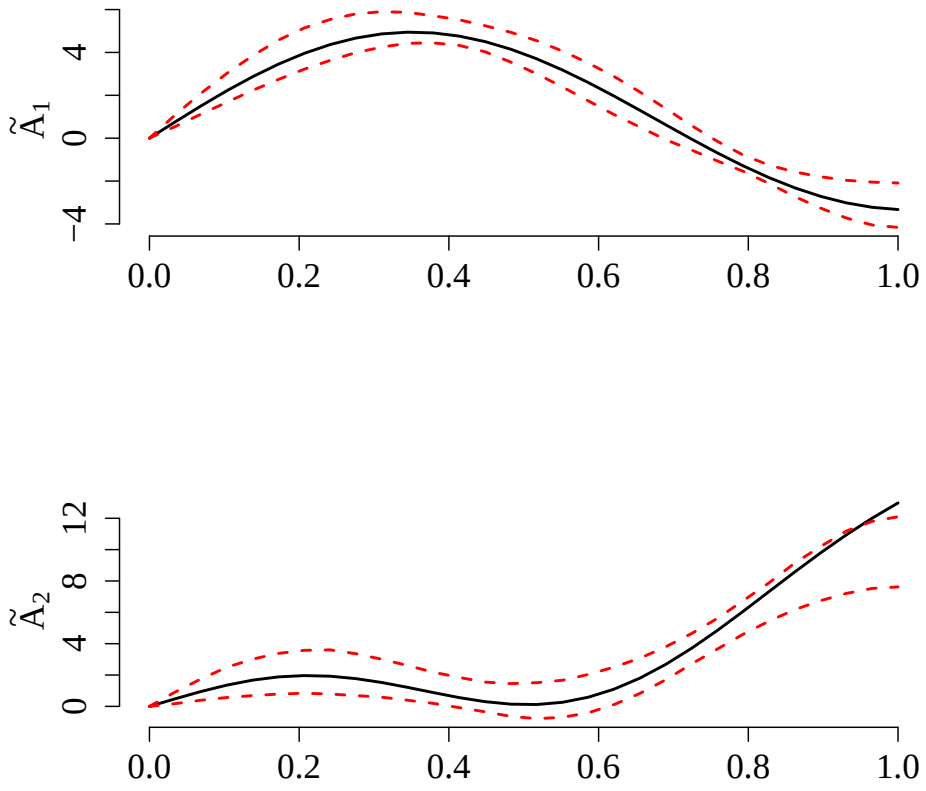

Figure 1.B.1: **Scenario 1: Estimated Regime Parameter Functions for Sample Size** $(n, T) = (50, 50)$**.** *Figure plots true parameter functions (solid black lines) versus pointwise* 5% *and* 95% *quantiles of parameter function estimates (dashed red lines).*

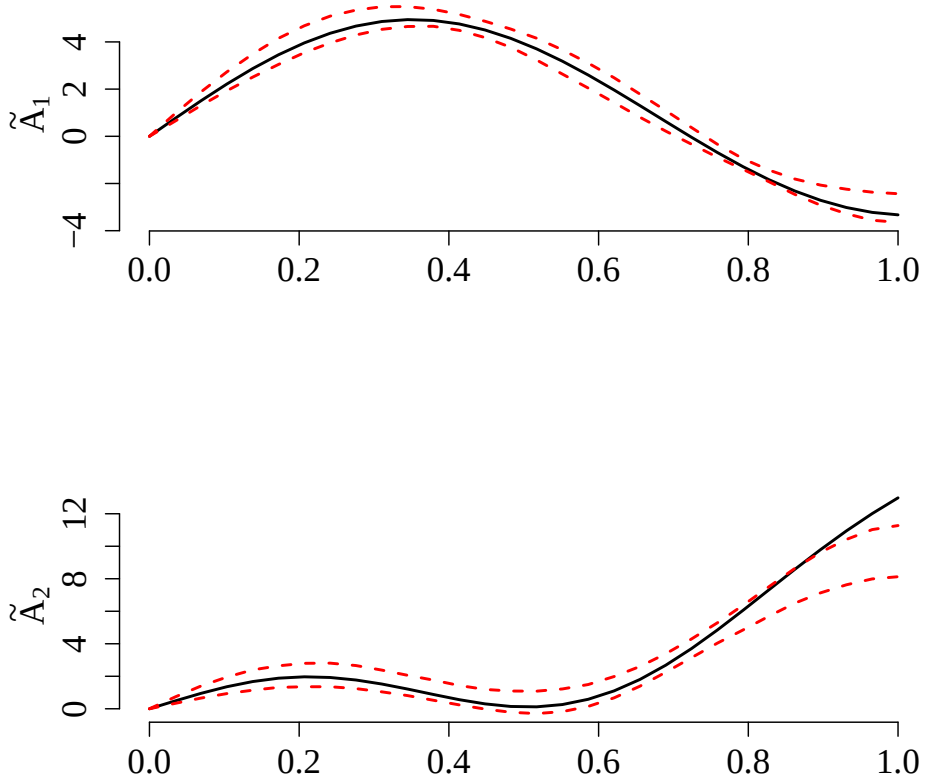

Figure 1.B.2: **Scenario 1: Estimated Regime Parameter Functions for Sample Size** $(n, T) = (100, 50)$**.** *Figure plots true parameter functions (solid black lines) versus pointwise* 5% *and* 95% *quantiles of parameter function estimates (dashed red lines).*

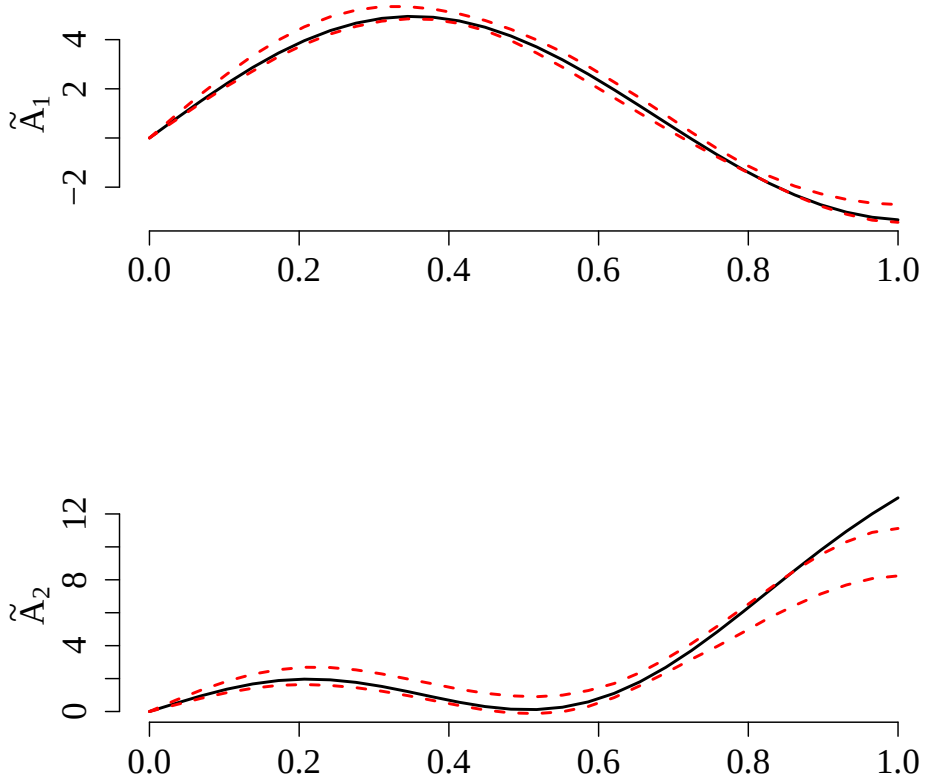

Figure 1.B.3: **Scenario 1: Estimated Regime Parameter Functions for Sample Size** $(n, T) = (150, 80)$**.** *Figure plots true parameter functions (solid black lines) versus pointwise* 5% *and* 95% *quantiles of parameter function estimates (dashed red lines).*

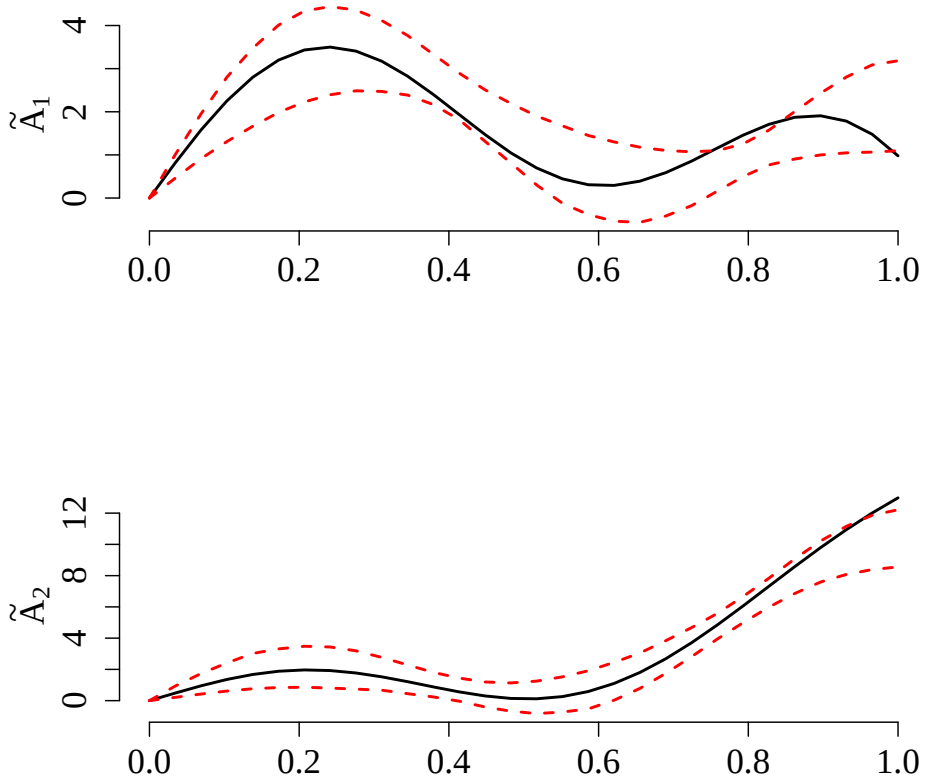

Figure 1.B.4: **Scenario 2: Estimated Regime Parameter Functions for Sample Size** $(n, T) = (50, 50)$**.** *Figure plots true parameter functions (solid black lines) versus pointwise* 5% *and* 95% *quantiles of parameter function estimates (dashed red lines).*

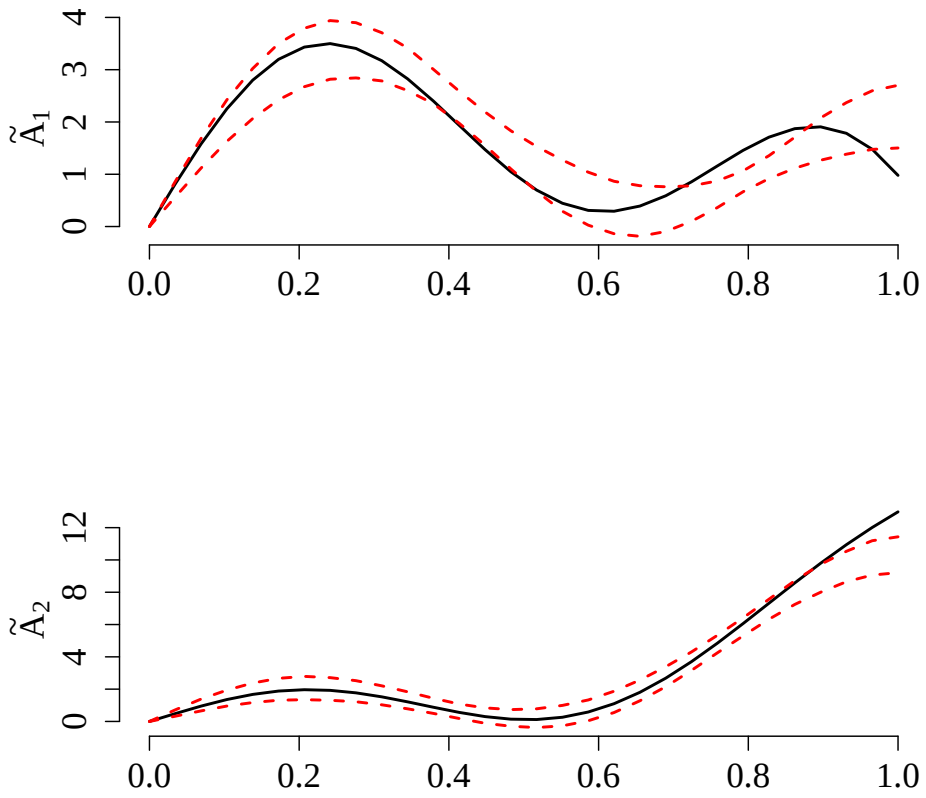

Figure 1.B.5: **Scenario 2: Estimated Regime Parameter Functions for Sample Size** $(n, T) = (100, 50)$**.** *Figure plots true parameter functions (solid black lines) versus pointwise* 5% *and* 95% *quantiles of parameter function estimates (dashed red lines).*

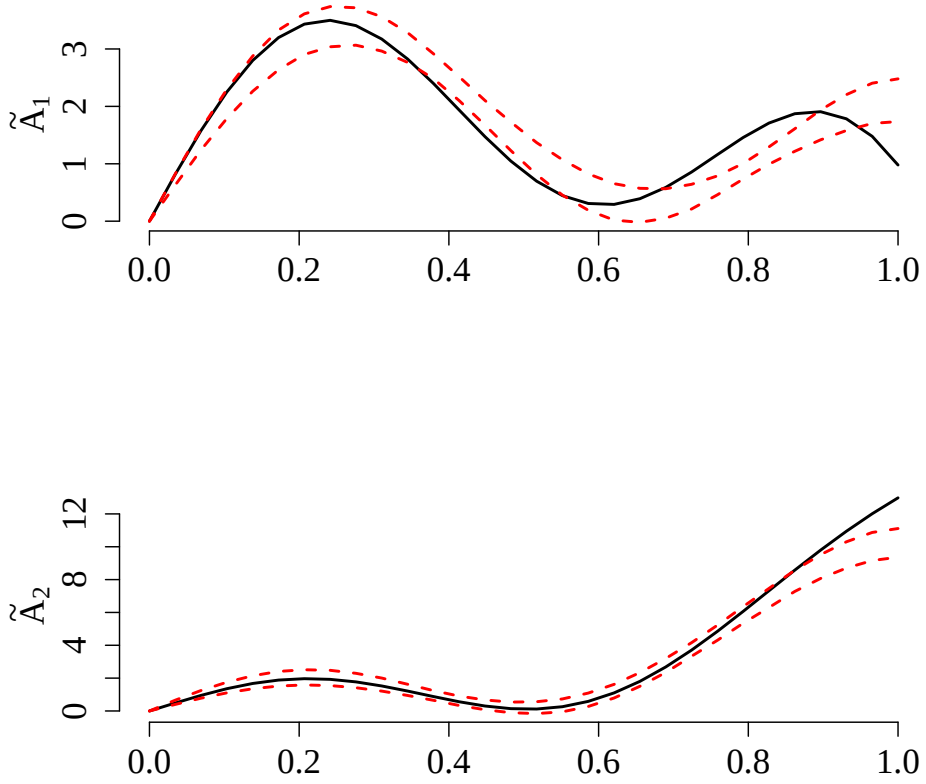

Figure 1.B.6: **Scenario 2: Estimated Regime Parameter Functions for Sample Size** $(n, T) = (150, 80)$**.** *Figure plots true parameter functions (solid black lines) versus pointwise* 5% *and* 95% *quantiles of parameter function estimates (dashed red lines).*

Chapter 2

Online Prediction of Intraday Stock Volatility

2.1 Introduction

Empirical second moments of stock price returns typically vary over time. This is the case for daily as well as intraday measures of return volatility. The complexity of such time variation is challenging, e.g. for algorithmic trading, where many types of algorithms require online predictions of intraday volatility measures. While conventional GARCH models are often adequate for predicting discrete daily volatility measures, they have limited use in an intraday context. This is often dedicated to pronounced intraday patterns of volatility (see for example Engle and Sokalska, 2012). Thus, for intraday algorithmic trading, approaches which explicitly take into account such patterns are a promising direction for research.

Motivated by these observations, I suggest a novel technique designed for the online prediction of intraday spot volatility during the trading day. Spot volatility is considered to be a smooth random curve, governing intraday variation in the second moment of stock price returns in continuous time. My methodology relies on separating the evolution of the level of the volatility curve from the evolution of its intraday shape. I assume the latter to be constituted by a time invariant pattern and irregularly recurring deviations from it. Online prediction of the volatility curve is translated into separately online forecasting the level and the shape of the spot volatility curve.

The suggested technique relates to two strands of the literature on intraday volatility modeling. First, it is similar to econometric works on component models for discrete time intraday volatility dynamics. Second, my approach is related to continuous time models for intraday volatility,

which are typically concerned with modeling functional data. In the literature on discrete time modeling, a closely related work is the one of Engle and Sokalska (2012). The authors introduce a component model governing the volatility of 10-minute returns during the trading day. The employed components are capable to separate daily, diurnal and intraday effects. Along with the returns, the intraday component is finally modeled as a conventional GARCH process. Diao and Tong (2015) use a very similar approach. Extending the model of Engle and Sokalska (2012) the authors suggest an explicit modeling of return dynamics along with a comparable type of component model for the second moment of 5-minute returns of the Chinese CSI 300 index. The intraday component of volatility is modeled in terms of an EGARCH-process. In summary, these two works explicitly model an intraday pattern of volatility as well as systematic deviations from it. Further elements supplementing this approach can be found in Andersen et al. (2012), Dette et al. (2016) and Christensen et al. (2017). In particular the work of Christensen et al. (2017) offers an assessment of the relative importance of diurnal and day-specific effects in intraday volatility. In an empirical study for the 30 constituents of the Dow Jones Industrial Average the authors confirm the important role of the diurnal pattern for intraday volatility, while substantial variation over time is left unexplained. Their findings allude to the potential of explicitly modeling the deviations from the time invariant pattern in intraday volatility. While the component models of Engle and Sokalska (2012) and Diao and Tong (2015) assess such problem in discrete time, functional data models are capable to approach it in continuous intraday time. For example recent work such as Müller et al. (2011) or Dahlhaus and Neddermeyer (2014) rely on measures of the functional *spot volatility*. The latter is the volatility term governing price variation in continuous time. Gabrys et al. (2013) examine the behavior of its structure over time and indicate potential instabilities.

Contributing to the above literature I suggest a novel, functional variant of the discrete time component model of Engle and Sokalska (2012). It explicitly disentangles the shape and the level of spot volatility curves. This is particularly beneficial to the ultimate aim of predicting the continuation of spot volatility once data was observed over a first part of the day. The key innovations of this chapter are the following:

1. Besides a vertical shift governing the level of spot volatility, the suggested model incorporates a global and a local shape component. The global shape component is invariant over trading days and describes a rough diurnal pattern. Deviations of the shape of the spot volatility curve from this global pattern are modeled as a local component. The latter potentially varies over trading days and is assumed to be irregularly recurring. This means that deviations in the shape of intraday volatility from its global pattern occur repeatedly over an

unknown set of trading days. This notion of the intraday volatility shape offers a simple and flexible way to deal with a rich set of systematic deviations from a constant diurnal pattern.

2. Online predictions of the spot volatility curve are constructed by combining online predictions of its level and its shape. As a central novelty of this chapter, the shape forecast is generated using a functional distance metric learning approach. This technique allows to quantify similarity of past intraday volatility curves with respect to their local shape components. Given this data-driven distance notion for volatility shapes, a weighted average of past available curves allows to approximate local (and global) shape features.

In an empirical study I confront this novel prediction technique with stock-level data from the German DAX 30 index. I test the performance of my predictions against three practically motivated benchmarks for each of the considered stocks. The results indicate superior performance of my predictions compared to the competitors. In particular, the relative performance improves as the amount of observed information during a day increases.

The remainder of this chapter is structured as follows. In Section 2.2, I begin with a brief review of the asset pricing framework, which defines spot volatility. After illustrating the practical use of such spot volatility curves, I formally define the prediction problem. Motivated by this problem, I introduce my econometric model and the related prediction scheme in Section 2.3. This is followed by the presentation of a benchmarking strategy to evaluate the accuracy of the forecasts in Section 2.4. I proceed with the empirical study in Section 2.5, where I discuss the construction of spot volatility curves from discrete return data as well as the final prediction results. Section 2.6 concludes.

2.2 Asset Pricing and Spot Volatility

The focus of this work is on the empirical exploration and prediction of the time variation in the second moment of intraday stock price returns. In order to prepare for an econometric assessment of this problem, I rely on a theoretical standard asset pricing model, which reviewed in this section. I consider a stock, which is traded at days indexed $t = 1, 2, \ldots$, where trading at each day takes place in continuous time, say, w.l.o.g., the unit interval $[0, 1]$. Its log-price at a day t at some intraday time $u \in [0, 1]$ is denoted $P_t(u)$. Along the lines of Müller et al. (2011) I assume that the log-price dynamics can be formulated according to the stochastic differential equation

$$\mathrm{d}P_t(u) = \mu_t(u)\mathrm{d}u + \Sigma_t(u)\mathrm{d}B_t(u), \quad u \in [0, 1], \tag{2.1}$$

where μ_t denotes a smooth random drift term and B_t a standard Brownian motion. The object Σ_t is the smooth spot volatility, which will be of main interest in this work. More formally Σ_t is assumed to be a random variable, taking values in $L^2([0,1])$, the space of square integrable functions over $[0,1]$. In addition to the assumption that Σ_t is independent of B_s for all $s = 1, 2, 3 \ldots$, I also require the smoothness assumptions M1-M4 in Müller et al. (2011) for the random variables Σ_t and μ_t to hold.

As a central feature, model (2.1) allows to characterize the conditional distribution of the log-return $P_t(v) - P_t(v-\Delta)$ between between times $(v-\Delta) \in (0, v)$ and $v \in [0,1]$, at some day t (see e.g. Barndorff-Nielsen and Shephard (2002)). It follows from the properties of the standard Brownian motion in (2.1) that

$$P_t(v) - P_t(v-\Delta) | \mu_{t,v,\Delta}, IV_{t,v,\Delta} \sim \mathcal{N}\left(\mu_{t,v,\Delta}, IV_{t,v,\Delta}\right), \tag{2.2}$$

where

$$\mu_{t,v,\Delta} := \int_{v-\Delta}^{v} \mu_t(u)\mathrm{d}u$$
$$\text{and} \quad IV_{t,v,\Delta} := \int_{v-\Delta}^{v} \Sigma_t^2(u)\mathrm{d}u.$$

The distribution stated in (2.2) implies in particular, that once the aggregated effect of the drift term, $\mu_{t,v,\Delta}$, is removed, the return distribution is exhaustively characterized by the so-called integrated volatility $IV_{t,v,\Delta}$. The latter in turn can be determined from the spot volatility $\Sigma_t(u)$, $u \in [v-\Delta, v]$. In terms of its use for intraday trading, an accurate approximation of (2.2) is of fundamental interest e.g. for risk management, order allocation or for the design of exit strategies. Illustrating the latter example, consider an investor holding a position in the stock of interest at *current time* $\tau \in [0,1]$. Suppose this investor wishes to obtain a decision about whether to hold or to sell his position within the next Δ units of time. One approach would be to implement an algorithm liquidating the existing position, if a certain p-quantile of the approximated distribution of $P_t(\tau+\Delta) - P_t(\tau)$ is below a pre-specified stop-loss threshold. Given the effect of the drift μ_t is negligible (which is a priori a reasonable assumption), the required p-quantile can be obtained from a volatility forecast characterizing the approximate return distribution as in (2.2). This example illustrates the use of spot volatility predictions for forecasting the future return distribution via the integrated volatility. In the above example the time horizon over which a decision has to be made might well vary over different investor needs such as risk management requirements. In light of this, a maximum degree of flexibility is achieved having an accurate prediction of the

whole remaining spot volatility curve $\Sigma_t(u)$, $u \in [\tau, 1]$, at that day available, rather than just some particular values of the integrated volatility IV.

The above example is just one out of several situations, in which it is desirable to obtain an accurate prediction of a missing piece of the spot volatility curve. Equivalently to obtaining this prediction in terms of Σ_t, I will develop my arguments for its logarithm

$$\sigma_t(u) := \log\left(\Sigma_t\right)(u)$$

in what follows. Working with log-volatility is common in the literature (see for example Engle and Sokalska, 2012 and Müller et al., 2011) and particularly appealing in my context, since multiplicative components in Σ_t translate into additive components in σ_t as will be discussed in the next section.

Given the bijective transformation[1] σ_t of the spot volatility Σ_t and neglecting the issue of curve construction from discrete data, the final prediction problem can be formulated as follows. Being at present time $\tau \in [\tau_{\min}, 1)$, $0 < \tau_{\min} < 1$ at day $T+1$, suppose that (i) a sample of T past log-spot volatility curves $\{\sigma_t : 1 \leq t \leq T\}$ and (ii) an estimate of the first part $\sigma^L_{T+1,\tau}$ of the curve σ_{T+1} on $[0, \tau]$ are available to the researcher. The objective is to predict from this information $\sigma^R_{T+1,\tau}$, the (logarithmic) spot volatility on the second part of the day, i.e. the curve σ_{T+1} on $[\tau, 1]$. The fact that present time τ is allowed to proceed through $(\tau_{\min}, 1)$, turns this situation into an online prediction problem. The lower bound $\tau_{\min}$ on τ can be viewed as governing a minimum amount of information, which is required to be available at day $T+1$. In what follows I will always suppose τ to be in $(\tau_{\min}, 1)$ without further reference.

Designing a suitable technique in order to approach the above forecasting problem is the central objective of my work. Such a technique consists of two building blocks. First, the spot volatility curves need to be recovered from discrete data. Second, the information in these past curves need to be processed in order to obtain a forecast for $\sigma^R_{T+1,\tau}$. While the issue of fitting spot volatility from discrete data is postponed to Section 2.5, I suggest a model and the corresponding forecasting scheme in the next section.

[1] In what follows I will refer to σ_t simply as the *spot volatility* for stylistic reasons. I only distinguish logarithmic spot volatility and spot volatility where both objects appear at the same time.

2.3 Econometric Modeling

The prediction problem defined in the previous section is known as the *curve continuation* problem in the functional data literature. A number of recent papers is concerned with this specific problem (cf. Goldberg et al., 2014, Kraus, 2015 and Recaredo and José, 2015), while a wider literature deals with the reconstruction of fragmented or partially observed functional data (cf. Delaigle and Hall, 2013, Kneip and Liebl, 2017 and Liebl and Rameseder, 2017). Instead of employing these mostly theoretically oriented procedures, my approach is motivated from the econometric literature on intraday volatility modeling. As argued in Section 2.1, recent work points at the potential of identifying a diurnal pattern of intraday volatility as well as systematic deviations from it. I suggest to explore such contributions to the intraday spot volatility curve, in order to tackle the above prediction problem. To do so, I will introduce a component model for σ_t in Part 2.3.1. The model's implications for $\sigma^L_{t,\tau}$ and $\sigma^R_{t,\tau}$ are subsequently employed for the design of a prediction scheme, which is discussed in Part 2.3.2.

Throughout this section, I assume, that a set of spot volatility curves, $\sigma^L_{t,\tau}, \sigma^R_{t,\tau}, \sigma_t$, $1 \leq t \leq T$ as well as $\sigma^L_{T+1,\tau}$ is available to the econometrician. For the sample up to day T

$$\sigma_t(u) = \begin{cases} \sigma^L_{t,\tau}(u) \text{ if } u \in [0,\tau] \\ \sigma^R_{t,\tau}(u) \text{ if } u \in [\tau,1] \end{cases}$$

defines the two sub-trajectories $\sigma^L_{t,\tau}$ and $\sigma^R_{t,\tau}$, $1 \leq t \leq T$ in analogy to day $T+1$.

2.3.1 A Component Model for Spot Volatility

In the discrete time context, Engle and Sokalska (2012) and Diao and Tong (2015) both use a multiplicative component model for the intraday volatility, which leads to an additive component model for the log-volatility. Their models can easily be transferred to the functional context, which I consider. More concretely, I postulate the logarithmic spot volatility curve σ_t at day t to evolve according to

$$\sigma_t(u) = \nu_t + s(u) + q_t(u) + e_t(u). \tag{2.3}$$

The first three summands on the right hand side of (2.3) are the model's systematic components while e_t denotes an error term. The role of these four terms can be summarized as follows:

(i) The term ν_t denotes a random, real-valued daily **level component**. It induces vertical shifts

of the spot volatility curve and is assumed to vary across trading days according to an ARMA(p,q)-process. ν_t can be understood as the slowly varying component in σ_t as the process varies only across but not within trading days.

(ii) The term $s(u) \in L^2([0,1])$ is the model's **global shape component**, which determines the rough form of the curve apart from vertical shifts. It is deterministic and assumed to be constant across different trading days.

(iii) The third component, $q_t \in L^2([0,1])$ is the **local shape component**. This function-valued random variable reflects smooth day-specific deviations in the shape of σ_t from the global shape component s.[2] As with s, q_t does not induce vertical shifts to the spot volatility curve, as made precise below.

(iv) The term e_t is a smooth random error function with values in $L^2([0,1])$. I assume this error to have the following properties:

(A.1) The functions $\{e_t : 1 \leq t \leq T\}$ are independent across days t and have a zero mean in the sense that, $E[e_t](u) = 0$ and $E[\mathrm{d}e_t(u)/\mathrm{d}u] = 0$ in the L^2-norm for all $1 \leq t \leq T$.

(A.2) e_t is independent of ν_s and q_s for any $1 \leq t, s \leq T$.

(A.3) For any $\tau \in (\tau_{\min}, 1)$ there exists a constant $c_{e,\tau} > 0$, such that

$$\int_0^\tau \left| \frac{\mathrm{d}e_t(u)}{\mathrm{d}u} - \frac{\mathrm{d}e_s(u)}{\mathrm{d}u} \right] \mathrm{d}u \leq c_{e,\tau} \text{ almost surely} \tag{2.4}$$

for any $1 \leq t \neq s \leq T$.

Assumptions A.1 and A.2 are standard assumptions as they appear e.g. in functional regression analysis. Intuitively Assumption A.3 can be thought of as a limitation of the error contribution to shape differences between different volatility curves.

The interplay of the three systematic components in (2.3) allows for a rich time variation of the stochastic spot volatility curve. This variation refers to two time scales. While the inter-day variation of σ_t is driven by ν_t, q_t and e_t, the intraday dynamics are determined by s, q_t and e_t. Apart from the error function e_t, the local shape component q_t is the only component which varies over both time scales. Such property makes q_t a flexible device to catch deviations of the spot volatility from a global shape.

[2]Note that the two shape components inherit the smoothness of σ_t which follows from the assumptions for Σ_t indicated in Section 2.2.

The way I model q_t below is inspired by the desire to capture shapes, which are specific to recurring *environments*. These environments are thought of as a convolution of regularly recurring phenomena such as weekday effects and irregularly recurring generic market situations, for example a bullish stock market paired with a bearish market for corporate bonds. In this particular example investors might have ceteris paribus an incentive to build up new or to strengthen existing positions in stocks by transferring money from the bond to the stock market. The resulting increased demand for stocks will most likely impact not only the first but also the second moment of their prices and in particular their intraday variation. Beyond this simplified example there are a number of situations, which occur repeatedly also in finite samples, given the considered time span is sufficiently large. More than that, I assume that there are in general just finitely many generic situations: Put differently, I assume that all shape-relevant generic situations can be classified into a finite number of categories.

In order to embed this intuition in model (2.3), I assume q_t to be a random variable with a support consisting of $J < \infty$ distinct elements $\{Q_j : 1 \leq j \leq J\}$, where J is a fixed quantity. Re-framing this assumption in terms of trading days, I suppose that there are nonempty sets $\mathcal{P}_1, \ldots, \mathcal{P}_J \subset \{1, \ldots, T, T+1\}$, which collect similar local shape components. In particular I assume the sets $\mathcal{P}_j$, $1 \leq j \leq J$, which I simply refer to as *environments*, to form a partition of the set of trading days $\{1, \ldots, T, T+1\}$. The local shape components differ across these environments and are constant within the environments[3] in the sense that

$$q_t = Q_j \text{ almost surely, if } t \in \mathcal{P}_j \tag{2.5}$$

and

$$\int_0^1 |Q_j(u) - Q_l(u)| \, \mathrm{d}u > C_Q > 0 \text{ for } j \neq l. \tag{2.6}$$

The constant C_Q is to be understood as a lower bound on the heterogeneity across different environments. Note that the L^1 distance in (2.6) is calculated over the entire intraday trading time $[0, 1]$. Beyond this I also suppose differences between the local shape components to satisfy the following local assumption:

(A.4) Consider any t, s such that $t \in \mathcal{P}_j$, $s \in \mathcal{P}_l$, $1 \leq j \neq l \leq J$. Suppose that for any

[3]Equality in Equation (2.5) is meant in terms of the L^2-norm.

$\tau \in (\tau_{\min}, 1)$, there exist constants $C_{1,\tau}, C_{2,\tau} > 0$ such that

$$\int_0^\tau \left| \frac{\mathrm{d}q_t(u)}{\mathrm{d}u} - \frac{\mathrm{d}q_s(u)}{\mathrm{d}u} \right] \mathrm{d}u > 2c_{e,\tau} + C_{1,\tau} \text{ almost surely} \tag{2.7}$$

and

$$\int_\tau^1 |q_t(u) - q_s(u)| \, \mathrm{d}u > C_{2,\tau} \text{ almost surely}, \tag{2.8}$$

where $c_{e,\tau}$ is the constant from Assumption A.3.

Intuitively, the two statements in Assumption A.4 guarantee the local shape component to be sufficiently different across environments on both parts of the day. In particular derivatives of q_t on the first half of the day are well distinguishable relative to the error term: The lower bound in (2.7) can be thought of as a restriction governing the signal-to-noise relation. The larger shape differences between errors are allowed to be, the more pronounced differences between distinct local shape components need to be.

Beyond the recurring nature of q_t, I alluded to the level-invariance of both, the local and the global shape components. They are assumed to exclusively affect the shape of σ_t, but not its level in terms of vertical shifts. More formally, I suppose that for any $1 \leq t \leq T + 1$

$$\int_0^1 s(u)\mathrm{d}u = 0 \quad \text{and} \quad \int_0^1 q_t(u)\mathrm{d}u = 0 \quad \text{almost surely.} \tag{2.9}$$

The condition in (2.9) is not very restrictive, as a necessary scaling is implicitly incorporated in the shift term ν_t.

The distribution of the curves σ_t is per se not strictly stationary given the above assumptions. In particular, the distribution of the local shape component might depend on time due to regularly recurring phenomena such as seasonal effects and further due to potentially non-stationary effects of market factors. Also, a time-variant covariance structure of the error e_t might contribute to deviations from stationarity. The lack of stationarity might be problematic for conventional estimation techniques. However, the model has two implications, which are particularly useful for prediction but do not rely on time-invariant distributions:

1. The components s and q_t are level-invariant in the sense of (2.9). This implies together with

the independence of e_t and ν_t that

$$E\left[\int_0^1 \sigma_t(u)\mathrm{d}u \middle| \nu_t\right] = \nu_t + E\left[\int_0^1 e_t(u)\mathrm{d}u \middle| \nu_t\right] = \nu_t. \tag{2.10}$$

In light of (2.10), the level component ν_t can be interpreted as the error-free logarithmic equivalent to the daily integrated volatility.

2. The component-induced shape of the spot volatility curve can be inferred - up to an error - from derivatives and thus irrespective of ν_t. To see this, I note that the derivative curve corresponding to σ_t reads as

$$\frac{\mathrm{d}\sigma_t(u)}{\mathrm{d}u} = \frac{\mathrm{d}s(u)}{\mathrm{d}u} + \frac{\mathrm{d}q_t(u)}{\mathrm{d}u} + \frac{\mathrm{d}e_t(u)}{\mathrm{d}u}, \quad u \in [0,1],\ 1 \leq t \leq T$$

and consequently

$$\left|\frac{\mathrm{d}\sigma_t(u)}{\mathrm{d}u} - \frac{\mathrm{d}\sigma_s(u)}{\mathrm{d}u}\right| = \left|\frac{\mathrm{d}q_t(u)}{\mathrm{d}u} - \frac{\mathrm{d}q_s(u)}{\mathrm{d}u} + \frac{\mathrm{d}e_s(u)}{\mathrm{d}u} - \frac{\mathrm{d}e_t(u)}{\mathrm{d}u}\right|, \quad u \in [0,1],\ 1 \leq t,s \leq T.$$

Hence, Assumptions A.2–A.4 imply that for any $\tau \in (\tau_{\min}, 1)$

$$\int_0^\tau \left|\frac{\mathrm{d}\sigma_t(u)}{\mathrm{d}u} - \frac{\mathrm{d}\sigma_s(u)}{\mathrm{d}u}\right| \begin{cases} \leq c_{e,\tau} & \text{if } t,s \in \mathcal{P}_j,\ 1 \leq j \leq J \\ > c_{e,\tau} + C_{1,\tau} & \text{if } t \in \mathcal{P}_j,\ s \in \mathcal{P}_l,\ 1 \leq j \neq l \leq J \end{cases} \tag{2.11}$$

almost surely. Both statements in (2.11) follow from the triangle inequality.

The preceding observations can be summarized as follows. First, the modeling of level and shape effects can be separated by integration and differentiation. Second, differences in derivative curves are potentially capable to uncover information about the underlying environments. Next, I will argue that the above two aspects can be used for the ultimate problem of forecasting the curve $\sigma^R_{T+1,\tau}$.

2.3.2 Prediction

The curves $\sigma^L_{t,\tau}$ and $\sigma^R_{t,\tau}$ inherit the structure of σ_t as given in (2.3). Formally a decomposition of the model's components into pre-τ and post-τ parts obtains according to

$$\sigma_t(u) = \begin{cases} \sigma^L_{t,\tau}(u) = \nu_t + s^L_\tau(u) + q^L_{t,\tau}(u) + e^L_{t,\tau}(u) \text{ if } u \in [0,\tau] \\ \sigma^R_{t,\tau}(u) = \nu_t + s^R_\tau(u) + q^R_{t,\tau}(u) + e^R_{t,\tau}(u) \text{ if } u \in [\tau,1], \end{cases} \tag{2.12}$$

where superscripts L and R indicate, in analogy to before, *left* and *right* pieces of the corresponding trajectories of s, q_t and e_t when split at τ. Based on these quantities, my strategy to predict $\sigma^R_{T+1,\tau}$ from $\sigma^L_{T+1,\tau}$ as well as curves σ_t, $1 \leq t \leq T$, can be structured in two steps. In a first step, I predict the two shape components, s^R_τ and in particular $q^R_{T+1,\tau}$. In a second step I forecast the level component ν_{T+1}. The two steps are made concrete below.

Predicting the Shape Components

Intuitively the suggested technique can be motivated by an optimal but infeasible procedure. Suppose for a moment that all environments $\mathcal{P}_j$, $1 \leq j \leq J$ were known and that $T+1 \in \mathcal{P}_j$ for some j. Given reasonably accurate level approximations

$$\hat{\nu}_t := \int_0^1 \sigma_t(u)\mathrm{d}u,\ 1 \leq t \leq T,$$

the average of level-adjusted curves

$$\frac{1}{\left|\mathcal{P}_j^-\right|} \sum_{t \in \mathcal{P}_j^-} \left(\sigma^R_{t,\tau}(u) - \hat{\nu}_t\right) = \left(s^R_\tau(u) + \frac{1}{\left|\mathcal{P}_j^-\right|} \sum_{t \in \mathcal{P}_j^-} q^R_{t,\tau}(u) \right) + \frac{1}{\left|\mathcal{P}_j^-\right|} \sum_{t \in \mathcal{P}_j^-} \left((\nu_t - \hat{\nu}_t) + e^R_{t,\tau}(u)\right) \tag{2.13}$$

$$= s^R_\tau(u) + Q^R_{j,\tau}(u) + \frac{1}{\left|\mathcal{P}_j^-\right|} \sum_{t \in \mathcal{P}_j^-} \left((\nu_t - \hat{\nu}_t) + e^R_{t,\tau}(u)\right) \tag{2.14}$$

with $\mathcal{P}_j^- := \mathcal{P}_j \setminus \{T+1\}$ and $Q^R_{j,\tau}(u) = Q_j(u)$, $u \in [\tau, 1]$, ideally resembles $\sigma^R_{T+1,\tau} - \nu_{T+1}$. The average on the right hand side of (2.14) constitutes the prediction error, which is thought of as being small given a sufficiently large number of summands $|\mathcal{P}_j^-|$.

The ideal prediction in (2.13) however, is infeasible as the environments $\mathcal{P}_1, \ldots, \mathcal{P}_J$ are unknown. Nevertheless, information about the environment trading day $T+1$ belongs to can be inferred from shapes of volatility curves on $[0, \tau]$, which is motivated by my observation in Equation (2.11). Thus, the key to approximating the average in (2.13) is to identify for which curves $\{\sigma^L_{t,\tau} : 1 \leq t \leq T\}$ shapes are close to the one of $\sigma^L_{T+1,\tau}$. Intuitively I approach this problem by quantifying a metric from the sample up to day T, which distinguishes different shapes in an optimal way, whereas optimality refers to the underlying prediction problem. Then, at day $T+1$, $\sigma^L_{T+1,\tau}$ is compared in this optimal metric to the curves $\{\sigma^L_{t,\tau} : 1 \leq t \leq T\}$. The resulting distances are subsequently employed to form a weighted average resembling the ideal one in (2.13).

More precisely, I suggest a functional distance metric learning procedure in the spirit of recent

developments in the machine learning literature. There, the quantification of an optimal distance between (vector-valued) data points is for example popularly applied for image recognition such as face classification (see, for example, Guillaumin et al., 2009). Two recent surveys on multivariate distance metric learning are offered by Kulis (2013) and Bellet et al. (2013). While the majority of works are concerned with (binary) classification problems, Weinberger and Tesauro (2007) and Xiao et al. (2009) consider distance metric learning for regression. I use their technique rather than the classification procedures as a starting point, since in small or medium-sized samples the number J of patterns might be relatively large compared to T. From a pragmatic viewpoint one can also argue that the finite support of q_t is a potentially rough approximation for buckets of *similar* curves: from this less restrictive perspective the underlying situation does not define a classification problem.

Motivated by these observations and the framework in Weinberger and Tesauro (2007), I define for each τ a semi-metric d_α between two volatility curves $\sigma^L_{t,\tau}, \sigma^L_{s,\tau}$ on $[0,\tau]$ according to

$$d_\alpha(\sigma^L_{t,\tau}, \sigma^L_{s,\tau}) := \int_0^\tau \alpha(u) \left| \frac{\mathrm{d}\sigma^L_{t,\tau}(u)}{\mathrm{d}u} - \frac{\mathrm{d}\sigma^L_{s,\tau}(u)}{\mathrm{d}u} \right| \mathrm{d}u. \tag{2.15}$$

Here, α is a smooth functional weighting parameter, which is restricted to be weakly positive on $[0,\tau]$, i.e. $\alpha(u) \geq 0$ for any $u \in [0,\tau]$.[4] In theory this parameter can be constant over the whole interval, as differences between local shape components are sufficiently pronounced compared to the error term as guaranteed by Assumption A.4. For finite samples, however, the introduction of a non-constant weighting potentially improves the distinction between different shapes. Given a fixed α, of which the choice is discussed in a moment, the average in (2.13) can be approximated by

$$\sum_{t=1}^{T} \omega_{T+1,t} \left(\sigma^R_{t,\tau}(u) - \hat{\nu}_t \right), \tag{2.16}$$

where

$$\omega_{T+1,t} := \frac{k_{T+1,t}}{\sum_{s=1}^{T} k_{T+1,s}} \tag{2.17}$$

$$\text{and } k_{T+1,t} := \kappa \left(d_\alpha(\sigma^L_{t,\tau}, \sigma^L_{T+1,\tau}) \right). \tag{2.18}$$

Here κ denotes a univariate second order kernel function. In the definition of the weights in (2.18) I follow Weinberger and Tesauro (2007) and do not incorporate a bandwidth parameter, as

[4] For simplicity of notation, I drop the dependence of α on τ.

the distance function d_α implicitly defines an appropriate scaling when the norm of α is allowed to vary.

The average in (2.16) is my ultimate prediction for the shape components of $\sigma^R_{T+1,\tau}$. If α guarantees for any t, the distance $d_\alpha(\sigma^L_{t,\tau}, \sigma^L_{T+1,\tau})$ to be sufficiently large, if t and $T+1$ are in different environments, and close to zero if these days are in the same environment, the shape forecast in (2.16) is close to its ideal oracle counterpart in (2.13). The introduction of a non-constant weight α should hence more sharply distinguish different environments from information on $[0, \tau]$. I suggest to choose α in a data-driven way, such that this desire is well reflected in the sample up to day T. For this purpose I select α as the minimizer of the loss function

$$\mathcal{L}_T(\alpha) := \sum_{t=1}^{T} \int_{[\tau,1]} \left| \frac{\mathrm{d}\sigma^R_{t,\tau}(u)}{\mathrm{d}u} - \sum_{s=1, s\neq t}^{T} \tilde{\omega}_{t,s} \frac{\mathrm{d}\sigma^R_{s,\tau}(u)}{\mathrm{d}u} \right| \mathrm{d}u, \tag{2.19}$$

which depends on the parameter of interest through the weights $\tilde{\omega}_{t,s}$. The latter are in analogy to (2.17) defined as

$$\tilde{\omega}_{t,s} := \begin{cases} k_{t,s} / \sum_{r=1, r\neq t}^{T} k_{t,r} & \text{if } t \neq s \\ 0 & \text{if } t = s. \end{cases} \tag{2.20}$$

$$\text{and } k_{t,s} := \kappa\left(d_\alpha(\sigma^L_{s,\tau}, \sigma^L_{t,\tau})\right) \tag{2.21}$$

similar to a cross validation procedure. Intuitively, the loss function $\mathcal{L}_T(\alpha)$ measures the accuracy with which the local shape component of a curve $\sigma^R_{t,\tau}$ can be predicted from the remaining curves. I use derivative functions rather than the original curves as this allows to abstract from the level component, which needed to be estimated and thus introduced an additional error.

The weighted average $\sum_{s=1,s\neq t}^{T} \tilde{\omega}_{t,s} \frac{\mathrm{d}\sigma^R_{s,\tau}(u)}{\mathrm{d}u}$ in (2.19) resembles the prediction from a functional kernel regression (see e.g. Section 5.4 in Ferraty and Vieu, 2006) for pairs of derivative curves

$$\left(\frac{\mathrm{d}\sigma^L_{t,\tau}(u)}{\mathrm{d}u}, \frac{\mathrm{d}\sigma^R_{t,\tau}(u)}{\mathrm{d}u} \right), \ 1 \leq t \leq T$$

using kernel κ, semi-metric d_α and a bandwidth equal to 1. There are two key differences of my approach to a conventional functional kernel regression. First, the support of q_t contains only $J < \infty$ elements, which reduces the burden from a curse of dimensionality, which is extreme in the case of (infinite) dimensional functional data. Second the distance notion in a functional kernel regression is typically chosen a priori, while it is selected in a data-driven way in my approach.[5]

[5] See also Section 13.6 in Ferraty and Vieu (2006) for a discussion of the role of the distance metric in the context of functional kernel regressions.

Turning to the implementation, note that minimizing $\mathcal{L}_T(\alpha)$ is in principle an infinite dimensional optimization problem. In order to translate such optimization into a feasible problem, I restrict the search for an optimal α, which I required to be smooth, to a finite dimensional space. A particularly parsimonious though quite flexible specification within the class of smooth, continuous and positive functions is the density of the beta distribution. Inspired by this, I suggest to search for an optimal parameter α through the set of functions

$$\mathcal{A} := \left\{\alpha \in L^2([0,\tau]) : \ \exists a_1, a_2 > 0 : \ \alpha(u) = u^{(a_1-1)}(1-u)^{(a_2-1)} \ \forall u \in [0,\tau]\right\}.$$

Functions in $\mathcal{A}$ do not integrate to unity, because there is no scaling constant as in the case of the density of a beta distribution.[6] The choice of α thus comes along with an implicit bandwidth choice for computing the weights in (2.17) and (2.18). Weighting functions in $\mathcal{A}$ are completely characterized by the two parameters a_1 and a_2, and so one may denote $\alpha(u) = \alpha_{a_1,a_2}(u)$. The optimization of the loss function in (2.19) consequently translates into a bivariate minimization problem. More specifically, I define the optimal weighting scheme as $\alpha^* = \alpha_{a_1^*,a_2^*}$, where $(a_1^*, a_2^*) := \arg\min_{a_1,a_2>0} \mathcal{L}_T(\alpha_{a_1,a_2})$, or equivalently

$$\alpha^* := \arg\min_{\alpha\in\mathcal{A}} \mathcal{L}_T(\alpha). \tag{2.22}$$

From this optimal α, the final shape prediction obtains as in (2.16), whereas the weights are determined from (2.17) and (2.18), using d_{α^*} as the underlying distance function. Denote the resulting optimal weights $\omega_{T+1,t}^*$, $1 \leq t \leq T$. In order to obtain a forecast of $\sigma_{T+1,\tau}^R$, this shape prediction is supplemented with a level prediction, which is discussed next.

Predicting the Level Component

Adding to the above prediction of the shape of $\sigma_{T+1,\tau}^R$, my strategy to predict the level effect ν_{T+1} relies on two sources of information. The first one is the level of the available sub-trajectory $\sigma_{T+1,\tau}^L$. The second source of information is given by the time series variation of $\hat{\nu}_t$ over trading days $t = 1, \ldots, T$. I combine these two sources of information in a parametric time series model. As the true, unobserved level component ν_t was postulated to follow an ARMA(p,q)-process, I start from an ARMA(1,1) model for $\hat{\nu}_t$. As indicated above, I enrich such model by incorporating the vertical position of $\sigma_{T+1,\tau}^L$ as an additional covariate. The resulting ARMAX regression reads as

$$\hat{\nu}_t = \beta_0 + \beta_1\hat{\nu}_{t-1} + \beta_2(\hat{\nu}_{t,\tau}^L - \hat{\nu}_{t-1,\tau}^L) + \beta_3\epsilon_{t-1} + \epsilon_t, \tag{2.23}$$

[6] Also functions in $\mathcal{A}$ differ from the beta density as they are defined on $[0,\tau]$, rather than the unit interval.

where ϵ_t is an iid error term and $\hat{\nu}^L_{t,\tau} := \tau^{-1} \int_0^\tau \sigma^L_{t,\tau}(u)\mathrm{d}u$ the approximated level of $\sigma^L_{t,\tau}$.[7] Apart from the updating, which is specific to the online prediction problem at hand, the use of ARMA-models is embedded in the tradition of forecasting realized volatility by time series models (cf. Andersen et al., 2006). The additional covariate $(\hat{\nu}^L_{t,\tau} - \hat{\nu}^L_{t-1,\tau})$ in (2.23) can be thought of as enriching the information from the last trading day $t-1$ in the autoregressive term by the partial information observed at day t. Such partial information at day t, i.e. the estimated level $\hat{\nu}^L_{t,\tau}$, carries substantial information about $\hat{\nu}_t$, as can be seen from

$$\hat{\nu}_t \approx \hat{\nu}^L_{t,\tau} \cdot \tau + \int_\tau^1 \sigma^R_{t,\tau}(u)\mathrm{d}u. \tag{2.24}$$

The relation in (2.24) is only approximate due to the construction of the curves $\sigma^L_{t,\tau}$,$\sigma^R_{t,\tau}$ and σ_t from discrete price data, as will be discussed in Section 2.5. For the same reason, as well as the presence of an error function e_t in my component model, the estimated levels of $\sigma^L_{t,\tau}$ and $\sigma^R_{t,\tau}$ most likely differ in contrast to the true latent level in (2.12).
Assuming the disturbances in (2.23) additionally to be Gaussian, the parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ can be obtained using a standard likelihood based estimation procedure (cf. Durbin and Koopman, 2001). Based on these estimates model (2.23) allows to predict ν_{T+1} ($\hat{\nu}_{T+1}$) from all available information up to time τ at day $T+1$ according to

$$\breve{\nu}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\nu}_T + \hat{\beta}_2(\hat{\nu}^L_{T+1,\tau} - \hat{\nu}^L_{T,\tau}) + \hat{\beta}_3 \hat{\epsilon}_T, \tag{2.25}$$

where $\hat{\epsilon}_T$ denotes the residual of period T. The forecast in (2.25) can be expected to be more accurate for larger values of τ. Intuitively, this is because the information contained in the updating term $(\hat{\nu}^L_{T+1,\tau} - \hat{\nu}^L_{T,\tau})$ increases in τ, which is also likely to be reflected in the magnitude of $\hat{\beta}_2$ relative to $\hat{\beta}_1$.

Combining Shape and Level Predictions

The combination of the preceding shape and level predictions is straightforward. The final prediction of $\sigma^R_{T+1,\tau}$ obtains as

$$\hat{\sigma}^R_{T+1,\tau} := \breve{\nu}_{T+1} + \sum_{t=1}^{T} \omega^*_{T+1,t} \left(\sigma^R_{t,\tau}(u) - \hat{\nu}_t\right). \tag{2.26}$$

The prediction mechanism from above certainly depends on the point in time τ, until which there is data available. In practice the parameters α and β_j, $j = 0, 1, 2, 3$ should thus be quantified for

[7] For simplicity of notation I suppress the dependence of β_j, $j = 0, 1, 2, 3$ as well as ϵ_t on τ, which indeed might differ across different values of τ.

all desired prediction horizons in advance. Due to the parsimonious parameterization, however, the computational burden related to this is low.
The forecast in (2.26) does not explicitly match the levels nor the derivatives of $\sigma^L_{T+1,\tau}$ and $\hat{\sigma}^R_{T+1,\tau}$ at τ. I do not incorporate such matching, because of the construction of $\sigma^L_{T+1,\tau}$, which can only be recovered from prices observed until time τ. Hence the point $\sigma^L_{T+1,\tau}(\tau)$ is likely to be estimated with low accuracy, which is obviously problematic for its use in the prediction scheme. The underlying phenomenon is illustrated in the empirical work in Section 2.5.

For a fixed τ, the prediction $\hat{\sigma}^R_{T+1,\tau}$ can retrospectively be compared to $\sigma^R_{T+1,\tau}$ in order to assess the predictive accuracy. I discuss my strategy to evaluate this accuracy in the next section.

2.4 Forecast Evaluation

Evaluating the accuracy of conventional volatility forecasts is typically difficult as the underlying object of interest is latent. As in the case of realized volatility however, an ex post estimate of the intraday volatility curve can be computed once all information at day $T+1$ has been observed. This allows to use a retrospective comparison of the prediction and the ex post estimate as a performance measure. Hence, I compute the deviation of my forecast $\hat{\sigma}^R_{T+1,\tau}$ in (2.26) from the ex post estimate $\sigma^R_{T+1,\tau}$ according to

$$\int_\tau^1 \left(\hat{\sigma}^R_{T+1,\tau}(u) - \sigma^R_{T+1,\tau}(u)\right)^2 \mathrm{d}u, \tag{2.27}$$

which serves as a prediction error. Studying absolute forecast errors, however, is typically limited to the search for patterns and the examination of error dynamics over time. Relative errors, i.e. absolute errors compared to errors from other prediction mechanisms, allow in addition to interpret magnitudes. Hence I assess the performance of predictions generated by the component model by benchmarking. More specifically, I compare the error in (2.27) to the forecast errors of three different competitors. In detail these benchmarks obtain as follows.

1. The first competitor is an adaption of the updated `RiskMetrics` (RM) volatility prediction technique as outlined in Zumbach (2007). While it was initially designed for scalar or multivariate measures, such as realized (co-) volatility, its adaption to the context of spot volatility curves is straightforward. Formally, the prediction for $\sigma^R_{T+1,\tau}$ obtains as

$$\hat{\sigma}^{R,RM}_{T+1,\tau}(u) = \sum_{k=1}^{K} g_k \tilde{\sigma}^{R,k}_{T+1,\tau}(u), \tag{2.28}$$

where the functions $\tilde{\sigma}^{R,k}_{T+1,\tau}$ are predictions from a number of K different exponentially weighted moving averages (EWMA's) from past available curves $\sigma^R_{T-j,\tau}$, $j = 0, 1, \ldots, M$. These averages enter the final prediction $\hat{\sigma}^{R,RM}_{T+1,\tau}$ with different weights, denoted g_k. The latter are, in analogy to Proietti (2014), computed for $k = 1, \ldots, K$ according to

$$g_k = \log(\gamma_0/\gamma_k) \left[\sum_{l=1}^{K} \log(\gamma_0/\gamma_l) \right]^{-1}$$
$$\text{with} \quad \gamma_k = \gamma_1 \rho^{k-1},\ 2 \leq k \leq K,\ \gamma_0 = 1560,\ \gamma_1 = 4 \text{ and } \rho = 2^{1/2}.$$

The single EWMA's $\tilde{\sigma}^{R,k}_{T+1,\tau}$ constituting (2.28) differ in the parameter governing the decay of the weights associated with past volatility curves. More concrete the k-th EWMA is defined according to

$$\tilde{\sigma}^{R,k}_{T+1,\tau}(u) = \sum_{j=0}^{M} \frac{(1-\lambda_k)^j}{\sum_{l=0}^{M}(1-\lambda_k)^l} \sigma^R_{T-j,\tau}(u)$$
$$\text{where} \quad \lambda_k = 1 - \exp\{-\gamma_k^{-1}\}.$$

Thanks to this simple form of the individual EWMA's, the final prediction $\hat{\sigma}^{R,RM}_{T+1,\tau}$ can be re-stated as a weighted average of the curves $\sigma^R_{T,\tau}, \ldots, \sigma^R_{T-M,\tau}$.[8] In terms of my component model, the accuracy of $\hat{\sigma}^{R,RM}_{T+1,\tau}$ will depend on how well the dynamics of the level and local shape components can be approximated by this simple autoregressive structure.

2. As the second competitor I use a variant of the component model in Section 2.3. It relies on two simplifying assumptions, which are (i) variation in the spot volatility is dominantly due to a level-effect and (ii) dynamics of the level follow a random walk (RW). Given these two assumptions a simple predictor for $\sigma^R_{T+1,\tau}$ can be computed according to

$$\hat{\sigma}^{R,RW}_{T+1,\tau} = \hat{\nu}_T + T^{-1} \sum_{t=1}^{T} \left(\sigma^R_{t,\tau}(u) - \hat{\nu}_t \right). \tag{2.29}$$

The average in (2.29) distinguishes in particular from my forecast in (2.26) as it does not incorporate any information from day $T + 1$.

3. A last predictor addresses another case of the model in (2.3). Perceiving that variation in the spot volatility across trading days is exclusively due to the error term e_t, a suitable forecast

[8] Cf. Proietti (2014).

of $\sigma^R_{T+1,\tau}$ obtained as

$$\hat{\sigma}^{R,Mean}_{T+1,\tau} = T^{-1} \sum_{t=1}^{T} \sigma^R_{t,\tau}(u). \tag{2.30}$$

The prediction in (2.30) is simply the mean function of available spot volatility curves evaluated on $[\tau, 1]$. Clearly, it does not incorporate any of the dynamic or local features of the original component model.

In summary, the three competitors can be interpreted as addressing three particular cases of the model. The first competitor illuminates the case in which the level and the local shape component evolve according to a simple autoregressive structure over trading days. The second and third benchmarks postulate reduced dynamics of the level and the shape components.
Regarding the choice of the above competitors, I rely on practically motivated approaches rather than functional data models such as Müller et al. (2011) or Kraus (2015). The three benchmarks can readily be applied to the curves, which are constructed for the use in my prediction framework from Section 2.3. In particular they do not require any other estimation or data processing steps as would be the case with functional principal components for example.

From the competing forecasts in (2.28), (2.29) and (2.30) I finally calculate relative errors according to

$$E^{(j)}_{T+1,\tau} := \frac{\int_\tau^1 \left(\hat{\sigma}^R_{T+1,\tau}(u) - \sigma^R_{T+1,\tau}(u)\right)^2 \mathrm{d}u}{\int_\tau^1 \left(\hat{\sigma}^{R,j}_{T+1,\tau}(u) - \sigma^R_{T+1,\tau}(u)\right)^2 \mathrm{d}u}, \quad j \in \{RM, RW, Mean\}. \tag{2.31}$$

The errors $E^{(j)}_{T+1,\tau}$ indicate a superior performance of the estimation approach from Section 2.5.3, whenever they take values smaller than one. Beyond $E^{(j)}_{T+1,\tau}$, the difference $1 - E^{(j)}_{T+1,\tau}$ can be interpreted as *relative accuracy gain* of my forecast compared to benchmark j. Based on these relative errors (accuracy gains) I will interpret my results in the empirical study, which is discussed next.

2.5 Stock Volatility in the German DAX 30

In the following empirical study I confront my component model with data from the German stock market. For this purpose I begin with a brief description of the data. In a next step I discuss an

approach to construct curves σ_t, σ_t^R and σ_t^L from the intraday price data. The curves computed that way are subsequently used to examine the predictive accuracy of the procedure suggested in Section 2.3.2. I test my method *separately* for 29 stocks, rather than just for a single one, in order to avoid spurious findings. For each single stock, I also illustrate how predictions change when current time τ proceeds.

2.5.1 Data

The data I employ are intraday stock prices for 29 constituents of the German DAX 30 index.[9] The sample covers 140 trading days between $03-06-2016$ and $16-12-2016$ as well as another 59 days between $08-09-2017$ and $01-12-2017$. Data for the first part of the sample, which is used as a training sample, were gathered from Bloomberg, while data for the second part, the test sample, were extracted from Thomson Reuters Eikon.[10] For every stock and every trading day I obtain high and low prices over 1 minute-intervals for every minute between 09:00 a.m. and 05:30 p.m. I re-scale time to the unit interval $[0,1]$ complying with the preceding sections. In what follows I denote the standardized time stamp of the j-th minute $l_j \in [0,1]$. The first minute is labeled $l_1 = 0$ and the last minute at a trading day is labeled $l_n = 1$, with $n = 8.5 \cdot 60 = 510$. For each 1-minute interval I proceed with the average between highest and lowest price, say $P_{t,j}$ for the j-th minute at trading day t. As I operate the same way for all stocks, I leave out indexes labeling the single assets.

2.5.2 Recovering Spot Volatility Curves

From the discrete price data I obtain spot volatility curves separately for each stock at each trading day. The component model I introduced is, without further restrictions, not compatible with principal-components-based methods due to the lack of stationarity. Hence, I proceed in a different, more practically oriented vein, which can be structured into three steps. In a first step I calculate log-returns and remove a potential trend. In a second step the log-transformation suggested in Müller et al. (2011) is applied to the squared detrended returns. The resulting transformed data can be interpreted as noisy discrete points on the logarithmic spot volatility curve. In a third step I fit a standard nonparametric regression curve through these points, which yields my final construction of the (log-) spot volatility curve.

Formally this three step procedure, of which the first two steps mainly follow Müller et al. (2011), obtains as follows. Given prices, $P_{t,1}, \ldots, P_{t,n}$ at day t, the corresponding (scaled) log-returns are

[9] Due to missing data in the training sample, I exclude the Deutsche Börse Stock.

[10] I would like to thank the Institute for Financial Economics and Stastics at the University of Bonn as well as the ZEF/ZEI in Bonn for allowing me to access the data.

denoted $r_{t,j} := n^{1/2}(\log(P_{t,j}) - \log(P_{t,j-1}))$. In order to remove the effect of a potential drift term as in (2.1), I compute a Nadaraya-Watson smoother with GCV-optimal bandwidth and Gaussian kernel through the points $\{(l_j, r_{t,j}), j = 2, 3, \ldots, n\}$. Denoting the resulting fit at design points l_j as $\hat{\mu}_t(l_j)$, I proceed with calculating de-trended returns $r'_{t,j} := r_{t,j} - \hat{\mu}_t(l_j)$. The latter are subsequently transformed according to

$$V_{t,j} := \log\left((r'_{t,j})^2\right) + 1.27,$$

which can be interpreted as noisy values of the logarithmic spot volatility curve at points l_j. For the last step, I use once more a Nadaraya-Watson regression with Gaussian kernel, this time through the points $\{(l_j, V_{t,j}),\ j = 2, 3, \ldots, n\}$ to obtain a smooth regression curve. Again, for bandwidth selection I rely on Generalized Cross Validation. The resulting curve is, by slight abuse of notation, denoted σ_t.[11]

From the constructed curves σ_t, I *sample* the sub-trajectories $\sigma^R_{t,\tau}(u)$ by simply evaluating σ_t on $[\tau, 1]$. Regarding the left, i.e. the initial part $\sigma^L_{t,\tau}(u)$, of the spot volatility curve, note that σ_t, is smoothed using a symmetric kernel. Thus, values of the fitted curve $\sigma_t(u)$, with $u \leq \tau$ close to τ include information from returns sampled after τ, i.e. returns indexed by $l_j > \tau$. This however does not reflect the situation in $T + 1$, where τ labels *current time*. Although information after τ is certainly available for past curves, I mimic the construction mechanism on $[0, \tau]$ at a day $T + 1$ also for curves in the training sample. In this spirit, I apply the above three step procedure again, this time to price data $\{P_{t,j} :\ 0 \leq l_j \leq \tau\}$, to obtain the curves $\sigma^L_{t,\tau}$. Due to the separate construction of $\sigma^L_{t,\tau}$ and $\sigma^R_{t,\tau}$, the points $\sigma^L_{t,\tau}(\tau)$ and $\sigma^R_{t,\tau}(\tau)$ not necessarily coincide, as I alluded to in Part 2.3.2. In the first place, this reflects the difference between latent population curves and their empirical counterparts, which are estimated from discrete price data. I illustrate the behavior of $\sigma^L_{t,\tau}$ and $\sigma^R_{t,\tau}$ close to τ as well as ten exemplary volatility curves σ_t in Figure 2.1.
I test the accuracy of the predictions from my component model for different levels of τ. More concrete, I consider values of τ in the set $\{0.4, 0.5, 0.6, 0.7\}$. Thus, the curves $\sigma^L_{t,\tau}$ are fitted for each τ separately, while the corresponding right-hand-side parts $\sigma^R_{t,\tau}$ can simply be obtained by *sampling* for each τ from the complete trajectory σ_t. Next I employ the curves $\sigma^L_{t,\tau}$ and $\sigma^R_{t,\tau}$ in the prediction context outlined in the preceding sections.

[11] In order to improve the readability, I do not distinguish explicitly between the true, latent curves and the curves constructed from discrete data.

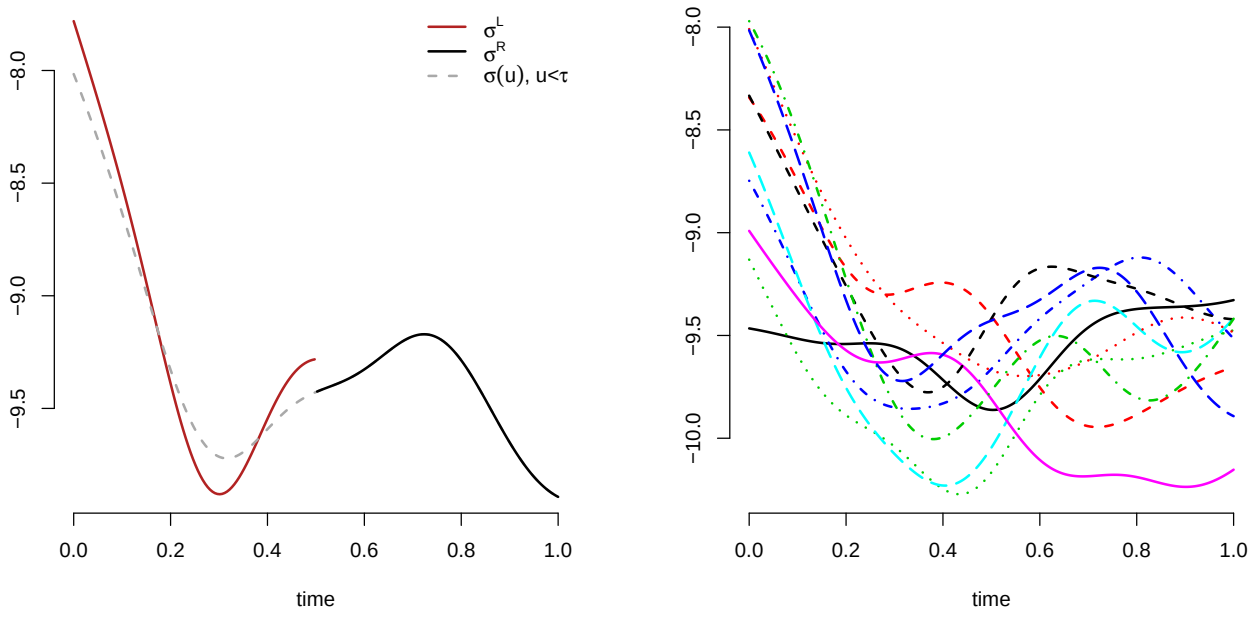

Figure 2.1: **Exemplary Volatility Curves.** *Left Panel: constructed logarithmic spot volatility curves* $\sigma^L_{t,\tau}, \sigma^R_{t,\tau}$ *and* σ_t *for the Adidas Stock at* $2016-06-03$. *Right Panel: ten exemplary logarithmic spot volatility curves for the Adidas stock between* $2016-06-03$ *and* $2016-06-16$.

2.5.3 Prediction

As indicated, I split the available data in a training and a test sample. The former collects the first available $T = 140$ trading days in the set $\mathbf{Tr} := \{1, \ldots, T\}$. The latter collects the last 59 trading days in the set $\mathbf{Te} := \{T+1, \ldots, T+59\}$. The training sample is used for two purposes. First I obtain for every considered level of τ the functional weight α^* and estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ from the curves in $\mathbf{Tr}$ as described in Section 2.3. Second, I use these curves to calculate the shape prediction in (2.26). For quantifying the functional weighting parameter and also for the final shape prediction, I choose a Gaussian kernel for κ in (2.18) and (2.21). Given the information from the training sample, I use my technique described in Section 2.3 to obtain forecasts for the curves in the test sample $\mathbf{Te}$. One-step-ahead level predictions are obtained iteratively through $\mathbf{Te}$, while the ARMAX parameters are not re-estimated. Also, I do not predict the first curve in the test sample because of the temporal gap between $\mathbf{Tr}$ and $\mathbf{Te}$.

I generate one-step-ahead predictions from the three competitors introduced in Section 2.4 in the following configuration. For the `RiskMetrics`-type prediction, I set the number of EWMA's equal to $K = 5$ and the window length to $M = 9$. In order to guarantee a fair comparison, I

calculate relative errors only for days $T+10,\ldots,T+59$ in **Te**. The EWMA-based forecasts thus rely exclusively on data from the test sample. Predictions $\sigma^{R,RW}_{T+h,\tau}$ from the random walk benchmark are calculated for $h=2,3,\ldots,59$. I iteratively update the level forecast and also the function valued term in (2.29) according to

$$\hat{\sigma}^{R,RW}_{T+h,\tau} = \hat{\nu}_{T+h-1} + (T+h-1)^{-1}\sum_{t=1}^{T+h-1}\left(\sigma^{R}_{t,\tau}(u) - \hat{\nu}_t\right), \text{ for } h=2,\ldots,59. \tag{2.32}$$

In analogy to this, I also update the third competitor, i.e. the mean function, when proceeding in the test sample. Such updating operates according to

$$\hat{\sigma}^{R,Mean}_{T+h,\tau} = (T+h-1)^{-1}\sum_{t=1}^{T+h-1}\sigma^{R}_{t,\tau}(u), \text{ for } h=2,\ldots,59. \tag{2.33}$$

Turning to the results, I will focus on medians of relative errors $E^{(j)}_{T+h,\tau}$ as the key performance indicator. In the first place I consider median relative errors on the stock level. For a comparison across different values of τ, I further consider medians of the pooled errors, where pooling proceeds over time and assets. I consider medians rather than means due to their robustness with respect to outliers, which are likely to occur asymmetrically as relative errors are bounded below by 0. My strategy to deal with different values of current time τ, is to begin the discussion with a reference case in which $\tau = 0.5$. Subsequently I examine changes in the results when τ takes values in $\{0.4, 0.6, 0.7\}$.

For the case $\tau = 0.5$ Figure 2.2 displays the key-results. Detailed stock-level findings are provided in Tables 2.B.4-2.B.6 in the appendix of this chapter.

The median relative errors visualized in Figure 2.2 document the success of my novel prediction technique. The most serious competitor is, as expected, the combination of exponentially weighted moving averages as in the `RiskMetrics` framework. Nevertheless this competitor is outperformed in median by up to 31% (BASF) which is, in the context of predicting stock volatility, a substantial amount. My predictions beat the EWMA-based competitor for 24 out of the 29 stocks. For 9 stocks the relative accuracy gain is greater than 10%, for 16 greater than 5%. The worst performance is delivered for the ProSiebenSat.1 Media and the Beiersdorf stocks (for both the median relative error is about 1.09). The average of median relative errors however is at 90% and thus well below one.

The solid performance of the predictions from my component model is even more visible for the random walk competitor. The relative accuracy gain ranges up to 56% in median (BASF), while for only 4 out of 29 stocks the performance is slightly worse [median errors: 1.01 (DTE), 1.02 (HEN),

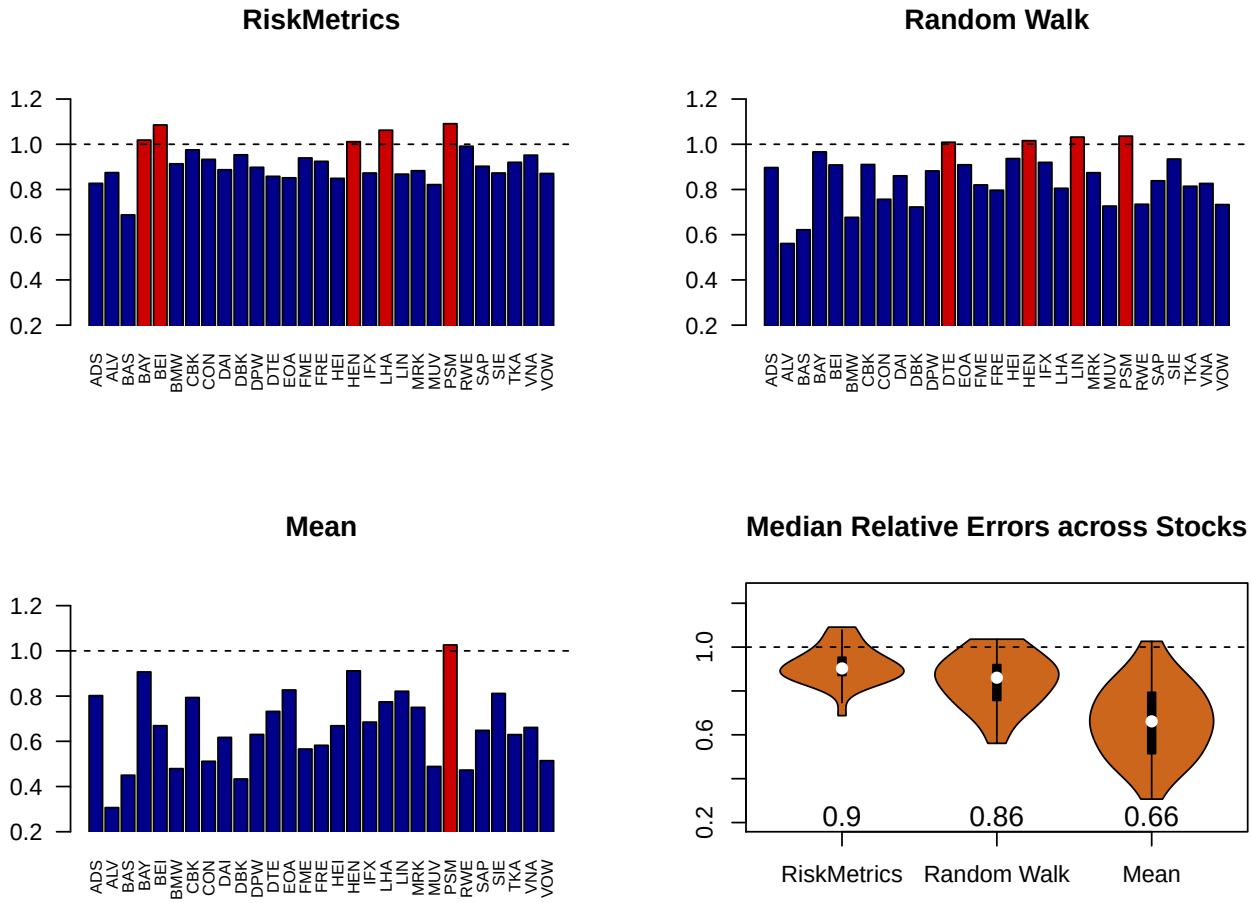

Figure 2.2: **Median Relative Performances for $\tau = 0.5$.** *Figure shows median relative errors for different competitors. The x-axis in the barplots is constituted by the different assets (see Table 2.A.1 in the appendix). Blue indicates bars below* 1*, red bars are above* 1*. The violin plots in the bottom-right panel depict the distribution of the medians in the barplots over assets. Numbers below the violins are means of medians.*

1.03 (LIN) and 1.04 (PSM)]. Averaged over the 29 stocks, the median relative error amounts to approximately 86%, or, equivalently, a performance gain of about 14%. This is 4 percentage points better than the corresponding number for the EWMA-based competitor. Compared to the latter, relative errors are, overall, also slightly less dispersed (see Table 2.1).

My forecasting technique most clearly outperforms the mean, which is the simplest competitor. For all but the ProSiebenSat.1 Media stock, predictions from my model are, in median, better. Median relative errors are for 6 stocks even below 50% and for 18 stocks below 70%. The best performance amounts to a median relative error of 31% for the Allianz stock, while the average of median relative errors is at 66%. Overall, my novel procedure thus appears to be substantially

	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
$\underline{\tau = 0.4}$								
RM	0.06	0.47	0.67	**0.93**	1.28	1.70	0.60	10.54
RW	0.14	0.50	0.67	**0.89**	1.11	1.44	0.44	5.23
Mean	0.06	0.30	0.47	**0.69**	0.99	1.36	0.52	7.42
$\underline{\tau = 0.5}$								
RM	0.07	0.44	0.65	**0.91**	1.25	1.74	0.60	11.02
RW	0.07	0.43	0.61	**0.86**	1.10	1.49	0.48	6.46
Mean	0.05	0.27	0.44	**0.67**	1.00	1.43	0.56	6.75
$\underline{\tau = 0.6}$								
RM	0.07	0.40	0.62	**0.89**	1.25	1.78	0.63	9.46
RW	0.08	0.36	0.53	**0.82**	1.12	1.71	0.59	9.30
Mean	0.05	0.25	0.39	**0.64**	1.03	1.60	0.64	8.28
$\underline{\tau = 0.7}$								
RM	0.05	0.33	0.57	**0.85**	1.27	1.85	0.71	16.28
RW	0.06	0.25	0.44	**0.76**	1.26	2.08	0.82	14.16
Mean	0.04	0.19	0.33	**0.62**	1.13	1.98	0.79	10.94

Table 2.1: **Summary of Pooled Relative Performances.** *Table summarizes the distribution of pooled relative errors for different competitors. IQR refers to the interquartile range. Pooling proceeds over stocks and periods in the test sample.*

more accurate than the mean for the considered prediction problem.

Turning to the relative performance for values of τ in $\{0.4, 0.6, 0.7\}$, I present the major results in Table 2.1 and Figure 2.3. In order to get a more tractable error measure I pool, for a given τ, relative errors for each competitor over trading days and assets. I also provide stock level results in analogy to the case of $\tau = 0.5$ in the appendix (see Tables 2.B.1-2.B.3, Tables 2.B.7-2.B.12 and Figures 2.B.1-2.B.3).

Overall the impression from the reference case $\tau = 0.5$ qualitatively transfers to the remaining cases. For all considered values of τ my predictions clearly outperform all competitors. In analogy to the above discussion, the `RiskMetrics`-type competitor is for all values of τ the most serious one. Relative accuracy gains are a bit higher for the random walk competitor and striking for the mean.

Interestingly, the relative performance of my predictions improves as τ increases. This holds true across all competitors. When comparing the cases $\tau = 0.4$ and $\tau = 0.7$ for example, the medians

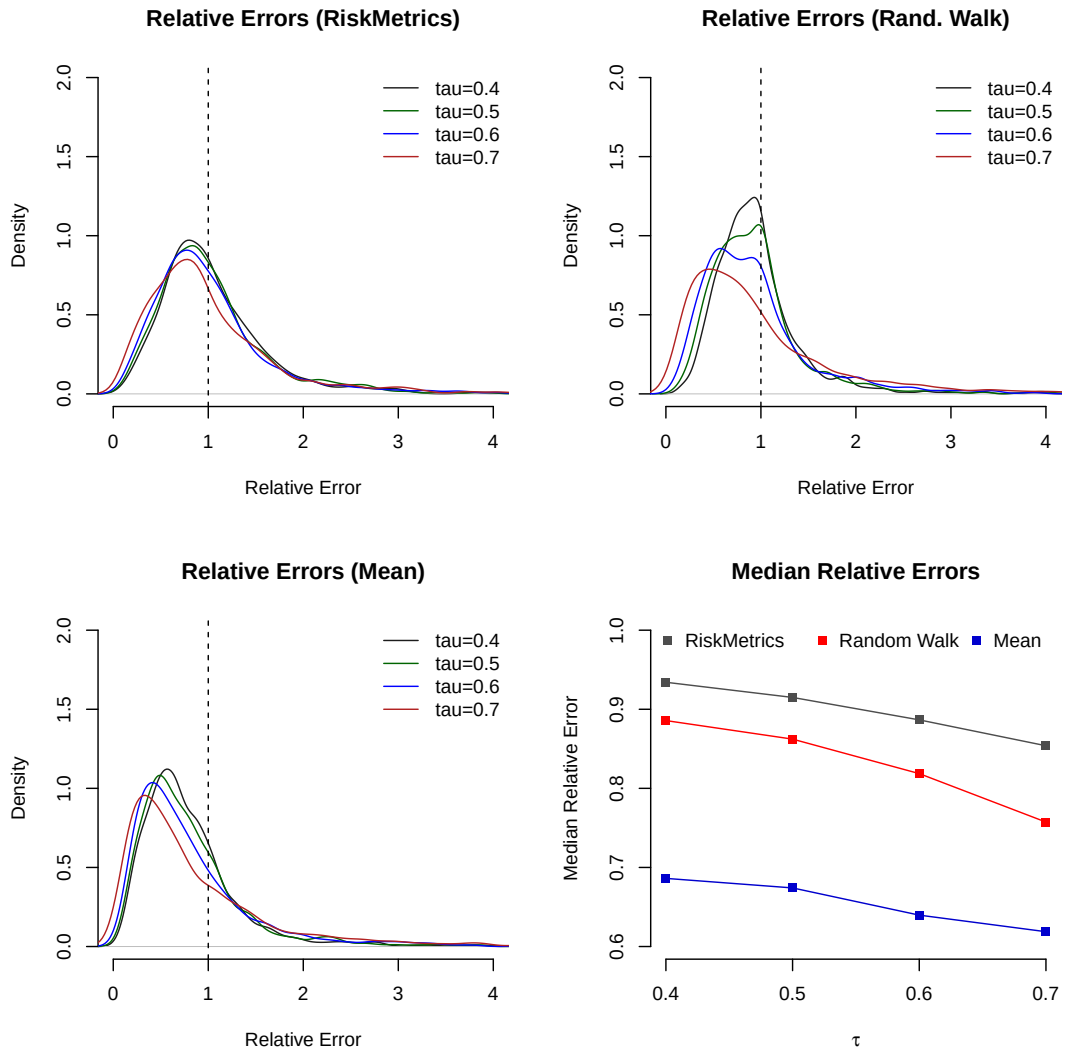

Figure 2.3: **Pooled Relative Performances across different τ's.** *Figure reports the distributions of pooled relative errors for different competitors. The first three panels show standard kernel density estimates of the corresponding empirical distributions.*

of pooled relative errors drop between 7 percentage points for the mean and 13 percentage points for the random walk competitor. This phenomenon appears to be monotonic in τ. At the same time however, interquartile ranges of the relative errors tend to increase, also monotonically in τ. The error distributions in Figure 2.3 offer a detailed view on the effect of increasing the value of τ. Supporting the above finding, probability mass shift towards zero as τ increases. This goes along with decreasing medians, but also with increasing probability mass in the right tail of the distributions.

These findings are also reflected in the number of assets for which a competitor is (in median) outperformed. As documented in Table 2.2 the number of successful cases grows in almost all

configurations as τ becomes larger.

τ	RiskMetrics	Random Walk	Mean
0.4	23	26	28
0.5	24	25	28
0.6	25	27	29
0.7	27	29	29

Table 2.2: **Successful Cases.** *Table reports for each τ and each competitor the number of stocks, for which the median relative error was below* 1.

The documented phenomenon of increasing performance of my technique for larger values of τ matches the intuition: in contrast to the competitors, the prediction in (2.26) takes the increasing information content in $\sigma^L_{T+1,\tau}$ explicitly into account. For increasing τ the importance of this information obviously increases, relative to the information from previous curves.

2.6 Conclusion

In this Chapter I considered the problem of predicting the smooth daily spot volatility curve of intraday stock price returns from partially observed information. For this purpose I introduced a novel component model. The model attributes the evolution of spot volatility over trading days to two systematic effects: a level effect and a shape effect. I model the former as a scalar level component which evolves according to a parametric time series process across trading days. I composed the shape effect by a global and a local shape component. The local shape component models random dynamic deviations from the deterministic curve-valued global shape component. Together they drive the form of the spot volatility curve beyond its vertical position. As a major novelty in this work, the local shape component was a recurring phenomenon, meaning that deviations of the spot volatility shape from its global shape component repeatedly occur over trading days.

I used the implications of my component model to design a two-step prediction mechanism to tackle the forecasting problem. This mechanism relies on separate level and shape predictions. The former was generated by an ARMA-model, which I enriched by an updating term. Such term incorporated already observed information at the day of interest. I constructed the shape prediction from a weighted average of past level-adjusted curves. The weights in this average were determined by a novel distance metric learning procedure. The resulting semi-metric between sub-trajectories of spot volatility curves allowed to identify similar local shape components across curves.

In an empirical study I tested the performance of this mechanism for 29 stocks in the German DAX 30 index. The results indicated a decent accuracy of the predictions, when benchmarked against three competitors. All of these were outperformed for the vast majority of stocks, for all considered amounts of partially observed information. I found that the relative performance of my predictions increased with the amount of available information at the day of interest: the more returns had been observed already during the day, the more accurate my predictions became relative to the competing approaches.

The current work points at a number of issues left for future research. In particular the exploration of theoretical properties of the introduced distance metric learning technique is an important topic. In this context I also consider an extension of this work to a panel data framework to be potentially beneficial. From an empirical viewpoint, the inclusion of additional explanatory variables appears to be particularly promising: it is likely that other variables such as the trading volume or liquidity measures carry predictive content for the continuation of the spot volatility curve.

Appendix

2.A Company Shortcuts

i	Shortcut	Full Name
1	ADS	Adidas
2	ALV	Allianz
3	BAS	BASF
4	BAY	Bayer
5	BEI	Beiersdorf
6	BMW	BMW
7	CBK	Commerzbank
8	CON	Continental
9	DAI	Daimler
10	DBK	Deutsche Bank
11	DPW	Deutsche Post
12	DTE	Deutsche Telekom
13	EOA	E.ON
14	FME	Fresenius Medical Care
15	FRE	Fresenius
16	HEI	HeidelbergCement
17	HEN	Henkel
18	IFX	Infineon Technologies
19	LHA	Deutsche Lufthansa
20	LIN	Linde
21	MRK	Merck
22	MUV	MunichRe
23	PSM	ProSiebenSat.1 Media
24	RWE	RWE
25	SAP	SAP
26	SIE	Siemens
27	TKA	ThyssenKrupp
28	VNA	Vonovia
29	VOW	Volkswagen

Table 2.A.1: **Company Shortcuts.**

2.B Stock Level Results

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.28	0.52	0.63	**0.87**	1.15	1.66	0.52	8.37
ALV	0.16	0.37	0.69	**0.93**	1.17	1.54	0.48	3.20
BAS	0.25	0.42	0.52	**0.74**	1.06	1.39	0.54	10.54
BAY	0.43	0.59	0.70	**0.97**	1.30	1.53	0.60	2.03
BEI	0.30	0.57	0.79	**1.16**	1.60	2.01	0.80	4.46
BMW	0.21	0.40	0.64	**0.99**	1.36	1.67	0.72	5.13
CBK	0.39	0.65	0.79	**1.01**	1.32	1.78	0.53	3.69
CON	0.29	0.40	0.68	**0.95**	1.27	1.51	0.60	3.03
DAI	0.11	0.47	0.62	**0.85**	1.41	1.95	0.79	4.74
DBK	0.19	0.44	0.68	**0.95**	1.32	1.69	0.63	4.19
DPW	0.33	0.56	0.64	**0.90**	1.42	2.08	0.78	4.05
DTE	0.14	0.49	0.63	**0.91**	1.20	1.57	0.57	2.92
EOA	0.13	0.48	0.63	**0.88**	1.27	1.73	0.64	2.37
FME	0.15	0.49	0.69	**0.88**	1.26	1.75	0.58	3.00
FRE	0.32	0.54	0.69	**1.00**	1.46	1.69	0.77	4.71
HEI	0.21	0.55	0.67	**0.90**	1.14	1.50	0.47	2.65
HEN	0.39	0.75	0.92	**1.02**	1.19	1.29	0.26	2.01
IFX	0.12	0.43	0.58	**0.90**	1.39	1.84	0.81	2.51
LHA	0.24	0.57	0.71	**1.06**	1.41	2.29	0.70	3.01
LIN	0.33	0.55	0.68	**0.94**	1.09	1.39	0.41	2.56
MRK	0.23	0.47	0.70	**0.92**	1.12	1.45	0.42	2.55
MUV	0.11	0.44	0.71	**0.87**	1.29	1.54	0.58	1.83
PSM	0.25	0.48	0.76	**1.12**	1.45	1.86	0.70	4.62
RWE	0.06	0.30	0.61	**1.02**	1.41	1.76	0.80	3.74
SAP	0.28	0.42	0.65	**0.85**	1.20	1.49	0.55	3.83
SIE	0.21	0.56	0.78	**0.96**	1.22	1.69	0.44	2.34
TKA	0.13	0.44	0.70	**0.93**	1.22	1.46	0.52	2.75
VNA	0.41	0.59	0.74	**0.99**	1.19	1.50	0.45	3.22
VOW	0.21	0.40	0.59	**0.89**	1.20	2.02	0.61	3.03

Table 2.B.1: L^2**-Errors Relative to Adapted RiskMetrics Competitor for** $\tau = 0.4$**.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the* `RiskMetrics`*-type competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.33	0.57	0.71	**0.86**	1.14	1.42	0.42	4.01
ALV	0.32	0.40	0.46	**0.67**	0.96	1.39	0.51	5.23
BAS	0.17	0.38	0.49	**0.73**	0.96	1.23	0.47	1.89
BAY	0.39	0.70	0.78	**0.97**	1.07	1.42	0.28	1.87
BEI	0.31	0.53	0.69	**0.93**	1.13	1.31	0.43	1.90
BMW	0.29	0.46	0.58	**0.78**	0.97	1.26	0.38	2.65
CBK	0.37	0.60	0.78	**0.94**	1.17	1.40	0.39	2.54
CON	0.21	0.45	0.64	**0.85**	1.14	1.68	0.50	2.26
DAI	0.24	0.53	0.65	**0.90**	1.12	1.73	0.47	3.33
DBK	0.14	0.48	0.61	**0.80**	0.98	1.20	0.36	2.37
DPW	0.31	0.52	0.70	**0.90**	1.02	1.27	0.32	1.59
DTE	0.36	0.57	0.70	**0.98**	1.18	1.56	0.48	4.55
EOA	0.41	0.61	0.77	**0.92**	1.25	1.47	0.47	2.75
FME	0.28	0.50	0.63	**0.88**	1.10	1.34	0.47	1.59
FRE	0.32	0.51	0.66	**0.83**	1.04	1.51	0.37	3.86
HEI	0.21	0.49	0.70	**0.93**	1.10	1.42	0.40	1.87
HEN	0.47	0.76	0.92	**1.02**	1.12	1.28	0.20	1.62
IFX	0.34	0.55	0.71	**0.93**	1.11	1.79	0.40	3.92
LHA	0.36	0.56	0.71	**0.89**	1.10	1.28	0.39	2.95
LIN	0.49	0.66	0.81	**1.01**	1.17	1.51	0.36	1.88
MRK	0.39	0.52	0.73	**0.89**	1.04	1.54	0.32	2.30
MUV	0.18	0.37	0.63	**0.89**	1.20	1.46	0.57	3.36
PSM	0.18	0.63	0.75	**1.01**	1.35	1.74	0.61	3.31
RWE	0.15	0.42	0.56	**0.83**	1.13	1.35	0.57	2.19
SAP	0.36	0.56	0.66	**0.85**	1.11	1.31	0.45	2.33
SIE	0.40	0.61	0.83	**0.97**	1.16	1.39	0.33	3.38
TKA	0.18	0.54	0.61	**0.80**	0.97	1.32	0.36	2.74
VNA	0.23	0.51	0.75	**0.88**	1.08	1.42	0.33	4.37
VOW	0.27	0.39	0.52	**0.76**	1.03	1.90	0.51	3.24

Table 2.B.2: **L^2-Errors Relative to Random Walk Competitor for $\tau = 0.4$.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from random walk / constant shape competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.12	0.44	0.59	**0.83**	1.16	1.67	0.57	2.74
ALV	0.12	0.18	0.23	**0.33**	0.63	1.12	0.40	7.42
BAS	0.11	0.21	0.29	**0.48**	0.92	1.21	0.63	4.14
BAY	0.43	0.56	0.74	**0.87**	1.05	1.30	0.31	2.54
BEI	0.20	0.35	0.52	**0.67**	0.98	1.20	0.46	3.64
BMW	0.16	0.23	0.31	**0.52**	0.75	1.03	0.44	2.05
CBK	0.37	0.51	0.63	**0.79**	1.01	1.10	0.39	2.03
CON	0.09	0.24	0.40	**0.52**	0.82	1.00	0.41	1.43
DAI	0.09	0.30	0.41	**0.63**	0.87	1.33	0.46	2.55
DBK	0.09	0.23	0.31	**0.44**	0.60	0.83	0.29	2.25
DPW	0.19	0.35	0.51	**0.65**	0.97	1.51	0.47	2.17
DTE	0.24	0.41	0.52	**0.76**	0.99	1.54	0.47	3.16
EOA	0.12	0.47	0.59	**0.86**	1.16	1.44	0.58	2.69
FME	0.20	0.26	0.38	**0.56**	0.78	0.90	0.39	1.49
FRE	0.13	0.37	0.44	**0.59**	1.05	1.40	0.61	3.32
HEI	0.21	0.41	0.53	**0.74**	0.96	1.32	0.43	2.69
HEN	0.32	0.68	0.75	**0.91**	1.13	1.34	0.37	2.39
IFX	0.11	0.31	0.42	**0.65**	1.05	1.58	0.63	2.33
LHA	0.19	0.43	0.56	**0.79**	0.96	1.26	0.39	1.59
LIN	0.34	0.56	0.65	**0.87**	1.03	1.28	0.38	3.82
MRK	0.26	0.47	0.57	**0.76**	0.98	1.30	0.41	3.32
MUV	0.06	0.23	0.36	**0.53**	0.75	1.11	0.39	2.45
PSM	0.20	0.43	0.78	**1.06**	1.57	2.70	0.80	3.94
RWE	0.09	0.25	0.36	**0.59**	0.86	1.35	0.50	3.21
SAP	0.18	0.27	0.42	**0.80**	1.06	1.40	0.64	2.81
SIE	0.46	0.57	0.70	**0.88**	1.04	1.40	0.35	2.14
TKA	0.20	0.36	0.46	**0.60**	0.82	1.69	0.36	4.70
VNA	0.12	0.39	0.50	**0.66**	0.91	1.15	0.42	2.21
VOW	0.15	0.25	0.36	**0.51**	0.78	1.17	0.42	1.66

Table 2.B.3: L^2**-Errors Relative to Mean for** $\tau = 0.4$**.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the mean (constant level / constant shape competitor) as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.26	0.45	0.68	**0.83**	1.11	1.70	0.43	8.57
ALV	0.08	0.31	0.65	**0.87**	1.34	1.69	0.68	2.92
BAS	0.15	0.39	0.50	**0.69**	1.22	1.84	0.72	11.02
BAY	0.26	0.60	0.72	**1.02**	1.24	1.62	0.53	3.63
BEI	0.23	0.55	0.81	**1.09**	1.66	2.48	0.86	4.50
BMW	0.22	0.40	0.55	**0.91**	1.43	1.81	0.88	2.49
CBK	0.30	0.56	0.77	**0.98**	1.30	1.69	0.53	3.85
CON	0.13	0.46	0.65	**0.93**	1.21	1.52	0.56	3.03
DAI	0.11	0.45	0.63	**0.89**	1.27	2.24	0.65	3.83
DBK	0.25	0.48	0.68	**0.95**	1.31	1.67	0.63	3.52
DPW	0.26	0.44	0.67	**0.90**	1.28	2.25	0.61	3.35
DTE	0.16	0.47	0.64	**0.86**	1.14	1.61	0.50	2.32
EOA	0.20	0.45	0.59	**0.85**	1.16	1.64	0.57	2.38
FME	0.13	0.47	0.65	**0.94**	1.29	1.56	0.64	2.66
FRE	0.22	0.54	0.68	**0.92**	1.44	1.80	0.76	4.35
HEI	0.20	0.44	0.66	**0.85**	1.24	1.36	0.58	2.34
HEN	0.36	0.79	0.91	**1.01**	1.21	1.48	0.30	1.91
IFX	0.18	0.41	0.63	**0.87**	1.28	1.80	0.65	3.99
LHA	0.22	0.53	0.65	**1.06**	1.52	2.45	0.87	3.48
LIN	0.24	0.51	0.69	**0.87**	1.11	1.35	0.43	2.63
MRK	0.29	0.49	0.66	**0.88**	1.14	1.59	0.48	4.91
MUV	0.15	0.43	0.62	**0.82**	1.09	1.44	0.46	1.76
PSM	0.22	0.46	0.74	**1.09**	1.44	2.06	0.71	6.59
RWE	0.07	0.29	0.49	**0.99**	1.41	1.85	0.92	2.93
SAP	0.18	0.38	0.61	**0.90**	1.11	1.38	0.51	2.78
SIE	0.18	0.42	0.72	**0.87**	1.13	1.58	0.41	2.37
TKA	0.11	0.45	0.71	**0.92**	1.15	1.58	0.44	3.97
VNA	0.26	0.51	0.72	**0.95**	1.23	1.49	0.50	4.68
VOW	0.16	0.37	0.59	**0.87**	1.27	1.95	0.68	2.55

Table 2.B.4: **L^2-Errors Relative to Adapted RiskMetrics Competitor for $\tau = 0.5$.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the `RiskMetrics`-type competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.25	0.50	0.66	**0.90**	1.13	1.43	0.47	3.85
ALV	0.12	0.31	0.40	**0.56**	0.86	1.71	0.46	3.88
BAS	0.20	0.31	0.38	**0.62**	0.99	1.32	0.61	2.00
BAY	0.49	0.71	0.82	**0.97**	1.13	1.24	0.31	1.47
BEI	0.28	0.47	0.68	**0.91**	1.09	1.37	0.40	2.65
BMW	0.26	0.36	0.49	**0.68**	0.99	1.19	0.50	2.49
CBK	0.37	0.54	0.79	**0.91**	1.12	1.38	0.33	2.13
CON	0.20	0.38	0.54	**0.76**	1.11	1.61	0.57	2.69
DAI	0.19	0.44	0.63	**0.86**	1.26	1.72	0.63	4.54
DBK	0.07	0.39	0.55	**0.72**	0.91	1.24	0.36	2.42
DPW	0.29	0.38	0.60	**0.88**	1.06	1.34	0.46	2.45
DTE	0.27	0.51	0.63	**1.01**	1.25	1.58	0.62	6.46
EOA	0.38	0.57	0.73	**0.91**	1.18	1.46	0.44	3.35
FME	0.28	0.44	0.52	**0.82**	1.07	1.29	0.54	1.56
FRE	0.31	0.47	0.63	**0.80**	0.99	1.72	0.35	3.03
HEI	0.18	0.46	0.64	**0.94**	1.05	1.41	0.41	1.92
HEN	0.43	0.73	0.87	**1.02**	1.17	1.51	0.30	2.04
IFX	0.18	0.52	0.62	**0.92**	1.08	1.97	0.47	4.25
LHA	0.41	0.56	0.65	**0.81**	1.11	1.54	0.46	3.93
LIN	0.34	0.61	0.85	**1.03**	1.14	1.41	0.29	3.69
MRK	0.47	0.59	0.73	**0.87**	1.04	1.53	0.31	3.42
MUV	0.16	0.38	0.54	**0.73**	1.11	1.39	0.57	4.35
PSM	0.19	0.57	0.83	**1.04**	1.37	1.93	0.54	3.79
RWE	0.12	0.37	0.49	**0.73**	0.96	1.20	0.47	2.02
SAP	0.32	0.43	0.55	**0.84**	1.04	1.37	0.48	2.76
SIE	0.41	0.67	0.78	**0.93**	1.09	1.31	0.31	4.00
TKA	0.16	0.45	0.65	**0.81**	1.02	1.46	0.38	4.12
VNA	0.11	0.45	0.68	**0.83**	1.11	1.75	0.43	3.08
VOW	0.19	0.35	0.51	**0.73**	1.08	1.73	0.56	4.00

Table 2.B.5: L^2**-Errors Relative to the Random Walk Competitor for** $\tau = 0.5$**.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from random walk / constant shape competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.16	0.39	0.53	**0.80**	1.20	1.67	0.66	2.74
ALV	0.06	0.17	0.23	**0.31**	0.59	1.34	0.36	6.75
BAS	0.13	0.19	0.28	**0.45**	0.90	1.19	0.62	4.54
BAY	0.42	0.62	0.72	**0.91**	1.07	1.44	0.34	1.91
BEI	0.21	0.37	0.52	**0.67**	1.06	1.52	0.54	3.66
BMW	0.13	0.20	0.31	**0.48**	0.76	1.05	0.45	1.91
CBK	0.37	0.45	0.65	**0.79**	1.03	1.24	0.38	1.69
CON	0.13	0.21	0.37	**0.51**	0.80	1.03	0.43	1.97
DAI	0.09	0.26	0.40	**0.62**	0.92	1.54	0.53	2.79
DBK	0.05	0.21	0.33	**0.43**	0.54	0.96	0.21	2.60
DPW	0.19	0.33	0.45	**0.63**	1.02	1.46	0.57	5.25
DTE	0.17	0.38	0.49	**0.73**	1.06	1.66	0.57	3.68
EOA	0.20	0.39	0.54	**0.83**	1.13	1.46	0.59	2.33
FME	0.20	0.27	0.36	**0.57**	0.73	0.96	0.38	1.43
FRE	0.12	0.35	0.45	**0.58**	1.04	1.48	0.59	3.89
HEI	0.17	0.39	0.52	**0.67**	0.98	1.32	0.46	2.32
HEN	0.33	0.66	0.75	**0.91**	1.18	1.50	0.42	2.76
IFX	0.18	0.28	0.40	**0.68**	1.15	1.55	0.75	2.70
LHA	0.20	0.39	0.53	**0.77**	0.96	1.26	0.43	2.98
LIN	0.24	0.49	0.62	**0.82**	1.09	1.39	0.47	6.39
MRK	0.33	0.47	0.54	**0.75**	0.97	1.48	0.43	3.86
MUV	0.08	0.23	0.30	**0.49**	0.75	1.26	0.44	2.25
PSM	0.19	0.51	0.78	**1.03**	1.48	2.43	0.70	3.83
RWE	0.09	0.23	0.34	**0.47**	0.87	1.45	0.54	3.39
SAP	0.16	0.28	0.41	**0.65**	1.07	1.27	0.66	2.40
SIE	0.39	0.56	0.69	**0.81**	1.06	1.22	0.37	2.31
TKA	0.15	0.32	0.47	**0.63**	0.97	1.93	0.50	3.50
VNA	0.07	0.34	0.50	**0.66**	0.96	1.35	0.46	2.64
VOW	0.11	0.22	0.32	**0.51**	0.78	1.29	0.46	2.36

Table 2.B.6: L^2**-Errors Relative to the Mean for** $\tau = 0.5$**.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the mean (constant level / constant shape competitor) as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.18	0.42	0.63	**0.79**	1.12	1.69	0.49	8.77
ALV	0.07	0.30	0.53	**0.92**	1.32	1.84	0.79	2.53
BAS	0.19	0.28	0.46	**0.67**	1.11	2.06	0.64	9.46
BAY	0.13	0.57	0.69	**1.03**	1.53	1.83	0.83	3.59
BEI	0.25	0.51	0.81	**1.01**	1.44	2.44	0.63	4.35
BMW	0.16	0.37	0.52	**0.88**	1.25	1.59	0.73	4.39
CBK	0.13	0.43	0.66	**0.92**	1.24	1.52	0.58	3.76
CON	0.11	0.38	0.64	**0.88**	1.07	1.40	0.44	3.29
DAI	0.15	0.39	0.59	**0.87**	1.35	2.26	0.76	5.29
DBK	0.16	0.44	0.67	**0.91**	1.34	1.97	0.67	4.33
DPW	0.09	0.44	0.60	**0.94**	1.23	1.83	0.63	3.40
DTE	0.19	0.51	0.64	**0.84**	1.11	1.60	0.46	4.40
EOA	0.15	0.35	0.53	**0.74**	1.03	1.50	0.51	2.98
FME	0.29	0.44	0.64	**0.89**	1.29	1.48	0.65	3.42
FRE	0.15	0.50	0.67	**0.96**	1.46	1.70	0.79	3.52
HEI	0.09	0.38	0.57	**0.78**	1.23	1.65	0.66	3.58
HEN	0.26	0.73	0.84	**1.03**	1.19	1.51	0.35	2.60
IFX	0.13	0.43	0.60	**0.83**	1.27	1.86	0.67	4.61
LHA	0.22	0.39	0.63	**0.98**	1.67	2.30	1.04	3.34
LIN	0.23	0.48	0.67	**0.89**	1.13	1.31	0.46	4.31
MRK	0.27	0.53	0.64	**0.84**	1.11	1.77	0.47	4.04
MUV	0.18	0.36	0.54	**0.78**	1.04	1.34	0.50	2.37
PSM	0.19	0.43	0.67	**1.08**	1.37	1.86	0.70	6.11
RWE	0.08	0.36	0.51	**0.98**	1.34	1.77	0.83	3.59
SAP	0.16	0.31	0.56	**0.83**	1.13	1.42	0.58	2.39
SIE	0.14	0.36	0.65	**0.86**	1.16	1.48	0.51	2.68
TKA	0.11	0.51	0.68	**0.86**	1.18	1.73	0.50	3.96
VNA	0.18	0.48	0.73	**0.89**	1.24	1.58	0.51	3.74
VOW	0.16	0.38	0.55	**0.94**	1.13	1.66	0.58	3.27

Table 2.B.7: L^2**-Errors Relative to Adapted RiskMetrics Competitor for** $\tau = 0.6$**.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the* `RiskMetrics`*-type competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.24	0.47	0.61	**0.89**	1.12	1.69	0.50	6.54
ALV	0.10	0.26	0.32	**0.50**	0.92	1.98	0.60	3.23
BAS	0.11	0.28	0.35	**0.55**	1.10	1.44	0.75	2.49
BAY	0.36	0.70	0.83	**0.99**	1.12	1.39	0.29	2.71
BEI	0.27	0.43	0.64	**0.90**	1.10	1.56	0.47	4.31
BMW	0.16	0.28	0.38	**0.59**	0.90	1.24	0.52	5.24
CBK	0.15	0.43	0.59	**0.95**	1.20	1.52	0.61	3.43
CON	0.08	0.26	0.43	**0.69**	1.00	1.43	0.57	5.06
DAI	0.14	0.32	0.51	**0.78**	1.33	2.02	0.82	5.87
DBK	0.10	0.27	0.47	**0.59**	0.96	1.82	0.49	3.46
DPW	0.12	0.28	0.50	**0.84**	1.08	1.59	0.58	3.86
DTE	0.23	0.50	0.70	**0.95**	1.46	2.00	0.76	9.30
EOA	0.19	0.46	0.59	**0.84**	1.09	1.45	0.49	3.59
FME	0.31	0.37	0.48	**0.77**	1.02	1.30	0.54	2.68
FRE	0.29	0.38	0.53	**0.71**	0.99	1.90	0.45	2.70
HEI	0.12	0.37	0.54	**0.78**	1.26	1.55	0.72	2.11
HEN	0.25	0.63	0.84	**1.01**	1.14	1.71	0.30	2.48
IFX	0.14	0.42	0.55	**0.88**	1.20	2.10	0.65	3.24
LHA	0.29	0.42	0.55	**0.83**	1.03	1.52	0.47	3.04
LIN	0.30	0.63	0.77	**1.00**	1.19	1.70	0.43	8.05
MRK	0.45	0.55	0.66	**0.79**	1.07	1.92	0.41	3.83
MUV	0.09	0.28	0.46	**0.66**	1.04	1.48	0.58	5.34
PSM	0.23	0.55	0.80	**1.07**	1.30	2.01	0.50	6.12
RWE	0.14	0.36	0.47	**0.67**	0.92	1.33	0.46	3.38
SAP	0.22	0.40	0.46	**0.72**	0.98	1.48	0.52	3.45
SIE	0.26	0.50	0.75	**0.92**	1.01	1.29	0.26	5.33
TKA	0.21	0.38	0.49	**0.81**	1.27	1.78	0.78	3.46
VNA	0.08	0.43	0.62	**0.76**	1.15	2.02	0.53	2.91
VOW	0.21	0.33	0.42	**0.65**	1.12	1.73	0.70	4.38

Table 2.B.8: **L^2-Errors Relative to Random Walk Competitor for $\tau = 0.6$.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from random walk / constant shape competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.16	0.35	0.48	**0.83**	1.18	1.86	0.70	5.66
ALV	0.05	0.16	0.20	**0.31**	0.62	1.64	0.41	5.18
BAS	0.08	0.17	0.23	**0.42**	1.02	1.75	0.79	5.69
BAY	0.35	0.55	0.71	**0.94**	1.14	1.46	0.43	5.16
BEI	0.17	0.33	0.46	**0.70**	1.11	1.94	0.65	3.79
BMW	0.08	0.17	0.24	**0.44**	0.61	1.04	0.37	2.11
CBK	0.14	0.34	0.55	**0.75**	0.97	1.35	0.42	2.71
CON	0.06	0.18	0.30	**0.46**	0.68	1.32	0.38	2.08
DAI	0.06	0.21	0.35	**0.56**	0.95	2.13	0.60	3.83
DBK	0.08	0.18	0.28	**0.43**	0.58	1.03	0.30	3.27
DPW	0.11	0.28	0.35	**0.71**	0.92	1.37	0.57	8.28
DTE	0.15	0.37	0.46	**0.71**	1.22	1.97	0.76	5.83
EOA	0.15	0.29	0.46	**0.69**	1.01	1.55	0.56	2.53
FME	0.22	0.26	0.35	**0.56**	0.66	1.17	0.31	2.68
FRE	0.19	0.27	0.43	**0.53**	1.00	1.50	0.57	3.38
HEI	0.08	0.28	0.41	**0.66**	1.06	1.44	0.65	3.15
HEN	0.17	0.57	0.73	**0.89**	1.30	1.68	0.58	3.33
IFX	0.12	0.19	0.41	**0.66**	1.15	1.59	0.74	2.72
LHA	0.19	0.30	0.48	**0.74**	0.91	1.24	0.43	7.55
LIN	0.28	0.42	0.55	**0.78**	1.20	1.81	0.65	6.09
MRK	0.36	0.42	0.49	**0.71**	1.01	1.96	0.51	6.45
MUV	0.06	0.19	0.28	**0.49**	0.72	1.63	0.43	2.73
PSM	0.16	0.47	0.75	**0.99**	1.31	2.28	0.57	3.49
RWE	0.12	0.26	0.33	**0.44**	0.83	1.61	0.50	4.65
SAP	0.14	0.25	0.30	**0.69**	1.05	1.46	0.75	4.71
SIE	0.22	0.36	0.62	**0.79**	1.01	1.30	0.39	2.17
TKA	0.16	0.26	0.42	**0.60**	1.12	2.37	0.70	3.63
VNA	0.06	0.31	0.46	**0.63**	0.93	1.64	0.47	2.92
VOW	0.11	0.21	0.29	**0.46**	0.86	1.22	0.57	2.91

Table 2.B.9: **L^2-Errors Relative to Mean for $\tau = 0.6$.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the mean (constant level / constant shape competitor) as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.18	0.40	0.53	**0.81**	1.16	1.81	0.63	8.40
ALV	0.10	0.23	0.59	**0.84**	1.44	1.68	0.85	4.40
BAS	0.15	0.29	0.43	**0.76**	1.13	2.40	0.70	10.32
BAY	0.18	0.37	0.61	**0.92**	1.32	1.59	0.71	4.37
BEI	0.13	0.37	0.68	**0.95**	1.63	2.71	0.95	7.12
BMW	0.06	0.18	0.45	**0.86**	1.09	1.60	0.64	16.28
CBK	0.06	0.37	0.59	**0.94**	1.46	2.54	0.86	5.27
CON	0.08	0.35	0.49	**0.76**	1.02	1.65	0.53	6.28
DAI	0.05	0.38	0.55	**0.82**	1.39	2.68	0.84	6.09
DBK	0.12	0.33	0.53	**0.85**	1.43	2.50	0.90	6.40
DPW	0.14	0.27	0.50	**0.88**	1.37	2.18	0.87	5.63
DTE	0.20	0.37	0.62	**0.89**	1.36	1.96	0.74	3.30
EOA	0.14	0.22	0.45	**0.74**	1.04	1.50	0.59	3.78
FME	0.12	0.34	0.55	**0.90**	1.12	1.46	0.58	3.29
FRE	0.15	0.48	0.64	**1.01**	1.51	1.94	0.86	3.99
HEI	0.14	0.36	0.55	**0.76**	1.15	1.61	0.60	4.73
HEN	0.29	0.72	0.85	**1.00**	1.35	1.89	0.50	3.00
IFX	0.10	0.33	0.53	**0.90**	1.37	1.73	0.84	5.78
LHA	0.12	0.28	0.64	**0.96**	1.53	1.79	0.90	7.28
LIN	0.10	0.38	0.54	**0.79**	1.12	1.40	0.58	4.18
MRK	0.20	0.39	0.63	**0.81**	1.23	1.85	0.60	4.07
MUV	0.12	0.30	0.48	**0.75**	1.01	1.36	0.53	3.71
PSM	0.22	0.44	0.69	**0.85**	1.28	1.81	0.60	3.28
RWE	0.05	0.24	0.39	**0.72**	1.21	1.55	0.82	5.29
SAP	0.10	0.28	0.49	**0.91**	1.36	1.99	0.86	2.55
SIE	0.18	0.41	0.57	**0.80**	1.15	1.66	0.58	6.11
TKA	0.14	0.33	0.60	**0.86**	1.38	2.19	0.79	5.90
VNA	0.20	0.45	0.63	**0.90**	1.11	1.61	0.48	5.98
VOW	0.13	0.35	0.51	**0.85**	1.15	1.75	0.64	3.01

Table 2.B.10: L^2**-Errors Relative to Adapted RiskMetrics Competitor for** $\tau = 0.7$**.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the* `RiskMetrics`*-type competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.17	0.38	0.51	**0.92**	1.36	1.95	0.85	6.51
ALV	0.12	0.16	0.23	**0.48**	0.86	2.21	0.63	5.10
BAS	0.07	0.19	0.30	**0.63**	1.20	1.86	0.90	8.60
BAY	0.10	0.53	0.65	**0.98**	1.20	1.36	0.55	3.92
BEI	0.13	0.23	0.53	**0.91**	1.41	2.64	0.87	6.62
BMW	0.06	0.16	0.26	**0.45**	0.73	1.49	0.47	5.75
CBK	0.06	0.33	0.55	**0.90**	1.40	2.21	0.85	14.16
CON	0.08	0.18	0.33	**0.58**	0.87	1.53	0.54	11.47
DAI	0.06	0.20	0.35	**0.68**	1.78	2.95	1.42	5.30
DBK	0.06	0.17	0.28	**0.51**	1.28	2.16	1.00	5.27
DPW	0.18	0.21	0.42	**0.69**	1.15	1.96	0.73	4.78
DTE	0.19	0.34	0.50	**0.95**	1.81	3.15	1.30	11.60
EOA	0.06	0.28	0.50	**0.83**	1.10	1.48	0.60	3.33
FME	0.11	0.27	0.41	**0.66**	1.10	2.63	0.69	3.09
FRE	0.16	0.34	0.49	**0.71**	1.15	2.05	0.66	3.62
HEI	0.18	0.28	0.47	**0.75**	1.19	1.67	0.72	3.77
HEN	0.22	0.61	0.78	**0.98**	1.54	1.92	0.76	2.68
IFX	0.12	0.30	0.50	**0.98**	1.39	2.08	0.88	6.45
LHA	0.17	0.29	0.44	**0.73**	0.97	1.50	0.52	2.24
LIN	0.16	0.41	0.69	**0.86**	1.16	1.83	0.47	6.63
MRK	0.24	0.41	0.58	**0.77**	1.55	2.55	0.97	6.62
MUV	0.09	0.17	0.37	**0.68**	1.21	1.88	0.84	11.43
PSM	0.17	0.54	0.76	**1.00**	1.47	2.26	0.72	3.42
RWE	0.17	0.25	0.33	**0.58**	0.89	1.49	0.56	4.29
SAP	0.13	0.20	0.43	**0.79**	1.29	1.94	0.86	4.10
SIE	0.15	0.43	0.65	**0.89**	1.11	2.18	0.47	7.23
TKA	0.22	0.30	0.44	**0.66**	1.61	2.81	1.17	11.55
VNA	0.10	0.31	0.52	**0.70**	1.15	2.11	0.63	6.20
VOW	0.08	0.22	0.33	**0.63**	1.04	1.93	0.71	4.34

Table 2.B.11: **L^2-Errors Relative to Random Walk Competitor for $\tau = 0.7$.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from random walk / constant shape competitor as in Section 2.4.*

Stock	Min	q(0.1)	q(0.25)	**q(0.5)**	q(0.75)	q(0.9)	IQR	Max
ADS	0.13	0.28	0.40	**0.78**	1.45	1.97	1.05	6.09
ALV	0.08	0.12	0.17	**0.31**	0.74	1.76	0.57	6.38
BAS	0.05	0.14	0.24	**0.38**	1.13	2.07	0.89	7.34
BAY	0.09	0.34	0.59	**0.92**	1.20	1.65	0.61	7.66
BEI	0.12	0.18	0.40	**0.70**	1.40	2.61	0.99	8.04
BMW	0.04	0.11	0.19	**0.33**	0.53	1.38	0.34	2.89
CBK	0.04	0.31	0.47	**0.72**	1.35	1.98	0.89	10.94
CON	0.06	0.13	0.27	**0.42**	0.71	1.36	0.44	3.76
DAI	0.04	0.16	0.27	**0.51**	1.21	2.76	0.94	5.65
DBK	0.05	0.13	0.19	**0.39**	0.73	1.35	0.54	7.92
DPW	0.12	0.18	0.29	**0.68**	1.21	1.73	0.92	7.46
DTE	0.16	0.26	0.39	**0.79**	1.45	2.53	1.06	5.48
EOA	0.05	0.18	0.41	**0.73**	1.02	1.40	0.62	6.24
FME	0.09	0.17	0.29	**0.46**	0.73	1.61	0.45	3.88
FRE	0.12	0.25	0.39	**0.64**	1.19	1.50	0.80	5.60
HEI	0.16	0.22	0.42	**0.60**	0.98	1.41	0.56	4.11
HEN	0.16	0.53	0.66	**0.92**	1.39	1.99	0.73	3.89
IFX	0.07	0.22	0.36	**0.74**	1.25	1.71	0.89	3.70
LHA	0.10	0.20	0.36	**0.69**	0.90	1.32	0.55	6.81
LIN	0.12	0.32	0.50	**0.71**	1.19	2.05	0.69	6.61
MRK	0.19	0.33	0.45	**0.64**	1.15	2.77	0.70	6.90
MUV	0.08	0.12	0.23	**0.47**	0.86	2.52	0.63	3.71
PSM	0.20	0.44	0.67	**0.96**	1.38	2.12	0.71	3.22
RWE	0.10	0.19	0.25	**0.39**	0.73	1.63	0.49	6.18
SAP	0.09	0.15	0.29	**0.72**	1.30	2.23	1.00	5.11
SIE	0.13	0.35	0.54	**0.77**	1.06	2.07	0.53	4.72
TKA	0.14	0.23	0.36	**0.51**	1.55	2.73	1.18	10.72
VNA	0.07	0.23	0.40	**0.62**	0.94	1.59	0.54	5.05
VOW	0.06	0.17	0.22	**0.45**	0.97	1.37	0.75	2.91

Table 2.B.12: L^2**-Errors Relative to Mean for** $\tau = 0.7$**.** *Table reports minimum,r-quantiles ($q(r)$), interquartile range and maximum of the L^2 prediction error relative to the error L^2 prediction error from the mean (constant level / constant shape competitor) as in Section 2.4.*

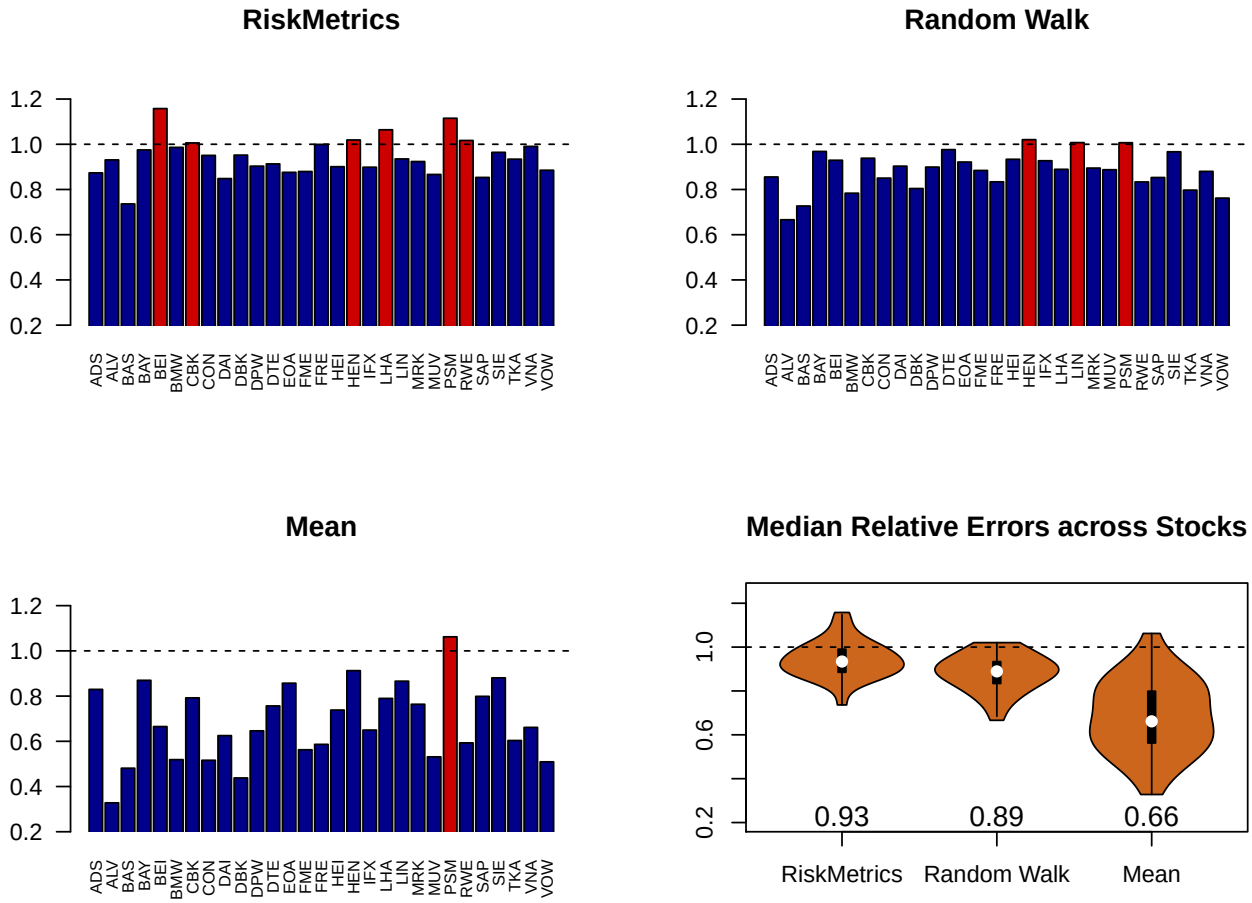

Figure 2.B.1: **Median Relative Performances for** $\tau = 0.4$**.** *Figure shows median relative errors for different competitors. The x-axis in the barplots is constituted by the different assets (see Table 2.A.1 in the appendix). Blue indicates bars below* 1*, red bars are above* 1*. The violin plots in the bottom-right panel depict the distribution of the medians in the barplots over assets. Numbers below the violins are means of medians.*

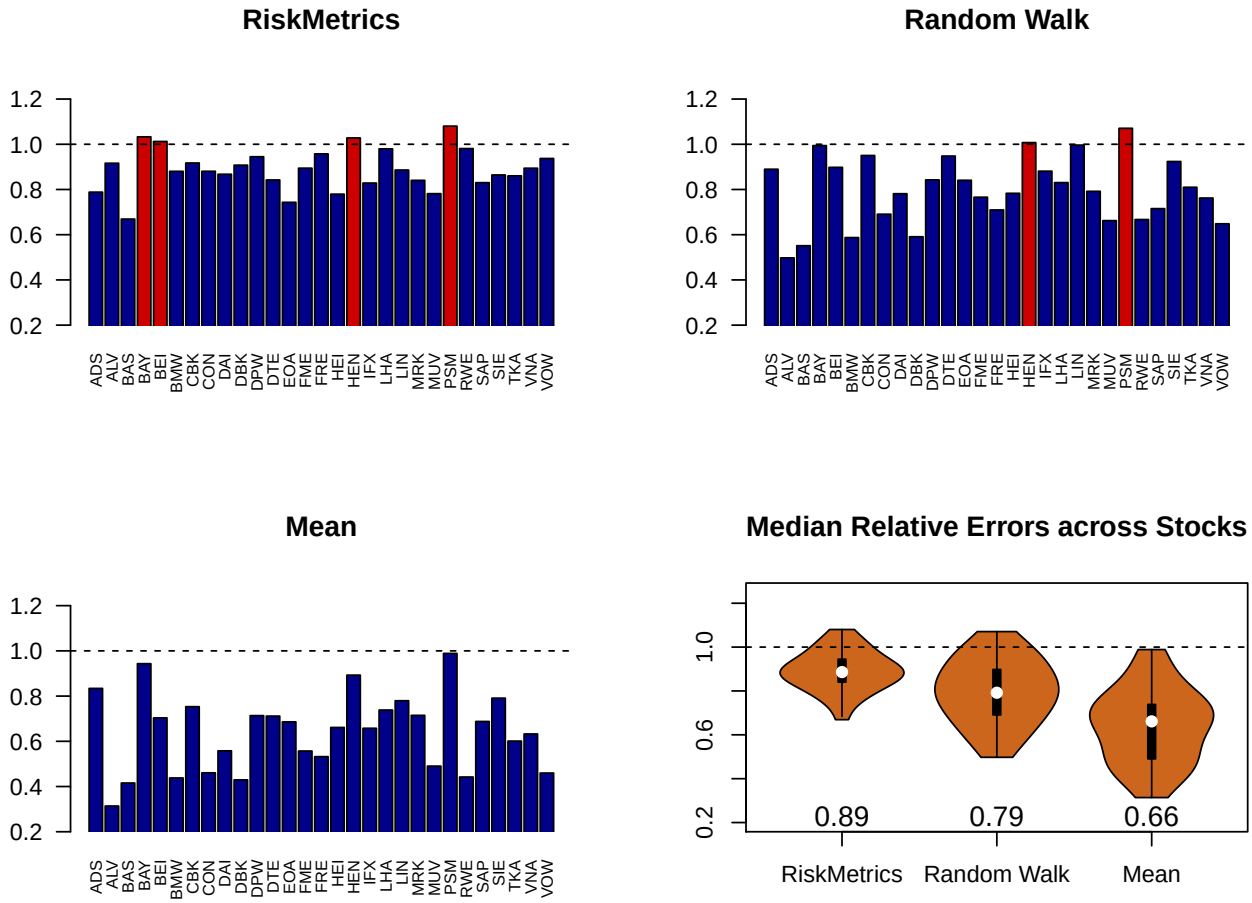

Figure 2.B.2: **Median Relative Performances for** $\tau = 0.6$**.** *The panels are to be interpreted as in Figure 2.B.1.*

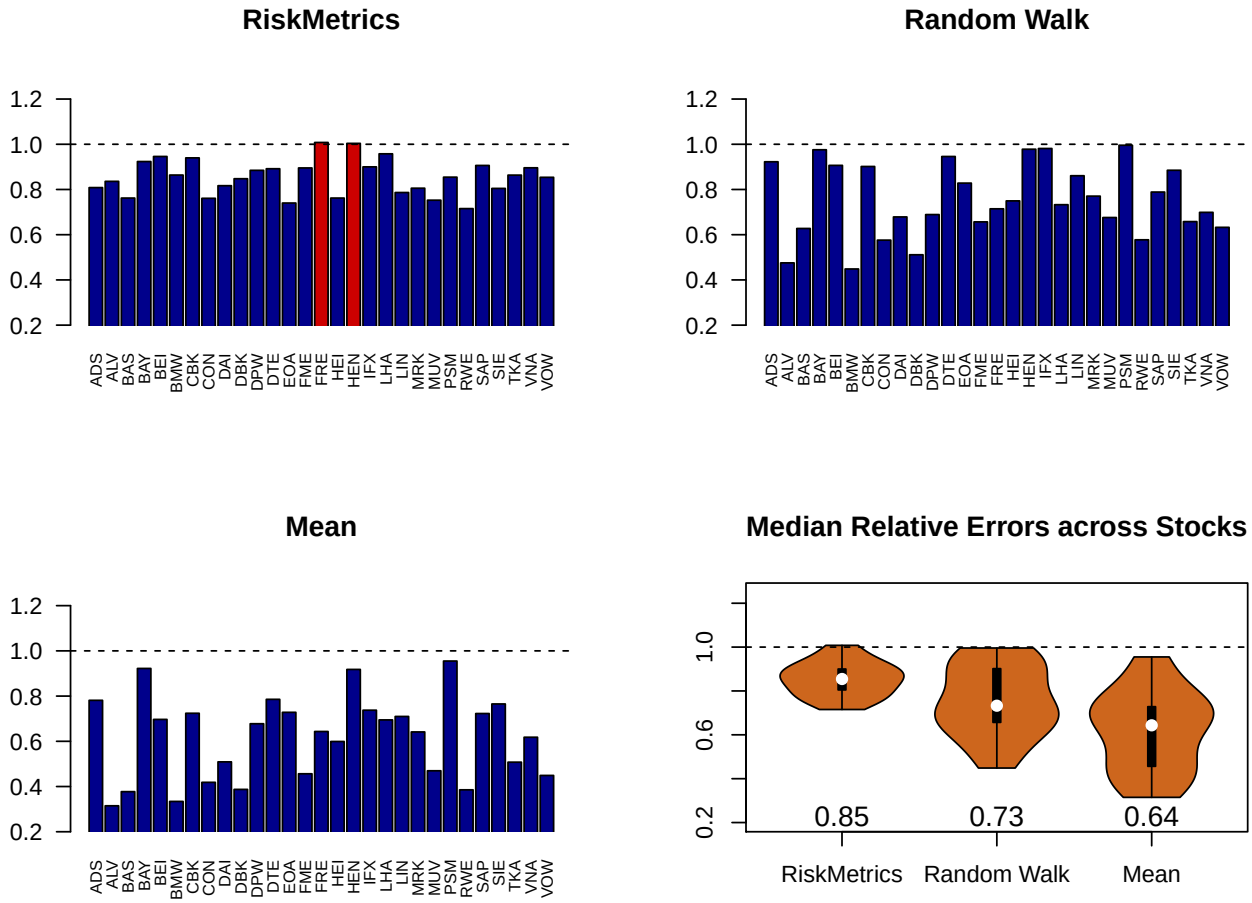

Figure 2.B.3: **Median Relative Performances for** $\tau = 0.7$**.** *The panels are to be interpreted as in Figure 2.B.1.*

Chapter 3

Heterogeneous Liquidity Effects in Corporate Bond Spreads

3.1 Introduction

Yield spreads of corporate bonds reflect multiple sources of risk. While credit risk is typically considered to be the most dominant driver, recent work points at the fundamental importance of liquidity risk.[1] The precise individual contributions of risk factors to the yield spread is of particular interest for investors, bond issuers as well regulatory authorities and central banks. Such information allows to improve risk management for bond portfolios, the timing and design of new bond issues as well as the regulatory assessment of asset riskiness and implementation of monetary policy.

According to Engle and Lange (2001), liquidity of an asset is *the ability to transact at low cost*. In their seminal work, Amihud and Mendelson (1986) introduced the idea that investors demand a premium for holding an illiquid asset. For bonds, this premium is to be viewed as one constituent of the yield spread. Liquidity premiums are typically difficult to extract from data as liquidity is latent. Furthermore, the way how a premium enters the yield spread is unknown and is likely to differ across assets.

The preceding remarks ultimately motivate the research questions addressed in this chapter: (i) in which way does idiosyncratic liquidity risk affect bond spreads, (ii) how do these effects vary over time, and (iii) what governs the differences of liquidity risk across assets? Using two different proxies for idiosyncratic bond liquidity, we find that their contribution to the yield spread is in many cases highly nonlinear, when controlling for market liquidity. This nonlinearity is particularly pronounced in times of low liquidity as it was the case, for example, during the 2008 financial

[1] See for example Longstaff et al. (2005).

crisis. However, the magnitude and range of the premium varies considerably among classes of bonds: in our data we identify one large class with moderate effects and one smaller class with large effects. These classes are among other characteristics distinguished by the spread level, the financial strength of the issuer as well as the size of the bonds. Interestingly, dynamic features of liquidity effects in the yield spreads seem to be quite similar for the majority of groups.

A number of recent studies has addressed the link between corporate bond liquidity and its contribution to yield spreads, e.g. Friewald et al. (2012), Dick-Nielsen (2012), Chen et al. (2007) and Houweling et al. (2005). The vast majority of these findings are gathered from trade data, which typically are temporally aggregated in order to construct a balanced panel, from which sub-samples are taken of bonds with similar time to maturity and rating class. The coefficients of the liquidity proxies in linear panel data regressions are then estimated for each sub-sample separately. Dick-Nielsen (2012) estimates linear panel regressions for a priori chosen subsamples employing a linear combination of liquidity measures as a new proxy. In particular, his results confirm the presence of liquidity effects on corporate bond spreads, varying over time, time to maturity, and rating classes. Based on similar data, Friewald et al. (2012) take a slightly different approach and estimate a linear panel model as well as Fama-MacBeth cross sectional regressions using a battery of classical liquidity measures. Effects of rating classes as well as changes in marginal effects of the liquidity measures in times of financial stress are covered using dummy variables. Their main finding confirms the presence of liquidity effects, which are particularly pronounced in times of financial turmoil.

While Friewald et al. (2012) and Dick-Nielsen (2012) aggregate all data to weekly, monthly and quarterly observations, Chen et al. (2007) employ even yearly data in a very similar framework. All studies basically find positive relations between measures of illiquidity and corporate bond spreads, while the magnitude of identified effects varies among different levels of temporal aggregation.

Contributing to this strand of literature, the three novelties of our work can be summarized as follows.

1. Our framework is capable to deal with nonlinear effects of idiosyncratic liquidity risk on yield spreads in excess to the effect of market liquidity. This allows to examine the interplay between different facets of bond illiquidity as addressed by different measures. For example, a deterioration of liquidity indicated by one measure might affect the yield spread differently in situations of high or low liquidity as indicated by another measure. Resulting mixed effects are a source of potential nonlinearity. Further, nonlinearities may also arise in a feasible regression even if the true liquidity-spread relation is linear, depending on the relation between the true but unobserved liquidity variable and its approximation.

2. We estimate liquidity effects based on *event time models*, thus avoiding high levels of temporal aggregation and loss of information. Despite the high degree of non-synchronous trading, common components of the regression functions can still be estimated using time series *and* cross sectional variation.

3. Our approach allows for substantial heterogeneity of liquidity effects across bonds. We avoid subjective divisions into subsamples by endowing the model with a latent group structure, which can be estimated from the data using standard cluster algorithms. In particular, there is no classification based on time-varying quantities such as time to maturity or the rating, avoiding inconsistencies when classifying bonds based on data over a large time horizon.

The remainder of this chapter is structured as follows. Section 3.2 introduces the formal model, presents estimation procedures, and addresses inferential questions and corresponding statistical tests. In Section 3.3 liquidity measures and their adaptation to the model framework are discussed. Section 3.4 offers an application of the outlined procedure to trade data from the US corporate bond market between 2004 and 2012. Finally, Section 3.5 concludes.

3.2 An Econometric Assessment of Liquidity Risk

In the following we present our semiparametric panel data model with a latent group structure in event time. For a formal presentation of the framework, index the available bonds by $i \in \{1, ..., N\} =: \mathbf{N}$. Suppose that for each bond there are observations over time available on a bond-specific time-scale. Precisely, for an $i \in \mathbf{N}$ denote the ordered set of corresponding time stamps as $\mathcal{T}_i$, with cardinality $T_i := |\mathcal{T}_i|$. These sets might well vary over the index i as is most likely the case with event times. For simplicity of notation we use a generic time index $t \in \mathcal{T}_i$ for a given $i \in \mathbf{N}$.

Given these time scales, suppose that for each bond $i \in \mathbf{N}$ a sample of time series observations $\{(y_{it}, \mathbf{x}_{it}, z_{it}) : \ t \in \mathcal{T}_i\}$ is available, where y_{it} denotes the spread, z_{it} is an L-vector of control variables, and the vector $\mathbf{x}_{it} \in \mathcal{I} \subset \mathbb{R}^m$ collects m idiosyncratic liquidity measures, with values in some set $\mathcal{I}$, independent of the index i.

3.2.1 Model Formulation and Estimation

We suggest incorporating nonlinear effects of idiosyncratic liquidity in a regression model according to

$$y_{it} = M_i(\mathbf{x}_{it}) + \gamma_i^\top z_{it} + \epsilon_{it}, \tag{3.1}$$

where $M_i : \mathcal{I} \to \mathbb{R}$ denotes an unknown smooth function. The term $M_i(\mathbf{x}_{it})$ represents the effect of idiosyncratic liquidity $\mathbf{x}_{it}$ on the yield spread, while the term $\gamma_i^\top z_{it}$ describes the effects of control variables z_{it}. These controls incorporate information on bond characteristics and the market environment, including market liquidity. Their bond-specific marginal effects are collected in the parameter vector γ_i. Regarding an intercept in model 3.1, we note that constants in $M_i(\mathbf{x}_{it})$ and $\gamma_i^\top z_{it}$ are certainly not separately identifiable.

For the above model we rely on the following set of standard assumptions. Importantly, we suppose that for all $i \in \mathbf{N}$, the random variables $\{(\epsilon_{it}, \mathbf{x}_{it}, z_{it}) : \ t \in \mathcal{T}_i\}$ are strictly stationary and ergodic. Further, the covariates are assumed to be exogenous in the sense that for any $i, j \in \mathbf{N}$, $\epsilon_{it} \perp\!\!\!\perp (\mathbf{x}_{js}, z_{js})$ for any $t \in \mathcal{T}_i, \ s \in \mathcal{T}_j$. We assume the errors to be independent across time and bonds, i.e. $\epsilon_{it} \perp\!\!\!\perp \epsilon_{js}$, for any $i, j \in \mathbf{N}$, $t \in \mathcal{T}_i$, $s \in \mathcal{T}_j$. Also, we require for any bond $i \in \mathbf{N}$ the errors $\{\epsilon_{it} : \ t \in \mathcal{T}_i\}$ to be normally distributed with mean 0 and bond specific variance σ_i^2. Suppose further that for both $\mathbf{x}_{it}$ and z_{it}, moments of sufficiently high order exist and that they are not perfectly correlated.

As a central feature of model (3.1) we postulate a latent group structure determining heterogeneity of liquidity effects. As in Vogt and Linton (2017) and Su et al. (2016), we assume that cross sectional units are organized in K groups, denoted $G_1, ..., G_K$, which form an unknown partition of the index set $\mathbf{N}$. Within a group, the nonparametric regression function is identical among assets, i.e. $\forall k = 1, ..., K$ it holds that

$$M_i = g_k \ \ \forall i \in G_k. \tag{3.2}$$

The functions $g_k : \mathcal{I} \to \mathbb{R}$ are the group specific liquidity effects - the quantity of main economic interest. Note that the parameters in γ_i might vary across bonds, also within a group. Hence, the *only* (a priori) common feature of two assets $i, j \in G_k$ is the liquidity effect g_k. Across groups we assume these functions to be sufficiently different in the L^2-metric.

Model (3.1) can be interpreted not only as a generalization of the classical linear panel regression model, but also of the longitudinal data models presented in Vogt and Linton (2017) and Su et al. (2016). It should be noted that the fully parametric linear models estimated in Friewald et al. (2012) and Dick-Nielsen (2012) are comparable to (3.1) in the sense that they also possess

a group structure, which is, however, determined by the rating class and time to maturity and is hence assumed to be known a priori.

Another important difference between the aforementioned models and (3.1) is the time dimension. As indicated, the time scales $\mathcal{T}_i$ collect bond-specific event times and are allowed to differ across assets. The precise meaning of an event, which depends on the trading activity of a bond, is introduced in Section 3.3. Reading (3.1) as a conventional panel data model, standard estimation techniques would suffer from unbalanced observations in practice, because trading is typically highly non-synchronous even at a daily level.

Our estimation strategy adapts the idea of Vogt and Linton (2017) to our setup. The procedure can be structured into three stages. The first stage is concerned with the estimation of the unknown parameters in $E[y_{it}|x_{it}, z_{it}]$ for each $i \in \mathbf{N}$ from time series variation. At the second stage we construct groups $G_1, \ldots, G_K$ from the results of the first stage. At a third stage common regression curves g_k are finally (re-) estimated using information from the two previous stages.

Compared to conventional panel data techniques, our strategy allows in particular to employ variation over the unbalanced time scales as well as variation over assets to estimate the common, i.e. group specific, component of the regression functions in (3.1). The implementation proceeds as follows.

First Stage

At the first stage, event time model (3.1) is estimated for each bond $i \in \mathbf{N}$ separately from corresponding time series data $\{(y_{it}, x_{it}, z_{it}) : t \in \mathcal{T}_i\}$. This is implemented running N times a penalized maximum likelihood approach. The smooth component of the regression is modeled using splines due to their numerical convenience and flexibility. We decide to approximate the functions M_i by thinplate regression splines (see Section 4.1.5 in Wood, 2006) of bond-specific dimension. The penalty order of the spline approximation for M_i is set to two for all bonds, ensuring that all functions M_i are estimated under the same notion of smoothness. Regarding the choice of the spline basis for M_i, note that multiplicative mixed effects between the m measures of illiquidity are allowed explicitly.

Given the spline approximation of the function M_i, our model in (3.1) can be restated as a multiple regression model under an identification constraint for the smooth terms.[2] Corresponding coefficients of the truncated basis expansion of M_i are collected in the vector λ_i.

The roughness penalty to be employed in the objective function is the well known scaled quadratic form $\nu_i \lambda_i^\top S_i \lambda_i$. Here S_i is a square matrix constituted by the second derivatives of elements in the corresponding spline basis[3] and $\nu_i > 0$ is a tuning parameter. The latter is determined using

[2] In particular, as in Wood (2006) pp. 163-164 we impose the constraints $\sum_{t \in \mathcal{T}_i} M_i(\mathbf{x}_{it}) = \sum_{t \in \mathcal{T}_i} b_i(t) = 0$.

[3] See Wood (2003) or the aforementioned section in Wood (2006) for details.

generalized cross validation. For efficiency of notation we collect all but the tuning parameter in the vector $\theta_i := [\lambda_i^\top, \gamma_i^\top, \sigma_i^2]^\top$ and denote the unpenalized Gaussian log-likelihood as $\mathcal{L}_i(\theta_i)$. The final objective function is the penalized likelihood given by

$$\mathcal{L}_{i,p}(\theta_i) = \mathcal{L}_i(\theta_i) - 1/2\nu_i \lambda_i^\top S_i \lambda_i. \tag{3.3}$$

The estimator $\hat{\theta}_i$ maximizes (3.3) respecting the imposed identification constraints. Then, the estimator $\hat{M}_i(x)$, $x \in \mathcal{I}$, is simply a linear combination of the basis functions evaluated at x. The weights in this linear combination are the coordinates in $\hat{\theta}_i$ estimating λ_i. Denoting the T_i-vector of residuals $\hat{\epsilon}_i$ and the influence matrix P_i, the estimator of the error variance is obtained as $\hat{\sigma}_i^2 = \frac{\hat{\epsilon}_i^\top \hat{\epsilon}_i}{T_i - tr(P_i)}$.

Second Stage

Along the lines of Vogt and Linton (2017), we suggest to employ the estimates $\hat{M}_i$ from the previous stage to classify bonds into groups G_k. For this purpose, a hierarchical cluster algorithm appears particularly suited as it (i) offers a natural notion of a distance between groups, and (ii) is computationally efficient even for a large number of bonds. For any cluster algorithm we need a suitable metric to measure the distance between two regression functions. Since these are smooth, the L^2 metric is an appropriate choice, which for two unknown population functions M_i and M_j is defined by

$$\Delta_{i,j} := \left(\int_{\mathcal{I}} (M_i(\mathbf{x}) - M_j(\mathbf{x}))^2 \mathrm{d}\mathbf{x} \right)^{1/2}. \tag{3.4}$$

In the same way, we define $\hat{\Delta}_{i,j}$ to be the L^2 distance between $\hat{M}_i$ and $\hat{M}_j$. Intuitively, given $i, j \in G_k$ these functions should be close to each other and so should be $\hat{\Delta}_{i,j}$ and $\Delta_{i,j} = 0$. Based on this metric, a hierarchical cluster algorithm yields a set of clusters, for a given number of K groups, including distances between clusters.

The statistics literature offers a number of cost criteria allowing to select the number of groups K from distances (see e.g. Charrad et al., 2014 for an overview). We suggest to employ one such cost criterion in order to obtain an estimate $\hat{K}$. Given this estimate, ultimate outcomes of the stage are estimated groups $\hat{G}_1, ..., \hat{G}_{\hat{K}}$.

Regarding implementation we note that, depending on the number of bonds N and the number of liquidity measures m, this stage might be computationally demanding, because it involves $O(N^2)$ computations of m dimensional integrals.

Third Stage

In a third step the group specific liquidity effects g_k are estimated. Following Vogt and Linton

(2017) this is implemented by averaging over estimated curves within groups. Unweighted averaging however might lead to inefficient estimation of g_k, because error variances σ_i^2 are allowed to vary over bonds i, also within groups G_k. We hence suggest to obtain an estimate $\hat{g}_k$ of the group function g_k at $\mathbf{x} \in \mathcal{I}$ by weighted averaging according to

$$\hat{g}_k(\mathbf{x}) := \sum_{i \in \hat{G}_k} w_{ik} \hat{M}_i(\mathbf{x})$$

where weights are defined as

$$w_{ik} := \frac{\hat{\sigma}_i^{-1}}{\sum_{j \in \hat{G}_k} \hat{\sigma}_j^{-1}}.$$

This weighting scheme ideally leads to an efficient estimator of $\hat{g}_k$, given the assumption of independent errors across i. Note that controlling for potential correlation is infeasible in our case as it would require temporally balanced data.

3.2.2 Inference

A large fraction of existing studies find significant linear liquidity effects from temporally aggregated data, but the results are specific to the level of aggregation. In our event time study, the type of liquidity effect is *a priori* unclear, which motivates to test the form of the effect from the data. To do so, we suggest an approximate (conditional) two-step testing procedure. It is designed to conclude whether there is (i) no effect, (ii) a linear or (iii) a nonlinear effect for each bond separately after the first stage regressions. In particular, we test in a first step whether liquidity effects are significant at all. Subsequently it is tested for which bonds the liquidity effect, conditional on being significant, deviates from linearity. Note that the first-mentioned test also allows a *pre-classification*: bonds that do not show any significant liquidity effect will be classified in a separate group, say $\hat{G}_0$. All regression functions $\hat{M}_i$ with $i \in \hat{G}_0$ are not used in the second and third stages.

Approaching the aforementioned procedure more formally, we consider the following two testing problems, **P1** and **P2**:

P1

$$\mathrm{H}_0^{(1)} : \; M_i(x) = 0 \; \forall x \in \mathcal{I}$$

versus

$$\mathrm{H}_1^{(1)} : \; \exists u \subset \mathcal{I} : \; M_i(x) \neq 0 \text{ for } x \in u.$$

For the second testing problem, define the set of all linear functions defined over $\mathcal{I}$ as

$$\mathcal{M} := \left\{ f : \mathcal{I} \to \mathbb{R} : \ \exists \beta \in \mathbb{R}^m : \ f(x) = \beta^\top x \ \ \forall x \in \mathcal{I} \right\}.$$

Formally, we wish to test whether the function M_i is in this set $\mathcal{M}$ or not, i.e.

P2

$$\begin{aligned} &\mathrm{H}_0^{(2)} : \ M_i \in \mathcal{M} \\ &\qquad \text{versus} \\ &\mathrm{H}_1^{(2)} : \ M_i \notin \mathcal{M}. \end{aligned}$$

Hypothesis $H_0^{(1)}$ is tested separately for all $i \in \mathbf{N}$. Analogously, the testing problem **P2** is considered separately for different bonds $i \in \mathbf{N} \setminus \hat{G}_0$. For both hypotheses, corresponding F-tests are readily available, because the spline approximation of M_i nests a zero function as well as a linear model. Denoting the (unscaled) deviances[4] of the null and alternative models as D_0 and D_1, the test statistic obtains as

$$F = \frac{(D_0 - D_1)/(df_1 - df_0)}{D_1/(T_i - df_1)}. \tag{3.5}$$

The quantities df_0 and df_1 are the model's (effective) degrees of freedom, defined as the trace of the corresponding influence matrix. For calculating F, we follow the advice in Wood (2006) and use estimates obtained from the maximization of the unpenalized likelihood in case of both testing problems. Under the null hypothesis, the test statistic in (3.5) is approximately $F_{T_i - df_1}^{df_1 - df_0}$-distributed for large $|T_i|$, which we use to obtain corresponding critical values and p-values.

3.3 Corporate Bond Liquidity

In the previous section we described a modeling approach allowing to identify the effect of an m-vector of idiosyncratic liquidity proxies, $\mathbf{x}_{it}$, on the yield spread given a bond specific event time scale. In this section we make the proxies in $\mathbf{x}_{it}$, the time scale as well as our control variable for market liquidity precise.

[4]For the definition, see Wood (2006), p.69 in the case of GLMs and p. 191 in the case of GAMs.

3.3.1 Idiosyncratic Liquidity

The literature offers a battery of liquidity proxies, designed to cover a range of different facets of illiquidity. Most common proxies in empirical studies are directly or indirectly measured bid-ask spreads, measures of the trading activity and measures of the price impact of transactions.
In this work we focus on $m = 2$ measures of idiosyncratic corporate bond illiquidity. The first one is a variant of the well-established Amihud (2002)-measure for the price impact. The second one is a measure of costs from so-called *imputed round-trip trades* as suggested by Feldhütter (2011).
In order to introduce these measures more formally, let us first discuss the considered time scale in more detail, recalling that one of our aims is to estimate liquidity effects at low levels of temporal aggregation.
As both of the above liquidity measures are calculated over fixed time intervals, temporal aggregation cannot be avoided completely. We consider intervals of one day for this purpose, which is a plausible compromise between a low level of aggregation (yielding larger T_i's) and exclusion of very thinly traded bonds (yielding smaller N). Thus, we finally define as an *event* (in the time scale sense) a calendar day at which for the considered bond there are sufficiently many transactions available to calculate the liquidity measures. We refer to such days as *active trading days*.

The liquidity proxy of Amihud (2002) explores how a trade changes the price relative to a dollar of traded volume. His measure of the price impact of a trade obtains as the absolute return of a trade divided by the traded volume, and is typically averaged over a fixed time horizon such as a day. Large price impacts are considered to indicate low liquidity. In order to achieve greater numerical stability, we consider a variant of this measure obtained by replacing the volume by the log-volume. In more formal terms, we label the temporally ordered transactions of a bond i, $\tau = 1, 2, \ldots, S_{it}$ at some active trading day t. Further we denote the reported price and volume on transaction τ as $p_{it,\tau}$ and $vol_{it,\tau}$, whereas $p_{it,0}$ is the last available price before day t. The corresponding return on transaction τ obtains as $r_{it,\tau} := (p_{it,\tau} - p_{it,\tau-1})/p_{it,\tau-1}$. Our variant of the Amihud (2002)-measure of price impact obtains as

$$x_{1,it} := \frac{1}{S_{it}} \sum_{\tau=1}^{S_{it}} \frac{|r_{it,\tau}|}{\log(vol_{it,\tau})}.$$

The second measure we use addresses illiquidity by inferring transaction costs from imputed round-trips as introduced by Feldhütter (2011). A round-trip trade describes the cycle of buying (selling) and re-selling (re-buying) a given amount of an asset within a short period of time. Indeed, empirically oftentimes trades with similar volume occur within a small time interval, after no other transactions were recorded for a longer period of time. According to Feldhütter (2011)

these trades are likely to be transactions constituting a round-trip-trade. Dick-Nielsen (2012) suggest to measure *imputed round-trip costs* (IRC), i.e. the transaction costs within such a cycle as the scaled range of transaction prices. For our purposes their measure can be calculated at day t for bond i according to

$$x_{2,it} := \frac{1}{|IRT_{it}|} \sum_{r \in IRT_{it}} \frac{\max_{\tau \in IRT_{it,r}} p_{it,\tau} - \min_{\tau \in IRT_{it,r}} p_{it,\tau}}{\max_{\tau \in IRT_{it,r}} p_{it,\tau}},$$

where the set IRT_{it} collects sets of transactions at that day sharing the same volume. For example if the volumes reported on transactions τ_1, τ_2 and τ_3 at day t coincide, the triple (τ_1, τ_2, τ_3) constitutes an imputed round-trip, if there is no further transaction with the same volume. The r-th element in IRT_{it}, i.e. the r-th imputed round-trip, is denoted $IRT_{it,r}$. As is also the case with our first measure $x_{1,it}$, large values of $x_{2,it}$ indicate low liquidity.

Model (3.1) enables us to discover rich mixed effects between the two liquidity measures. Besides theoretical moment conditions, it was assumed that $\mathbf{x}_{it}$ takes values in $\mathcal{I}$, a subset of $\mathbb{R}^m$, where in our case $m = 2$. This assumption seems to be weak, but for practical purposes it deserves closer attention: a reliable estimation of M_i over all $\mathcal{I}$ in a first stage regression will require the data to cover also the boundaries of the interval sufficiently well. Simply defining $\mathcal{I}$ as set spanned by minima and maxima of x_1 and x_2 over all bonds and active trading days will usually not lead to such a situation. We therefore suggest to standardize the liquidity measures according to

$$\mathbf{x}_{j,it} := \frac{x_{j,it} - \min_{t \in \mathcal{T}_i} x_{j,it}}{\max_{t \in \mathcal{T}_i} x_{j,it} - \min_{t \in \mathcal{T}_i} x_{j,it}}, \quad j = 1, 2. \tag{3.6}$$

Hence, every bond also covers boundary regions of $\mathcal{I}$, which is the unit plane $[0, 1]^2$. In what follows, we collect the two liquidity proxies in a vector $\mathbf{x}_{it} = [\mathbf{x}_{1,it}, \mathbf{x}_{2,it}]^\top$, serving as argument of M_i.

3.3.2 Market Liquidity

The above measures allow to examine the effect of idiosyncratic liquidity on the yield spread, which is the object of main interest in our work. In most situations not only idiosyncratic but also market liquidity is priced. Hence, we suggest to control for the market environment, including a market liquidity proxy in the set of control variables collected in z_{it}. For this purpose we rely on the intraday price dispersion as suggested by Han and Zhou (2006). As argued in their work, price dispersion within a day is, in the absence of news about fundamentals, mainly due to the bid-ask

spread and thus a well suited proxy for liquidity.
Formally we obtain a market-aggregate of their measure as follows. For a bond i at day $t \in \mathcal{T}_i$ denote the empirical r-quantile of transaction prices $\{p_{it,\tau} : 1 \leq \tau \leq S_{it}\}$ as $\mathcal{Q}_{it}(r)$. Inspired by the measure in Han and Zhou (2006), we calculate from these quantiles the scaled price dispersions $d_{it} := (\mathcal{Q}_{it}(0.9) - \mathcal{Q}_{it}(0.1))/\bar{p}_{it}$, where the denominator is the mean price. The individual measures d_{it} are then aggregated to obtain our market liquidity proxy according to

$$D_t := \frac{1}{|\mathbf{N}_t|} \sum_{i \in \mathbf{N}_t} d_{it},$$

where the index set $\mathbf{N}_t$ collects all bonds, for which a measure d_{it} can be calculated at day t. In analogy to the idiosyncratic liquidity measures, large values of D_t are to be interpreted as an indication of low market liquidity.

3.4 Empirical Evidence: The US Corporate Bond Market

We confront the approach outlined in the previous sections with data for the US corporate bond market. The primary reason is the availability of rich data on corporate bond transactions. As imposed by the US Financial Regulatory Authority, almost all important brokers and dealers in the US are obliged to report every transaction. This data has been used recently e.g. by Dick-Nielsen (2012) or Friewald et al. (2012). In this section, after a description of the data and its processing, we discuss the empirical estimation results for our approach.

3.4.1 Data

We employ trade data for a large number of US corporate bonds gathered from TRACE (Trade Reporting and Compliance Engine). Furthermore, we employ Thomson Reuters' Eikon, CUSIP and COMPUSTAT data to obtain bond and issuer characteristics as well as returns of the Dow Jones Industrial Average index. Spreads are constructed using reference yields from the daily sovereign yield curve, obtained from the data set provided by Gurkaynak et al. (2006).
The considered sample covers the time between 2004-01-01 and 2012-12-31. Inspired by Dick-Nielsen (2012), we discard earlier observations as the coverage of TRACE was much lower in its initial stage, at which dominantly large investment grade bonds were taken into account.
In order to deal with errors in the data, we first employ the filter suggested in Dick-Nielsen (2009). Then, inspired by the median filter, we cope with potential outliers deleting all trades for which

the reported price deviates by more than 10% from its daily median.[5] For a given bond, we also remove trading days with less than three transactions. At any point in time, only bonds with time to maturity between one and thirty years are considered. Furthermore, we remove securities with less than 50 active trading days.
From the remaining observations, we calculate for each bond a daily spread by averaging over the calculated intraday spreads. The liquidity measures are computed from the intraday observations as outlined in Section 3.3. Finally, our sample comprises 4512 bonds of 1167 issuers. Thus, the number of assets is smaller than in other studies, but the sample still covers the full range of bond characteristics.

Interest Rate and Equity Market Controls

We employ three control variables capturing level, slope and curvature effects of the risk free yield curve. Each one is calculated from the daily US sovereign yield curve reported in Gurkaynak et al. (2006). Shifts in the level are reflected by the mean yield over all available maturities, while the slope is calculated as the spread between yields at 29 years and 1 year time to maturity. A curvature effect is captured by the so-called butterfly spread, defined as the difference between the yield at time to maturity 15 years and the mean from yields corresponding to 29 years and 1 year time to maturity.
As corporate bond yields typically exhibit a dependence on equity dynamics, we also consider daily returns of the Dow Jones Industrial Average index.

Bond Characteristics and Issuer Information

The bond characteristics used in the analysis are the time to maturity (maturity dates), dummies signaling embedded call options, (fixed) coupon rates, Standard & Poor's (S&P) bond ratings as well as the original amount issued. The considered information on the issuing firms comprises the S&P quality rank, Fitch long term credit ratings and industry classifications.
A summary table (Table 3.A.1) is provided in the appendix.

3.4.2 Estimating Liquidity Effects

We follow the three-step procedure as described in Section 3.4.2 using model (3.1) in levels.
In Section 3.3 we introduced a standardization of the liquidity measures, the effect of which can be inferred from Figure 3.A.1 in the appendix. Marginal distributions of $\mathbf{x}_{1,it}$ and $\mathbf{x}_{2,it}$ are both heavily right-skewed when pooling the proxies over assets and time. This indicates the dominance of situations with high idiosyncratic liquidity relative to peaks of illiquidity. Such peaks are for

[5] Cf. Friewald et al. (2012) and references therein.

the vast majority of bonds during the 2008 financial crisis or within the following year. The correlation between $\mathbf{x}_{1,it}$ and $\mathbf{x}_{2,it}$ amounts to 55% when pooling data over bonds i and days t.

First Stage: Bond Specific Event Time Regressions

For the first stage regressions, we collect the aforementioned interest rate and equity market control variables, the bond's time to maturity, three autoregressive terms as well as the market liquidity proxy described in Section 3.3 in the vector of controls z_{it}. We employ a bond specific complexity of the spline approximation for M_i, the basis dimension of which is chosen to depend on the sample size according to $\lfloor 10 \cdot T_i^{2/9} \rfloor$.[6]

To reduce the marginal impact of single data points, we remove outliers as identified by Cook's distances. After deleting such outliers, we relabel lags accordingly. Furthermore, the liquidity measures are again transformed according to (3.6) in order to ensure that minimum and maximum are at zero and one, respectively.

A battery of specification tests is applied to the fitted regressions. First of all, we run Augmented Dickey Fuller (ADF) tests on the residual series. Using a specific lag length of $\lfloor (T_i - 1)^{1/3} \rfloor$ and no trend or constant, the presence of a unit root is rejected for 91 per cent of the assets at the 5 per cent level. The results of tests for residual autocorrelation are summarized in Table 3.1.

	$p > 0.10$	$p > 0.05$	$p > 0.01$
Ljung-Box	0.62	0.70	0.81
Box-Pierce	0.63	0.71	0.82

Table 3.1: **Residual Correlation.** *Share of estimated models for which the tests deliver a p-value as indicated. The lag lengths for the two Box tests are both set to ten.*

The Box tests indicate that for a majority of the regressions there is negligible autocorrelation in the residuals. For example, at the 1 percent level this holds true for more than 81 per cent of the residual series.

	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)
Adjusted R^2	0.63	0.86	0.94	0.97	0.98
Explained Deviance	0.68	0.88	0.95	0.98	0.99

Table 3.2: **Adjusted R^2's and Explained Deviances of first Stage Regressions.** *Table reports quantiles of the empirical distribution of adjusted R^2's and explained deviances over different bonds.*

Table 3.2 reports classical goodness-of-fit measures. As 90 per cent of the regressions yield an adjusted R^2 higher than 63 per cent, we regard the quality of the corresponding fits as globally

[6] See the discussion and corresponding reference in Wood (2006), p. 157.

satisfactory, a view supported by the explained deviances. To more clearly isolate the importance of the idiosyncratic liquidity effect, we run the tests shown in Section 3.2 for problems **P1** and **P2**. Table 3.3 summarizes the results for **P1**.

	$p < 0.10$	$p < 0.05$	$p < 0.01$	$p < 0.001$
Share of Regressions	0.77	0.72	0.63	0.52

Table 3.3: **Test Results for Problem P1.** *Quantiles of the empirical distributions of p-values across regressions for the F-Test shown in (3.5) in Section 3.2.*

For about 77% of the bonds, the idiosyncratic liquidity effect is significant at the 10 per cent level. For almost half of the bonds it is even significant at the 1 per cent level. Whether this effect is linear is tested using the F-test shown in Section 3.2. The corresponding results are provided in Table 3.4.

	$p < 0.10$	$p < 0.05$	$p < 0.01$	$p < 0.001$
F-Test (Full Sample)	0.67	0.61	0.49	0.37
F-Test (Reduced Sample, $p < 0.1$)	0.60	0.56	0.46	0.36
F-Test (Reduced Sample, $p < 0.05$)	0.57	0.54	0.54	0.36

Table 3.4: **Test Results for Problem P2.** *Quantiles of the empirical distributions of p-values across regressions for the F-Test in (3.5) with parametric (linear) liquidity contribution (additive, no mixed effects) in the null model. The reduced samples consist of all bonds for which there is a significant liquidity effect at the level indicated in parentheses.*

The findings reported in Table 3.4 show an obvious tendency: for the majority of the bonds for which there is a significant liquidity effect, this effect is significantly nonlinear. Note however that this finding does not imply that $\hat{M}_i$ is nonlinear in all directions as the tested hypothesis is that of *joint linearity*.

It may also be the case that, depending on the value of the remaining component in $\mathbf{x}_{it}$, one component of liquidity has linear effects in some situations and non-linear ones in others. This is of particular economic interest: for example, is the effect of the price impact higher in times of high or low transaction-costs in round-trips? The suggested model (3.1) allows to answer such questions explicitly from data.

Figure 3.1 as well as Figures 3.B.1 and 3.B.2 in the appendix illustrate the contributions of the measures $\mathbf{x}_1$ and $\mathbf{x}_2$ to the yield spreads in different situations. First, Figure 3.1 shows the effect along the *main diagonal*, i.e. the evaluation of $\hat{M}_i$ at liquidity measures taking all the same values. Second, 3.B.1 shows the effects of the measure of price impact, $\mathbf{x}_1$, in cases of high or low imputed round-trip costs. Finally, Figure 3.B.2 shows the effect of the IRC's for situations of high or low intraday price impact.

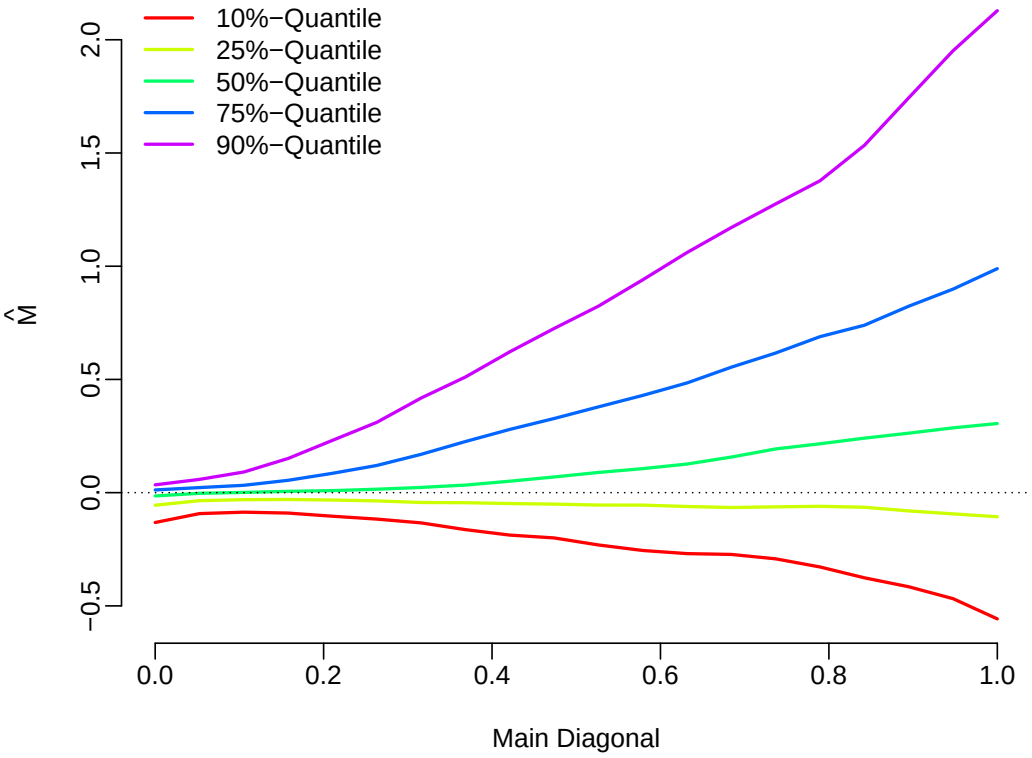

Figure 3.1: **Effects of Idiosyncratic Liquidity along the Main Diagonal.** *Figure shows (pointwise) quantiles of* $\hat{M}_i([x,x])$ *for different values of* x*, calculated over bonds* i*, having significant liquidity effects at the* 10 *per cent level according to test* ***P1****.*

The cross-section distributions of the main diagonal values of $\hat{M}_i$ (Figure 3.1) reveal substantial differences of liquidity effects, which become more pronounced as both coordinates in $\mathbf{x}_{it}$ approach 1. More concretely, the distributions are more dispersed in this situation and probability mass shifts to the right, meaning that when both considered measures indicate lower liquidity, the liquidity contribution to the yield spread increases for the majority of bonds. Interestingly there is however also a fraction of bonds for which the effect of idiosyncratic illiquidity remains negligible or even decreases in times of low idiosyncratic liquidity. As Figures 3.B.1 and 3.B.2 suggest, the impact of a single idiosyncratic liquidity measure strongly depends on the level of the other measure. In particular heterogeneity of liquidity effects among bonds is much higher if one single measure indicates low liquidity.

Second Stage: Classification of Regression Curves

In the first stage, a number of 3254 bonds were found to have significant liquidity effects at the 5 per cent level. As indicated in Section 3.2, all other bonds are *pre-classified* to be members of the group $\hat{G}_0$. In order to avoid obtaining groups having only one or two members, we further discard bonds for which $\hat{M}_i$ takes *extreme* values. To be precise, we calculate values of $\hat{M}_i$ on a grid in $\mathcal{I}$ containing 400 points and delete those assets having values of $\hat{M}_i$ which are below the 1% quantile or above the 99% quantile of the distribution of the pooled values. Focusing on the resulting set of 2936 bonds, estimated curves $\hat{M}_i$ are clustered using Ward's algorithm, based on the dissimilarity measure in (3.4).[7]

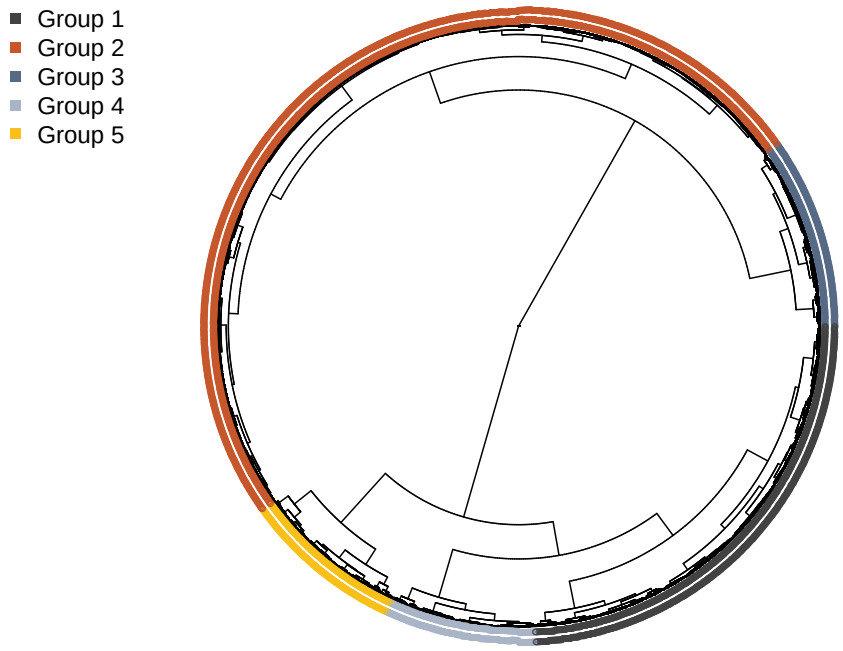

Figure 3.2: **Radial Dendrogram from Clustering Liquidity Effects.** *The dendrogram displays group and distance structures for estimated liquidity effects $\hat{M}_i$. Classification results are indicated by the color scheme.*

To estimate the number of groups, we employ the *McClain* index as described e.g. in Charrad et al. (2014). We set the lower and upper bounds to be 3 and 15, avoiding the appearance of

[7]Regarding implementation, we used the algorithm outlined in Murtagh and Legendre (2014).

very small or very large groups. For comparison, note that in Dick-Nielsen (2012) the number of classes resulting from rating/time to maturity buckets is determined to be 15 (both, in pre- and post-subprime crisis samples). According to the McClain criterion, we find the optimal number of clusters to be $\hat{K} = 5$. This appears to be quite robust to the choice of the cost function: the *Silhouette* and *Dunn* indexes as described in Charrad et al. (2014) suggest $\hat{K} = 5$ and $\hat{K} = 7$. The five clusters constructed by Ward's algorithm are depicted in Figure 3.2.

The algorithm identifies two large classes, $k = 1, 2$, accounting together for about 74% of the assets. The remaining bonds are found to constitute three more groups of smaller size. Remarkably, as can be seen from Figure 3.2, groups 2 and 3 are considered to be close to each other but have a large distance to the remaining clusters. Another bucket is formed by the remaining three classes $k = 1, 4, 5$.

Table 3.B.1 in the appendix summarizes the basic bond characteristics within the different groups. Groups 2 and 3 comprise on average larger bonds than the remaining groups what concerns the initial amount issued. Classes 4 and 5 collect the smallest bonds on average. Table 3.B.2 and Figures 3.B.3-3.B.5 in the appendix report the distribution of maturity years and shares of investment-grade / speculative-grade assets and issuers within the different classes. Bonds in groups 2 and 3 appear to mature substantially later than in the other classes, in which the average maturity year is between 3 years ($k = 1, 5$) and 5 years ($k = 4$) smaller. Beyond that, bonds in the largest group $k = 2$ also receive the best credit ratings on average. In sharp contrast to that, the fifth group collects bonds with substantially worse average rating. Numerically, shares of investment grade bonds range from 68% (group 5) to 91% (group 2), while the shares in the remaining groups are between 75% and 85%. In lines with this observation issuers in the second group receive on average the highest quality ranks, while firms in the fifth group receive the lowest quality ranks (see Table 3.B.3 in the appendix).

Interestingly, as reported in Table 3.B.4 in the appendix, the level of spreads seems to be somewhat smaller in the largest group $k = 2$. This is contrasted by the remaining groups revealing in tendency higher levels: in particular the fifth group contributes to this result. The dissimilarity of groups 2 and 5 also transfers to the trading activity. Bonds in the second group are most actively traded while bonds in the last group are least actively traded as reported in Table 3.B.5 in the appendix. Regarding the number of trades, the remaining groups are located between these two extremes, whereas the third group appears to be closer to the second and the first and fourth group closer to the fifth class. As we alluded to before, the cluster algorithm seems to confirm this result also with respect to the dissimilarity between the different regression functions, as can be seen from Figure 3.2.

Furthermore, Table 3.B.6 reports the distribution of issuers' industry classifications within the different groups. These distributions however do not show a striking pattern which would allow

to distinguish groups.

Third Stage: Estimation of Group specific Liquidity Effects

Given the classification of the second stage, we obtain group specific curves $\hat{g}_k$ as the weighted average of estimated regression functions $\hat{M}_i$ belonging to group k. The resulting estimates $\hat{g}_1, ..., \hat{g}_5$ are depicted in Figures 3.3 and 3.4. Further Figure 3.B.6 in the appendix shows corresponding heat plots.

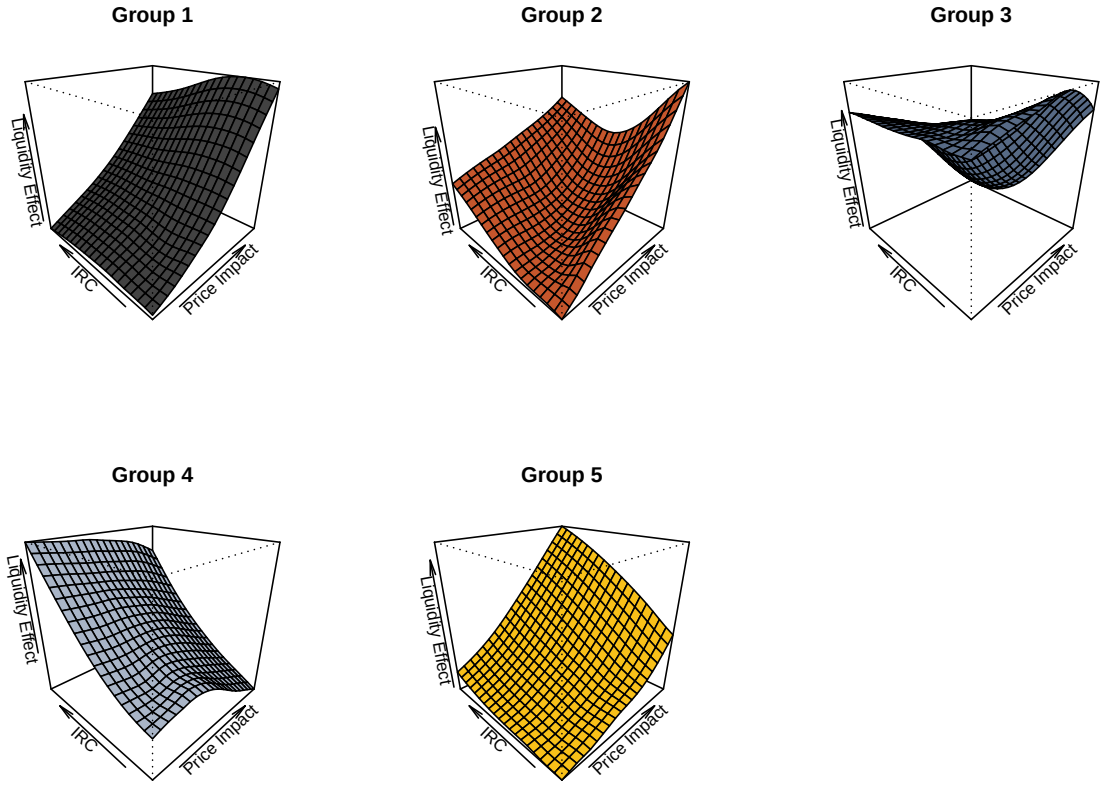

Figure 3.3: **Estimated Regression Functions by Group.** *Figure reports estimates* $\hat{g}_1, \ldots, \hat{g}_5$. *The units on the (vertical) z-axis differ across panels of the figure.*

The effects of idiosyncratic liquidity on the yield spreads are most pronounced in group 5 as becomes visible in Figure 3.4. Classes 1 and 4 also show substantial effects, while the liquidity contribution to the yield spread is only moderate in classes 2 and 3.

What concerns shapes of the functions $\hat{g}_k$, our model distinguishes between three types of pricing illiquidity. For groups 1 and 4 dominantly one of the two illiquidity proxies is priced, while both are nearly symmetrically priced in groups 2 and 5. For the third group the liquidity effect is am-

biguous. While the first group and the fourth group share a similar generic type of liquidity effect, the exact pricing is opposite: in class $k = 1$ bond spreads are almost exclusively effected by illiquidity as measured by the price impact. For the fourth group liquidity effects are largely driven by the imputed round-trip costs. For groups 2 and 5 both measures appear to jointly govern the contribution of illiquidity to the spread. In particular the liquidity effect increases nonlinearly along the main diagonal, i.e. in situations, in which both of our measures indicate the same magnitude of illiquidity. For the second group the liquidity effect is not as symmetric as in the fifth group: in case of low costs from imputed round-trips, the Amihud (2002)-measure has a particularly large impact on the liquidity effect. In the third group the impact of the liquidity proxies on the yield spread is highly nonlinear. For some values of the measures it is even slightly counter-intuitive, as e.g. the price impact has a decreasing marginal effect on the yield spread in some situations.

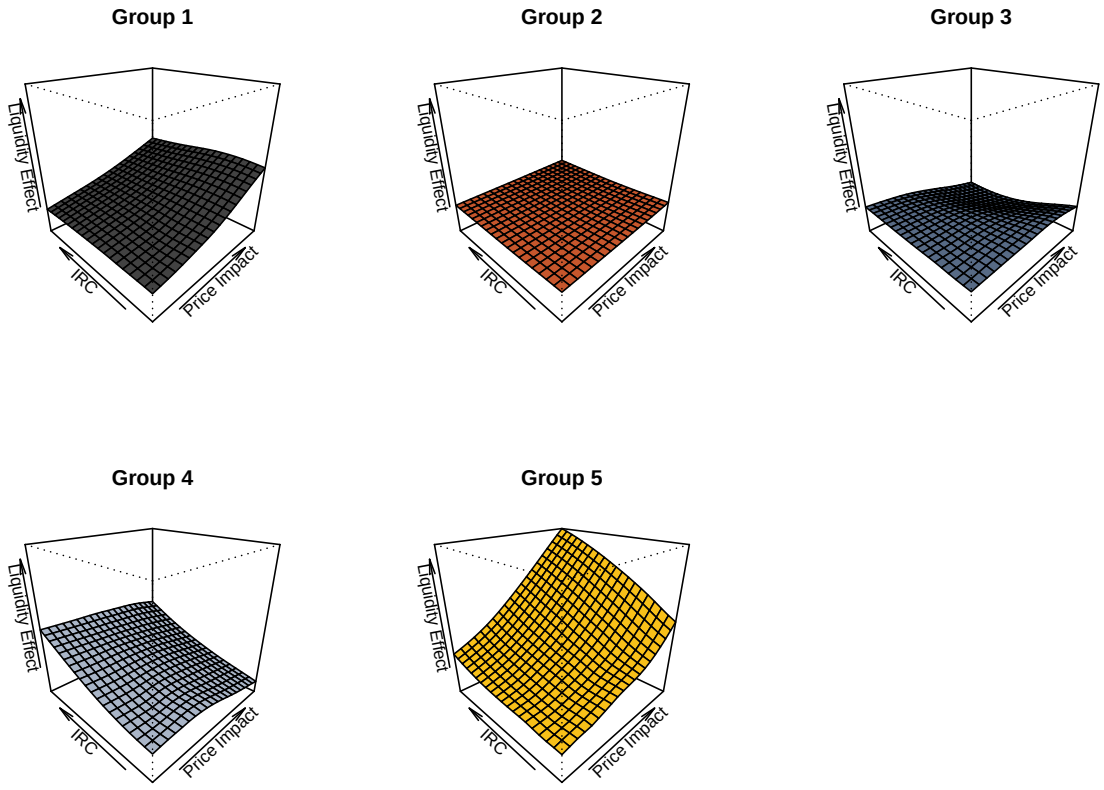

Figure 3.4: **Estimated Regression Functions by Group: Common Scale.** *Figure reports estimates $\hat{g}_1, \ldots, \hat{g}_5$ on a common scale on all axes.*

The estimates $\hat{g}_k$ now also allow us to assess differences in liquidity effects between groups nu-

merically. Central questions of economic relevance are twofold: first, do the effects vary equally strong for different degrees of illiquidity among groups? And second: which combination of liquidity measures maximizes the distinction between groups? The first question can simply be tackled considering the difference between ranges of function values, i.e.

$$R_k := \max_{x \in \mathcal{I}} \hat{g}_k(x) - \min_{x \in \mathcal{I}} \hat{g}_k(x) \tag{3.7}$$

for $k = 1, ..., \hat{K}$. We addressed the second question already visually and quantify these observations by comparing functions $\hat{g}_k$ and $\hat{g}_l$ according to the location of their maximum distance, i.e.

$$D_{k,l} := \arg\max_{x \in \mathcal{I}} \left| \hat{g}_k(x) - \hat{g}_l(x) \right|. \tag{3.8}$$

The results for R_k as well as locations of maxima and minima are reported in Table 3.5, the locations maximum differences in Table 3.6.

	Group 1	Group 2	Group 3	Group 4	Group 5
R_k	1.33	0.14	0.98	1.63	3.39
arg max	(1,0.21)	(1,0)	(0.63,0)	(0,1)	(1,1)
arg min	(0,1)	(0,0)	(1,0.89)	(1,0)	(0,0)

Table 3.5: **Range of Function Values, Maximizer and Minimizer of $\hat{g}_k$.** *Table reports rounded values. In rows two and three, the first coordinate refers to the measure of price impact, the secod one to the imputed round-trip costs.*

The most striking observation is that group functions have, in most cases, optima as well as maximum differences at the boundaries of their support. Except for the fourth group the maximum is for all groups attained at a large value of the price impact measure. Group 5, which was identified to have the largest liquidity effect, attains a minimum at $(0, 0)$ and a maximum at $(1, 1)$. This behavior matches the intuition: large values of $\mathbf{x}_1$ and $\mathbf{x}_2$ mean high illiquidity, which tends to imply large (idiosyncratic) liquidity premiums.

The range of function values over $\mathcal{I}$ is comparably small for the largest group. The opposite is true for the groups with more substantial liquidity effects, $k = 1, 4, 5$, in particular for the fifth group. Between groups 5 and 2, for example, the difference can be quantified with a scaling factor close to 24. For those bonds which tend to have higher spreads, shorter maturities, worse ratings and smaller sizes, the effects of liquidity vary much stronger.

Differences as quantified by $D_{k,l}$ confirm the intuition gathered from above: remarkably, for almost all groups, differences to other groups are most pronounced for extreme values of the id-

	Group 1	Group 2	Group 3	Group 4
Group 2	(1,0.26)			
Group 3	(1,0.63)	(1,0.89)		
Group 4	(1,0)	(0,1)	(0.95,1)	
Group 5	(1,1)	(1,1)	(1,1)	(1,0.74)

Table 3.6: **Locations of maximum Level Differences** $D_{k,l}$**.** *Table reports rounded locations of level differences as in (3.8).*

iosyncratic liquidity measures. The magnitude of liquidity contributions particularly often discriminates groups in situations of a high intraday price impact.

3.4.3 Time Variation of Liquidity Effects

Our estimation procedure revealed substantial differences of magnitudes, shapes and interpretations of the functions $\hat{g}_k$ between groups. Whether this translates into different function values over time certainly depends on the evolution of liquidity proxies over time. We assess the dynamics of liquidity and its effects of the yield spreads as follows.
What concerns the dynamics of the liquidity proxies, we calculate a daily average liquidity measure

$$\overline{\mathbf{x}}_{j,kt} := \frac{1}{|\mathbf{N}_{kt}|} \sum_{i \in \mathbf{N}_{kt}} \mathbf{x}_{j,it}, \quad j = 1, 2 \tag{3.9}$$

for each group $k \in \{1, ..., \hat{K}\}$. The set $\mathbf{N}_{kt} \subset \hat{G}_k$ comprises all members of group k for which there is an observation at day t. Figures 3.B.7 and 3.B.8 in the appendix visualize the group specific time series of the two measures $\overline{\mathbf{x}}_{j,kt}$, smoothed by averaging within rolling windows.
In analogy we compute the time series of average liquidity effects within an estimated group k according to

$$\overline{\hat{g}_k(\mathbf{x})}_t := \frac{1}{|\mathbf{N}_{kt}|} \sum_{i \in \mathbf{N}_{kt}} \hat{g}_k(\mathbf{x}_{it}).$$

Again, to improve the visualization, we calculate rolling means for each series. Figure (3.5) allows to compare average idiosyncratic liquidity effects between groups.
Figures 3.B.7 and 3.B.8 indicate very similar dynamics of the liquidity proxies over time. Both measures jump to a global peak during the financial crisis. Interestingly, in the run-up to the crisis our proxies already show a slight increase starting from the bankruptcy of New Century Financial. Turning to the resulting impact on the yield spreads, the most striking observation is that the series

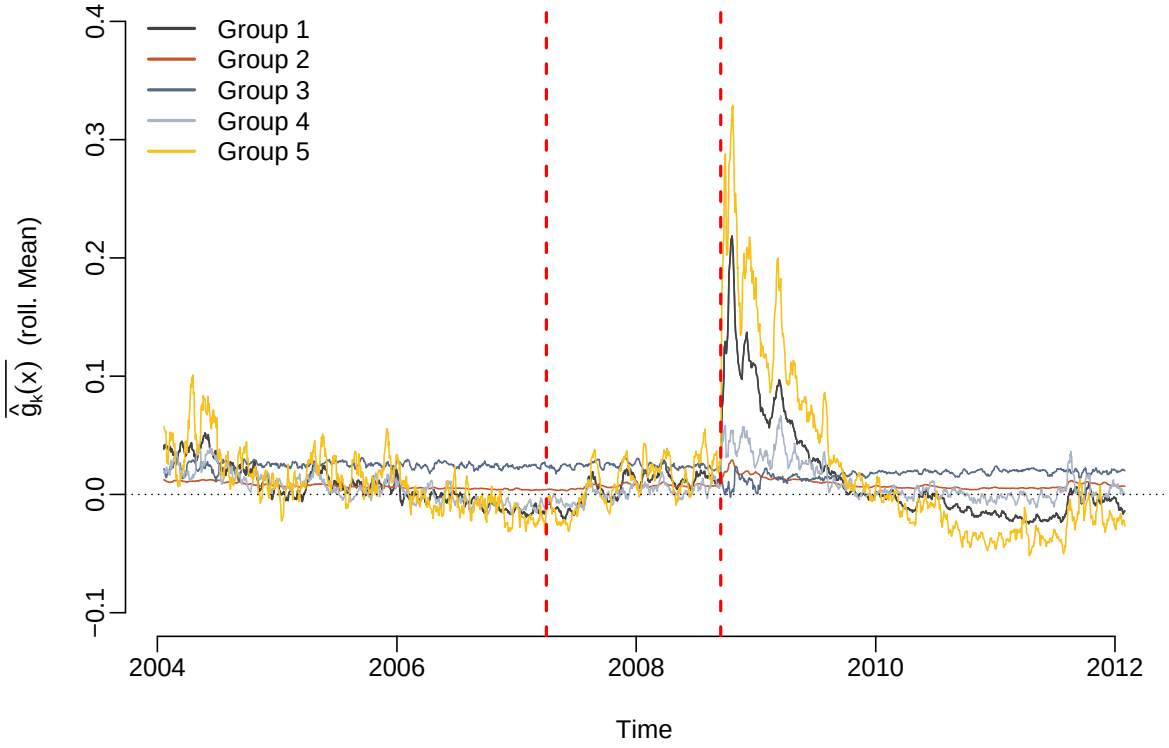

Figure 3.5: **Effect of Idiosyncratic Liquidity.** Figure reports smoothed values of $\overline{\hat{g}_k(\mathbf{x})}_t$ over time. Smoothing proceeds according to a (left-sided) rolling mean with a window length of 10 trading days. Vertical dashed lines are at $2007-04-02$ (New Century Financial bankruptcy) and $2008-09-15$ (Lehman Brothers bankruptcy).

corresponding to groups $k = 1, 4, 5$ tend to have a very similar behavior up to a scaling factor. This impression is indeed particularly supported by the 2008 financial crisis: for all such groups, the liquidity contribution increases quite exactly at Lehman Brothers' bankruptcy. The magnitude of this increase differs among groups and, in particular, is more than thirty basis points for the fifth group. Note that this increase is in excess to the effect of (declining) market liquidity, which we controlled for in our regression model.

Albeit the absolute levels differ among groups, timing and direction of changes are quite similar between most groups. This is underlined by the empirical correlation between function values among groups. As can be seen from Figure 3.B.9 in the appendix, correlations are between 49 and 87 per cent, when discarding the third group. Average liquidity effects of the latter are slightly negatively correlated to the series from the other groups. This emphasizes the ambiguous role of

the third group.
In light of the above observations—apart from the third group—it appears that liquidity effects (i) are similar among bonds up to scale, and (ii) differ substantially only in times of financial turmoil.

3.5 Conclusion

In this work we introduced a novel way of estimating idiosyncratic liquidity effects in corporate bond yield spreads from an unbalanced panel of securities. The suggested semiparametric longitudinal data model enables us to avoid high degrees of temporal aggregation as typically found in the related literature. We allow for a latent group structure in the cross section dimension to govern similarities and dissimilarities in the effects of idiosyncratic illiquidity. The group structure can be estimated from the data. The actual liquidity effects are allowed to reflect a potentially nonlinear interplay between different facets of bond illiquidity.
In our empirical study, we estimated idiosyncratic liquidity effects as well as groups of assets sharing similar effects from transaction data of about 4500 bonds traded in the US. Controlling for market illiquidity, we found significant and significantly nonlinear effects for the majority of assets. Based on the similarity of estimated liquidity effects in the L^2 metric, we classified assets into five groups. The largest group collected high-rated bonds mostly having moderate spread levels, longer maturities and high quality issuers. The effects of illiquidity on the yield spread were found to be quite small, even in the 2008 financial crisis. This was particularly sharply contrasted by one smaller group. Such group contained high-yield bonds with worse average ratings and smaller times to maturity. At the beginning of the crisis, idiosyncratic liquidity contributions jumped by more than 30 basis points in excess to the effect of market liquidity.
An analysis of group specific effects showed that, in particular for extreme values of the liquidity proxies, functions differed among groups. The groups to distinguish in that respect were particularly the largest and the smaller high-risk group. Although the effects of illiquidity differed considerably in their functional form, function values were substantially and positively correlated over time for four out of five groups.

Appendix

3.A Data Summary

Quantiles	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)
Spread (%)	0.39	0.75	1.59	3.31	6.00
Trades/Day/Bond	3.00	5.00	9.00	18.00	34.00
Active Trad. Days/Bond	75.00	119.00	224.00	443.25	755.90
Mean Rate Yield Curve (%)	3.09	3.69	4.21	4.65	4.90
Slope Yield Curve (%-pts.)	-0.19	0.50	2.63	3.93	4.37
Butterfly Spread Yield Curve (%-pts.)	0.16	0.42	1.32	1.78	2.02
DJIA Returns (%)	-1.25	-0.47	0.05	0.55	1.21

Table 3.A.1: **Summary Statistics.** *Table reports p-Quantiles, $Q(p)$, calculated from the empirical distribution of the pooled sample.*

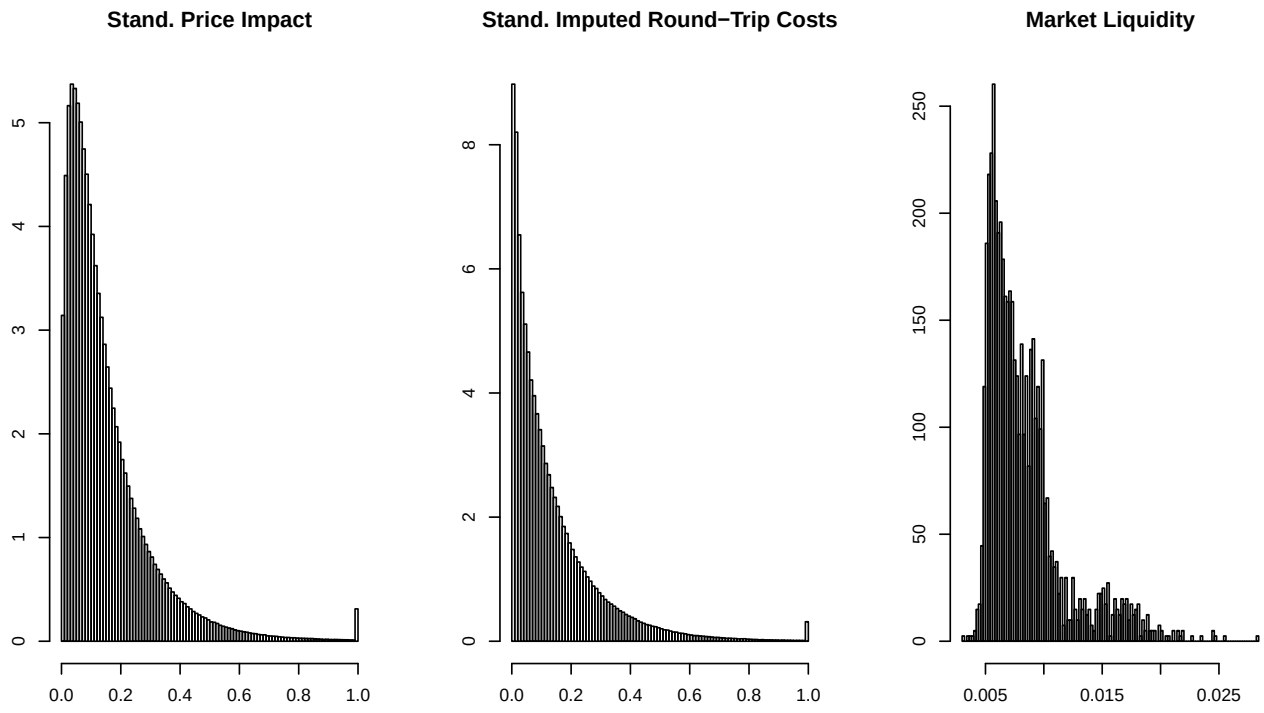

Figure 3.A.1: **Distribution of Liquidity Measures.** *Distributions are calculated from pooled data.*

3.B Additional Results

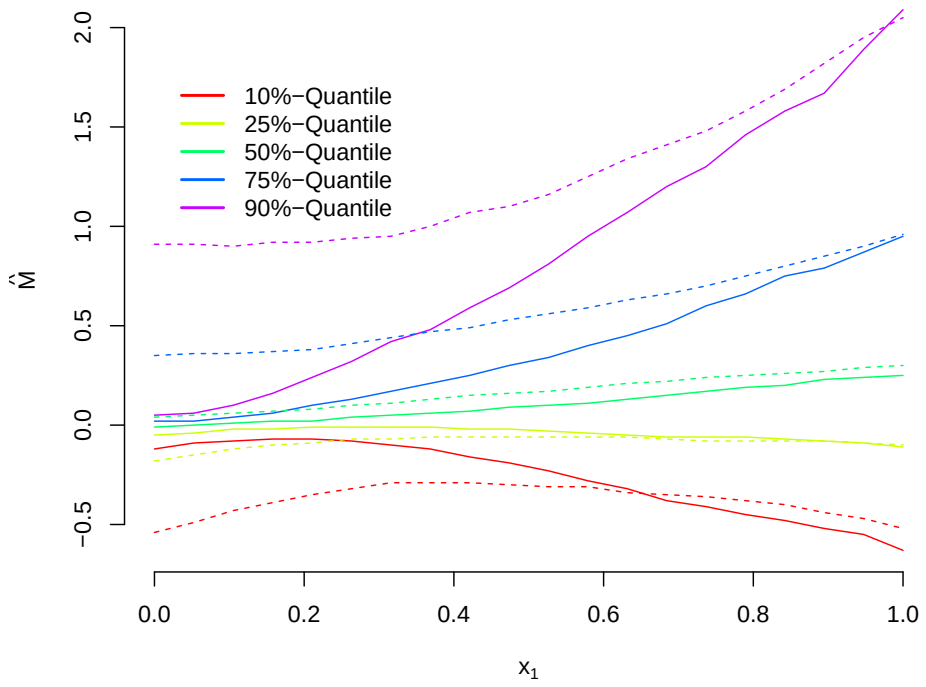

Figure 3.B.1: **Idiosyncratic Liquidity Effect: Price Impact.** *Figure shows (pointwise) quantiles of* $\hat{M}_i([\mathbf{x_1}, \mathbf{x_2}])$ *for* $\mathbf{x_2} = 0.1$ *(solid line) and* $\mathbf{x_2} = 0.9$ *(dashed lines), calculated over bonds* i *having significant liquidity effects at* 10 *per cent according to test* ***P1****.*

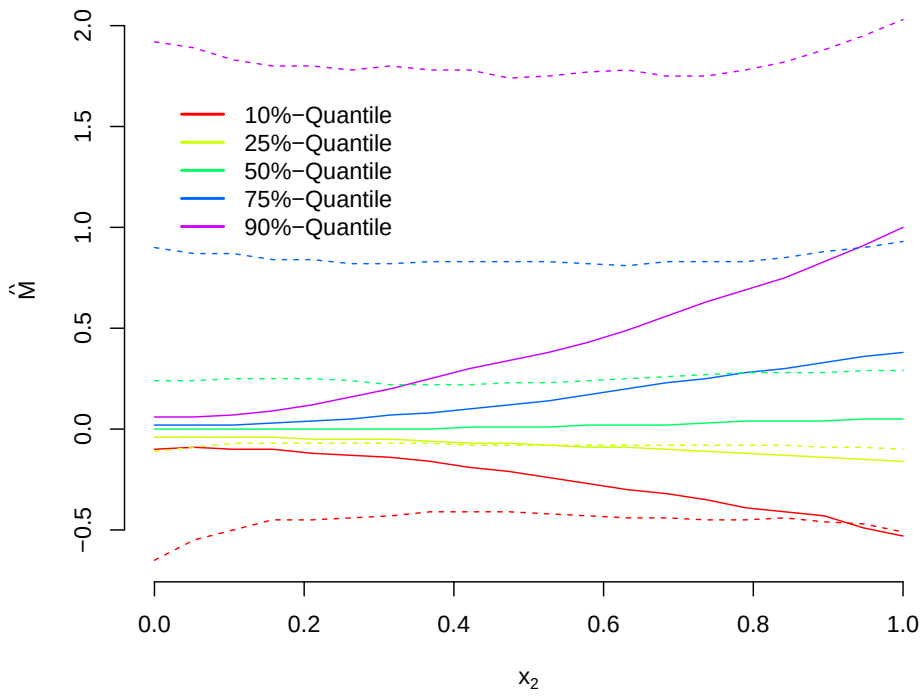

Figure 3.B.2: **Idiosyncratic Liquidity Effect: Imputed Round-Trip Costs.** *Figure shows (pointwise) quantiles of* $\hat{M}_i([\mathbf{x_1}, \mathbf{x_2}])$ *for* $\mathbf{x_1} = 0.1$ *(solid line) and* $\mathbf{x_1} = 0.9$ *(dashed lines), calculated over bonds* i *having significant liquidity effects at* 10 *per cent according to test* ***P1****.*

	Group 1	Group 2	Group 3	Group 4	Group 5
Number of Bonds	708	1475	286	228	239
Share of Fixed Rate Bonds	0.92	0.95	0.91	0.90	0.87
Share of Callable Bonds	0.64	0.71	0.75	0.58	0.63
Median Coupon Rate*	6.36	5.75	6.25	6.38	6.38
Median Amt. Issued (Mio USD)	350	500	375	300	300

Table 3.B.1: **Bond Characteristics in Estimated Groups** *Coupons: *where applicable.*

	Group 1	Group 2	Group 3	Group 4	Group 5
Mean Maturity Year	2015	2018	2018	2013	2015
Share of Investment Grade Issuer*	0.74	0.80	0.80	0.78	0.73
Share of Speculative Grade Issuers*	0.26	0.20	0.20	0.22	0.27
Share of Investment Grade Bonds**	0.79	0.91	0.82	0.85	0.68
Share of Speculative Grade Bonds**	0.21	0.09	0.18	0.15	0.32

Table 3.B.2: **Maturity Dates and Ratings by Group.** *Shares reflect only Issuers/Bonds for which a rating was available. *According to S&P Long Term Issuer Rating, **According to Fitch Rating*

	Group 1	Group 2	Group 3	Group 4	Group 5
A+	0.04	0.05	0.05	0.05	0.04
A	0.07	0.08	0.08	0.09	0.05
A-	0.07	0.09	0.08	0.05	0.05
B+	0.17	0.16	0.16	0.16	0.15
B	0.16	0.20	0.20	0.17	0.14
B-	0.18	0.15	0.19	0.17	0.21
C	0.07	0.04	0.03	0.07	0.10
D	0.04	0.02	0.04	0.04	0.05
No Rank	0.20	0.21	0.17	0.19	0.21

Table 3.B.3: **S&P Quality Ranks of Firms by Group.**

Quantiles	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)
Group 1	0.47	0.83	1.70	3.17	5.41
Group 2	0.40	0.65	1.20	2.19	3.75
Group 3	0.58	1.10	1.93	3.29	5.55
Group 4	0.48	0.76	1.55	3.75	6.71
Group 5	0.45	0.95	2.09	4.25	7.25

Table 3.B.4: **Spreads by Group in %.**

Quantiles	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)
Group 1	3	5	8	16	30
Group 2	4	6	11	21	39
Group 3	3	5	9	17	30
Group 4	3	5	9	17	33
Group 5	3	5	8	15	28

Table 3.B.5: **Number of Trades per Bond and Day by Group.**

	Group 1	Group 2	Group 3	Group 4	Group 5
Accommodation and Food Services	0.02	0.02	0.04	0.01	0.02
Admin., Support, Waste Manag. and Remediation Serv.	0.02	0.01	0.01	0.01	0.00
Agriculture, Forestry, Fishing and Hunting	0.00	0.00	0.00	0.01	0.00
Arts, Entertainment, and Recreation	0.01	0.00	0.02	0.02	0.02
Construction	0.02	0.01	0.17	0.18	0.04
Finance and Insurance	0.15	0.16	0.01	0.01	0.17
Health Care and Social Assistance	0.02	0.02	0.08	0.08	0.02
Information	0.09	0.08	0.32	0.34	0.05
Manufacturing	0.31	0.33	0.07	0.07	0.30
Mining, Quarrying, and Oil and Gas Extraction	0.07	0.07	0.01	0.01	0.06
Other Services (except Public Administration)	0.00	0.00	0.00	0.01	0.03
Professional, Scientific and Technical Services	0.02	0.01	0.05	0.02	0.02
Real Estate and Rental and Leasing	0.02	0.01	0.04	0.06	0.03
Retail Trade	0.05	0.05	0.11	0.04	0.05
Transportation and Warehousing	0.04	0.04	0.02	0.14	0.12
Utilities	0.13	0.15	0.04	0.02	0.03
Wholesale Trade	0.03	0.02	0.01	0.01	0.02

Table 3.B.6: **Industries in Estimated Groups.** *Distribution of industries within groups. Companies are classified according to NAICS keys.*

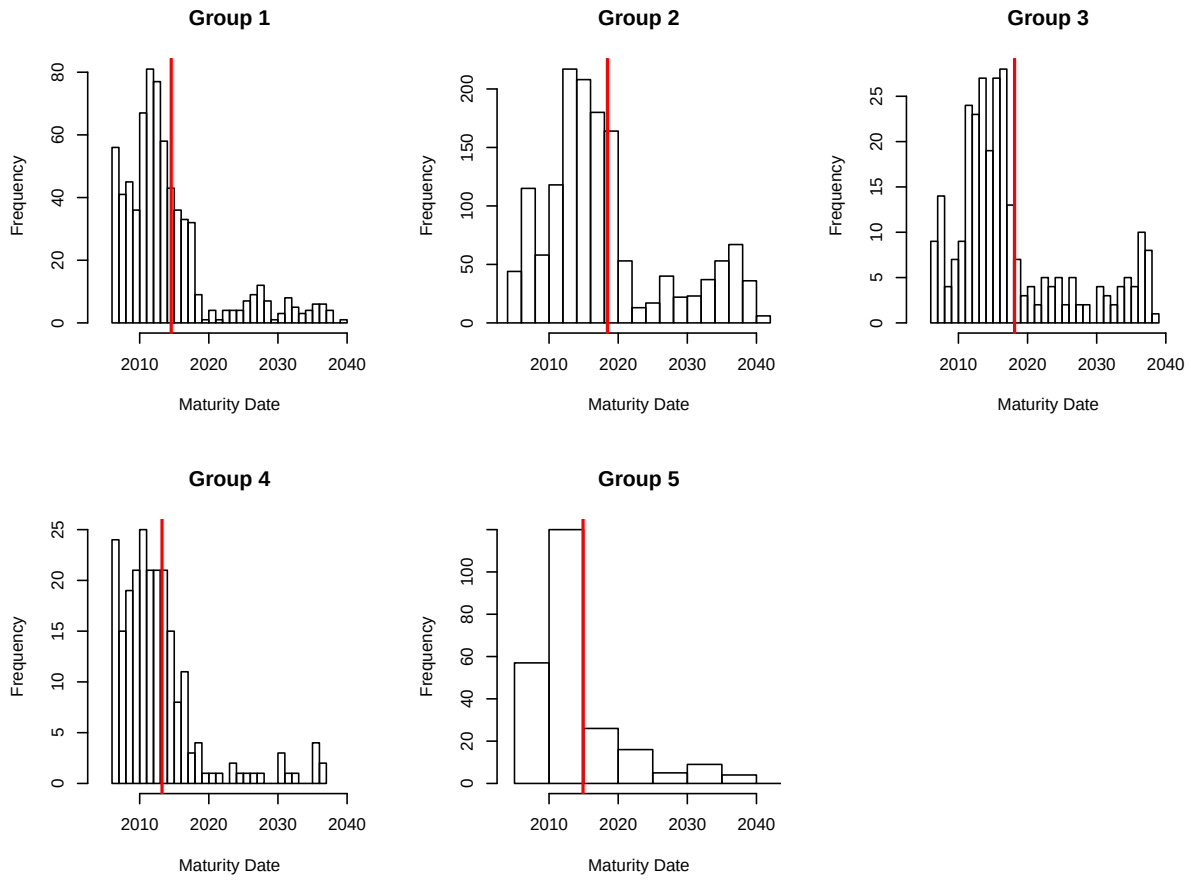

Figure 3.B.3: **Distribution of Maturity Years by Group.** *Red vertical lines are at the mean maturity years.*

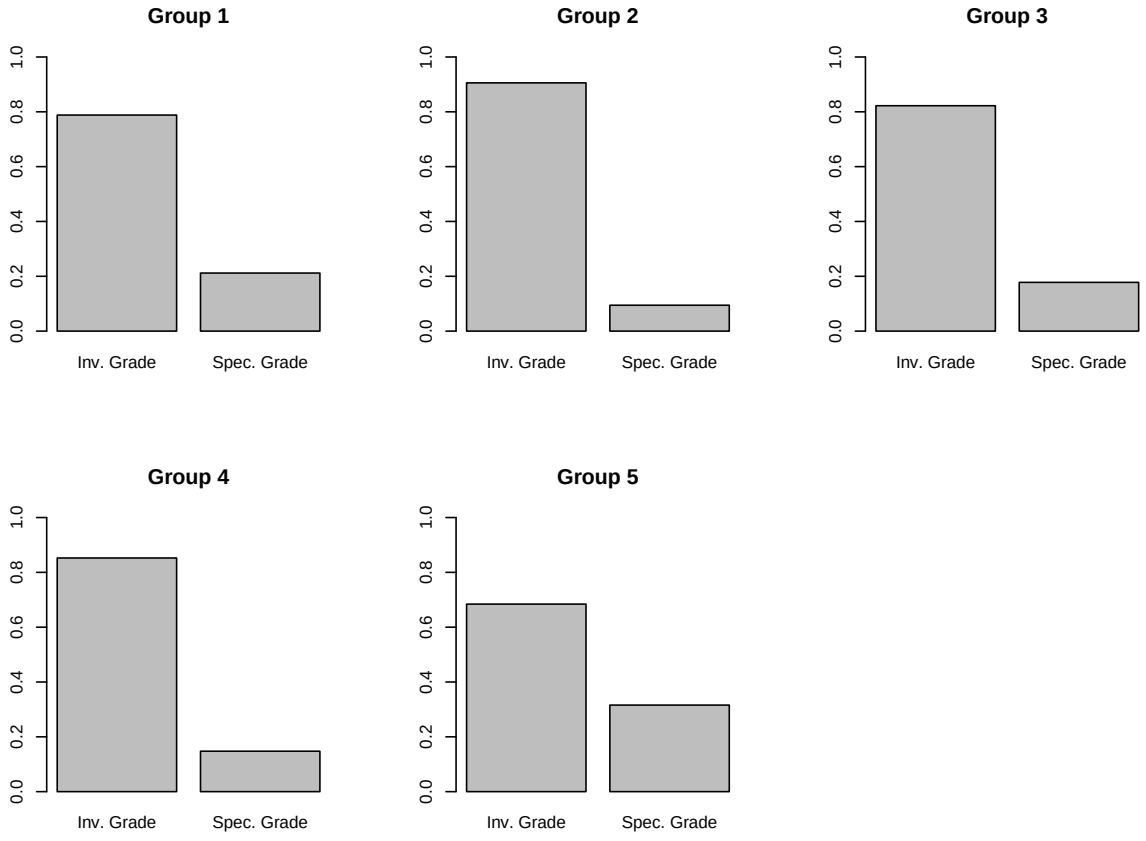

Figure 3.B.4: **Distribution of Fitch Bond Ratings by Group.** *Bonds without available rating were discared in the calculation.*

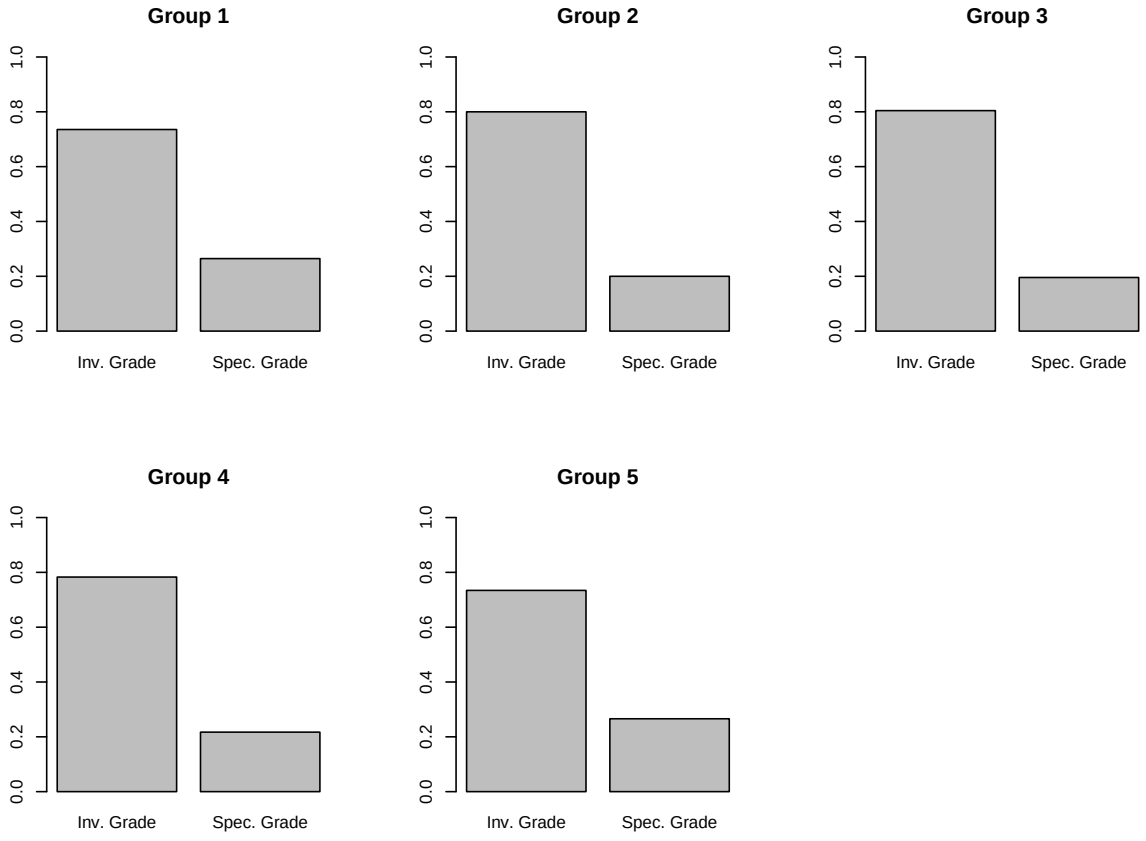

Figure 3.B.5: **Distribution of S&P Issuer Ratings by Group** *Issuers without available rating were discared in the calculation.*

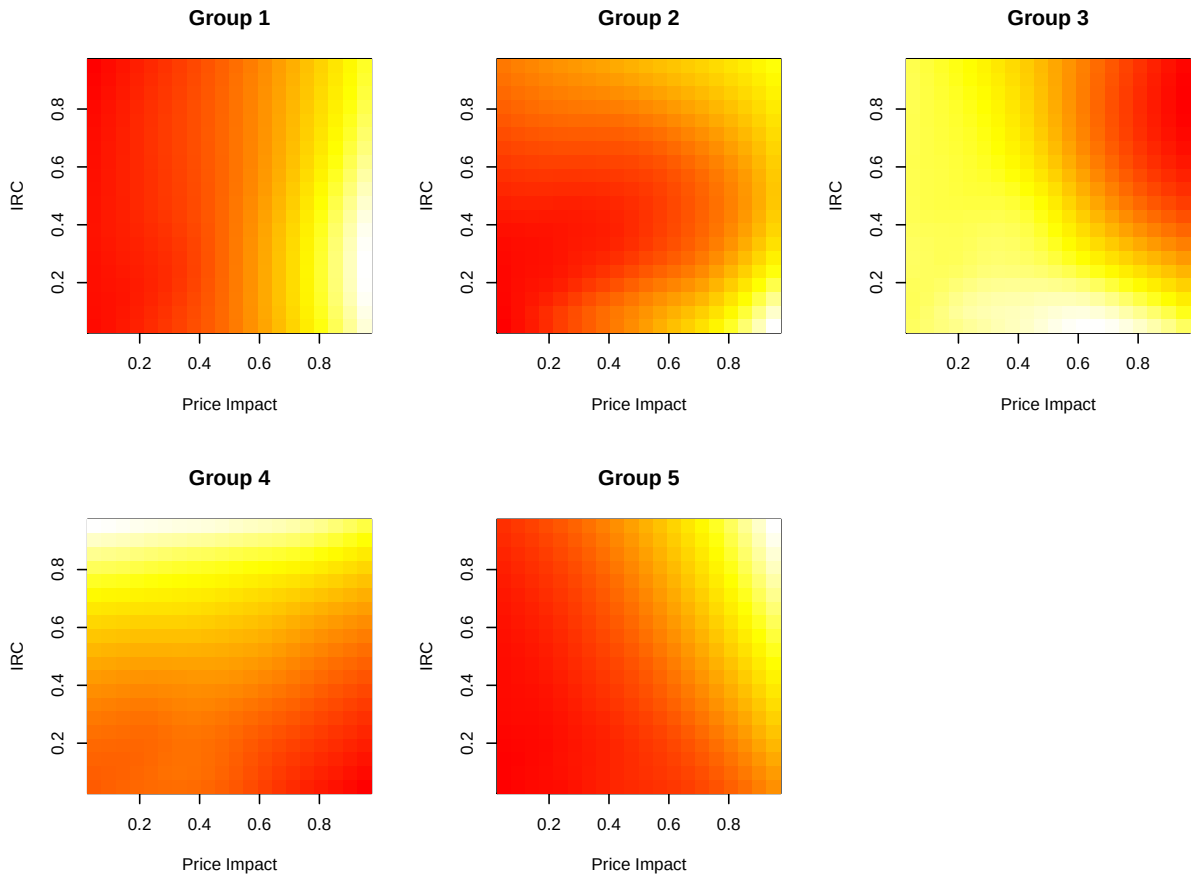

Figure 3.B.6: **Heat plots of estimated Liquidity Effects by Group.**

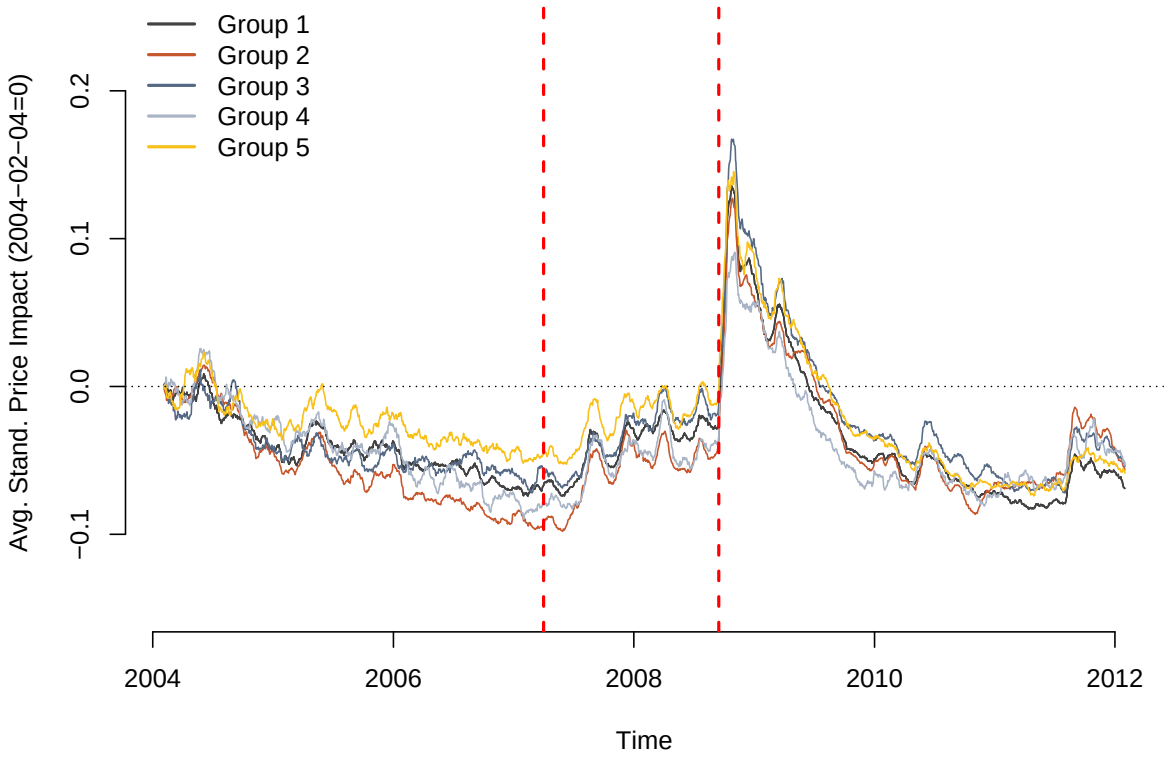

Figure 3.B.7: **Price Impact Measure over Time by Group.** *Means are calculated in rolling windows of length* 10 *days. Vertical dashed lines are at* $2007-04-02$ *(New Century Financial bankruptcy) and* $2008-09-15$ *(Lehman Brothers bankruptcy).*

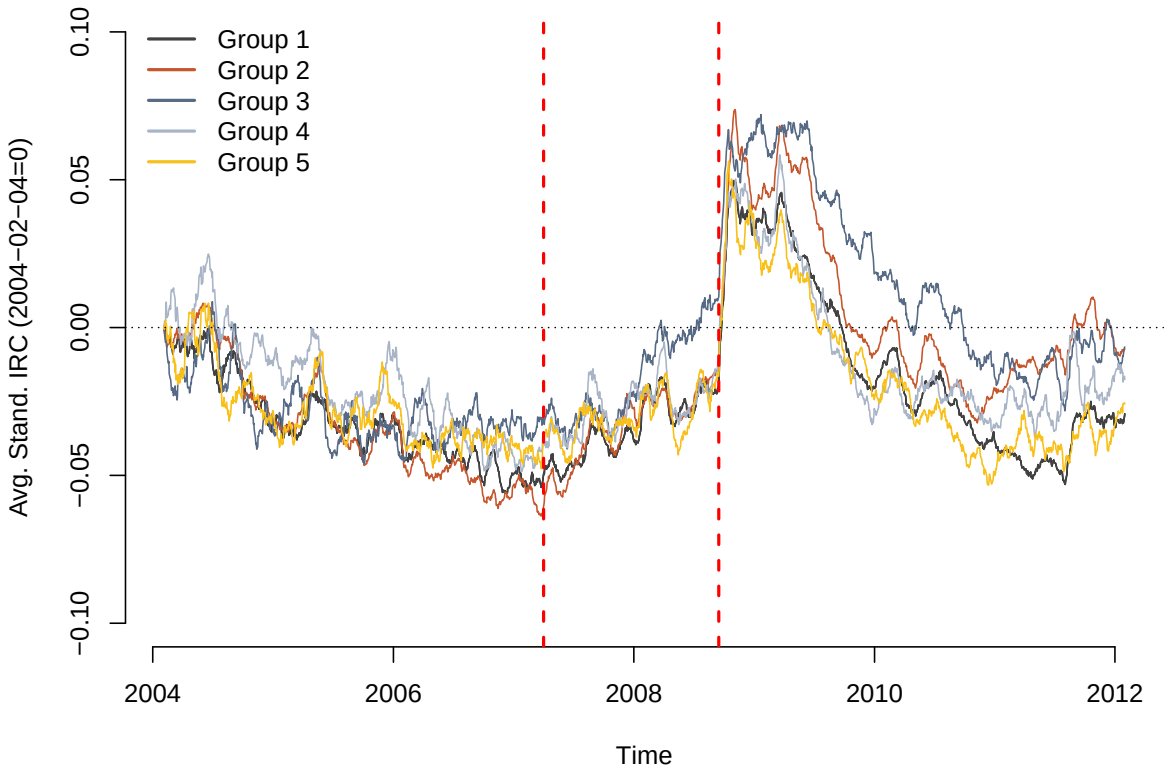

Figure 3.B.8: **Imputed Round-Trip Costs over Time by Group.** *Means are calculated in rolling windows of length* 10 *days. Vertical dashed lines are at* $2007-04-02$ *(New Century Financial bankruptcy) and* $2008-09-15$ *(Lehman Brothers bankruptcy).*

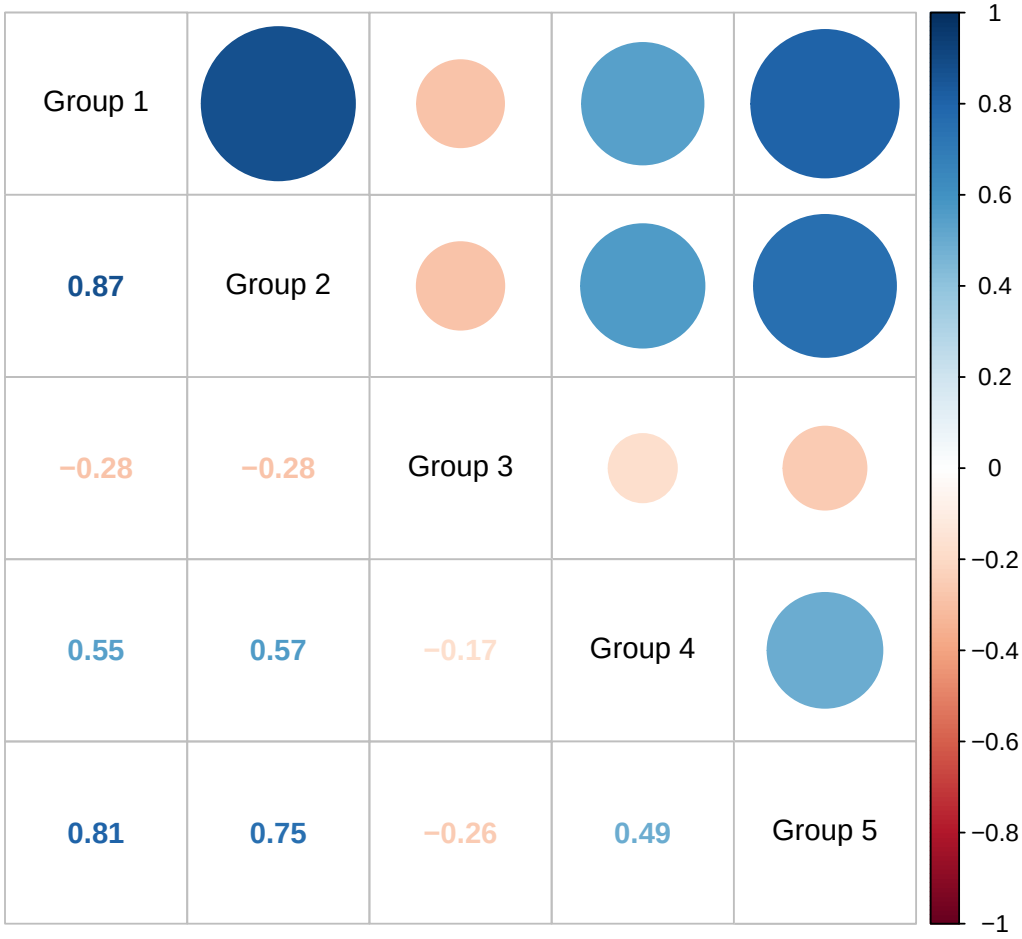

Figure 3.B.9: **Correlation of Liquidity Effects between Groups.** *Figure reports the correlation matrix between time series of mean liquidity effects* $\overline{\hat{g}_k(\mathbf{x})}_t$ *across groups.*

References

Ahn, S.C. and Horenstein, A.R. (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica*, 81 (3): 1203–1227.

Amihud, Y. (2002). Liquidity Biases in TRACE. *Journal of Financial Markets*, 5 (1): 31–56.

Amihud, Y. and Mendelson, H. (1986). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics*, 17 (2): 223–249.

Andersen, T., Bollerslev, T., Christoffersen, P. and Diebold, F. (2006). Volatility and Correlation Forecasting. In: *Handbook of Economic Forecasting*, Elliott G., Granger CWJ., Timmermann A. (Editors). Elsevier, 777-âĂŞ878.

Andersen, T. G., Dobrev, D. and Schaumburg, E. (2012). Jump-Robust Volatility Estimation using Nearest Neighbor Truncation. *Journal of Econometrics,* 169 (1): 75–93.

Aneiros-Pérez, G. and Vieu, P. (2006). Semi-Functional Partial Linear Regression. *Statistics & Probability Letters,* 76 (11): 1102–1110.

Aneiros-Pérez, G. and Vieu, P. (2008). Nonparametric Time Series Prediction: A Semi-Functional Partial Linear Modeling. *Journal of Multivariate Analysis,* 99 (5): 834–857.

Aneiros-Pérez, G. and Vieu, P. (2013). Testing Linearity in Semi-Parametric Functional Data Analysis. *Computational Statistics,* 28 (2): 413–434.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance,* 61 (1): 259–299.

Aue, A., Norinho, D.D. and Hörmann, S. (2015) On the Prediction of Stationary Functional Time Series. *Journal of the American Statistical Association,* 110 (509): 378–392.
Factor Models. *Econometrica,* 70 (1): 191–221.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society: Series B*, 64 (2): 253–280.

Bellet, A., Habrard, A. and Sebban, M. (2013). A Survey on Metric Learning for Feature Vectors and Structured Data. *arXiv preprint arXiv:1306.6709.*

Caliński, T. and Harabasz, J. (1974). A Dendrite Method for Cluster Analysis. *Communications in Statistics,* 3 (1): 1–27.

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software,* 61 (6): 1–36.

Chen, L., Lesmond, D.A. and Wei, J. (2007). Corporate Yield Spreads and Bond Liquidity. *Journal of Finance*, 62 (1): 119–149.

Christensen, K., Hounyo, U. and Podolskij, M. (2017). Is the Diurnal Pattern Sufficient to Explain the Intraday Variation in Volatility? A Nonparametric Assessment. *CREATES Research Paper, Department of Economics and Business Economics, Aarhus University*, 2017-30.

Dahlhaus, R. and Neddermeyer, J.C. (2014). Online Spot Volatility-Estimation and Decomposition with Nonlinear Market Microstructure Noise Models. *Journal of Financial Econometrics*, 12 (1): 174–212.

Delaigle, A. and Hall, P. (2013). Classification using Censored Functional Data. *Journal of the American Statistical Association*, 108 (504): 1269–1283.

Dette, H., Golosnoy, V. and Kellermann, J. (2016). The Effect of Intraday Periodicity on Realized Volatility Measures. *Discussion Paper 55/2016, SFB 823, University of Dortmund.*

Diao, X. and Tong, B. (2015). Forecasting Intraday Volatility and VaR using Multiplicative Component GARCH Model. *Applied Economics Letters*, 22 (18): 1457–1464.

Dick-Nielsen, J. (2009). Liquidity Biases in TRACE. *Journal of Fixed Income*,19 (2): 43–55.

Dick-Nielsen, J. and Feldhütter, P. and Lando, D. (2012). Corporate Bond Liquidity before and after the Onset of the Subprime Crisis. *Journal of Financial Economics*, 103 (3): 471–492.

Ding, H., Liu, Y., Xu, W. and Zhang, R. (2017). A Class of Functional Partially Linear Single-Index Models. *Journal of Multivariate Analysis*, 161: 68–82.

Durbin, J. and Koopman, S. J. (2001). Time Series Analysis by State Space Methods. *Oxford University Press.*

Murtagh, F. and Legendre, P. (2001). Predicting VNET: A model of the Dynamics of Market Depth. *Journal of Financial Markets*, 4(2): 113–142.

Engle, F.R. and Sokalska, M.E. (2012). Forecasting Intraday Volatility in the US Equity Market. Multiplicative Component GARCH. *Journal of Financial Econometrics*, 10 (1): 54–83.

Fama, E.F. and French, K.R. (1995). Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance*, 50 (1): 131–155.

Feldhütter, P. (2011). The Same Bond at Different Prices: Identifying Search Frictions and Selling Pressures. *The Review of Financial Studies*, 25(4), 1155–1206.

Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. *Springer Science & Business Media.*

Friewald, N. and Jankowitsch, R. and Subrahmanyam, M.G. (2012). Illiquidity or Credit Deterioration: A Study of Liquidity in the US Corporate Bond Market during Financial Crises. *Journal of Financial Economics*, 105 (1): 18–36.

Fu, F. (2009). Idiosyncratic Risk and the Cross-Section of Expected Stock Returns. *Journal of Financial Economics*, 91 (1): 24–37.

Gabrys, R., Hörmann, S. and Kokoszka, P. (2013). Monitoring the Intraday Volatility Pattern. *Journal of Time Series Econometrics*, 5 (2): 87–116.

Goldberg, Y., Ritov, Y. and Mandelbaum, A. (2014). Predicting the Continuation of a Function with Applications to Call Center Data. *Journal of Statistical Planning and Inference*, 147: 53–65.

Guillaumin, M., Verbeek, J. and Schmid, C. (2009). Is that You? Metric Learning Approaches for Face Identification. *In: Proceedings of the 12th International Conference on Computer Vision, IEEE*, 498–505.

Gurkaynak, R.S., Sack, B. and Wrigth, J.H. (2006). The U.S. Treasury Yield Curve: 1961 to the Present. *Finance and Economics Discussion Series, 2006-28. Federal Reserve Board.*

Hafner, C.M. and Walders, F. (2017). Heterogeneous Liquidity Effects in Corporate Bond Spreads. *The Journal of Fixed Income,* 26 (4): 73–91.

Hall, P. and Horowitz, J.L.(2007). Methodology and Convergence Rates for Functional Linear Regression. *The Annals of Statistics*, 35 (1): 70–91.

Hall, P. and Hosseini–Nasab, M.(2006). On Properties of Functional Principal Components Analysis. *Journal of the Royal Statistical Society: Series B*, 68 (1): 109–126.

Han, S. and Zhou, H. (2006). Nondefault Bond Spread and Market Trading Liquidity. *Federal Reserve Boad*.

Herskovic, B., Kelly, B., Lustig, H. and Van Nieuwerburgh, S. (2016). The common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications. *Journal of Financial Economics*, 119 (2): 249–283.

Houweling, P., Mentink, A. and Vorst, T. (2005). Comparing possible Proxies of Corporate Bond Liquidity. *Journal of Banking & Finance*, 29 (6): 1331–1358.

Horváth, Lajos and Reeder, Ron (2012). Detecting Changes in Functional Linear Models. *Journal of Multivariate Analysis*, 111: 310–334.

Horváth, Lajos and Kokoszka, Piotr (2012). Inference for Functional Data with Applications. *Springer Series in Statistics.*

Hou, K. and Loh, R.K. (2016). Have we solved the Idiosyncratic Volatility Puzzle? *Journal of Financial Economics*, 121 (1): 167–194.

Hörmann, Siegfried and Kokoszka, Piotr (2010). Weakly Dependent Functional Data. *The Annals of Statistics*, 38 (3): 1845–1884.

Hsing, T. and Eubank, R. (2015). Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. *John Wiley & Sons*.

Inglot, T. (2010). Inequalities for Quantiles of the Chi-Square Distribution. *Probability and Mathematical Statistics*, 30 (2): 339–351.

Kneip, A. and Liebl, D. (2017). On the Optimal Reconstruction of Partially Observed Functional Data. *arXiv preprint arXiv:1710.10099.*

Kneip, A., Poß, D. and Sarda, P. (2016). Functional Linear Regression with Points of Impact. *The Annals of Statistics*, 44 (1): 1–30.

Kokoszka, P., Miao, H. and Zhang, X. (2014). Functional Dynamic Factor Model for Intraday Price Curves. *Journal of Financial Econometrics*, 13 (2): 456–477.

Kong, D., Xue, K., Yao, F. and Zhang, H.H. (2016). Partially Functional Linear Regression in High Dimensions. *Biometrika*, 103 (1): 147–159.

Kraus, D. (2015). Components and Completion of Partially Observed Functional Data. *Journal of the Royal Statistical Society: Series B*, 77 (4): 777–801.

Kulis, B. (2013). Metric Learning: A Survey. *Foundations and Trends in Machine Learning*, 5 (4): 287–364.

Lian, H. (2011). Functional Partial Linear Model. *Journal of Nonparametric Statistics*, 23 (1): 115–128.

Liebl, D. and Rameseder, S. (2017). Partially Observed Functional Data: The Case of Systematically Missing Parts. *arXiv preprint arXiv:1711.07715.*

Liebl, D. and Walders, F. (2018). Parameter Regimes in Partial Functional Panel Regression. *arXiv preprint arXiv:1709.05786.*

Longstaff, F.A., Mithal, S. and Neis, E. (2005). Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market. *The Journal of Finance*, 60 (5): 2213–2253.

Lu, Y. and Du, J. and Sun, Z. (2014). Functional Partially Linear Quantile Regression Model. *Metrika*, 77 (2): 317–332.

Murtagh, F. and Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering method: Which Algorithms implement Ward's Criterion? *Journal of Classification*, 31 (3): 274–295.

Müller, H.G., Sen, R. and Stadtmüller, U. (2011). Functional Data Analysis for Volatility. *Journal of Econometrics*, 165 (2): 233–245.

Peng, Q.Y. and Zhou, J.J. and Tang, N.S. (2016). Varying Coefficient Partially Functional Linear Regression Models. *Statistical Papers*, 57 (3): 827–841.

Proietti, T. (2014). Exponential Smoothing, Long Memory and Volatility Prediction. *Centre for Economic and International Studies Tor Vergata: Research Paper Series*, 12 (7).

Ramsay, J.O. and Silverman, B.W. (2005). Functional Data Analysis. Second Edition. *Springer Series in Statistics.*

Recaredo, J. and José, R. (2015). Prediction Bands for Functional Data Based on Depth Measures. *Working paper No. 24606, Universidad Carlos III de Madrid. Departamento de Estadística.*.

Reiss, P.T. and Goldsmith, J. and Shang, H.L. and Ogden, R.T. (2016). Methods for Scalar-on-Function Regression. *International Statistical Review*, 85 (2): 228-249.

Schipper, M., Taylor, J.M.G. and Lin, X. (2008). Generalized Monotonic Functional Mixed Models with Application to Modelling Normal Tissue Complications. *Journal of the Royal Statistical Society: Series C*, 57 (2): 149–163.

Shin, H. (2009). Partial Functional Linear Regression. *Journal of Statistical Planning and Inference*, 139 (10): 3405–3418.

Shin, H. and Lee, M.H. (2012). On Prediction Rate in Partial Functional Linear Regression. *Journal of Multivariate Analysis.* 103 (1): 93–106.

Su, L. and Shi, Z. and Phillips, P.C.B. (2016). Identifying Latent Structures in Panel Data. *Econometrica*, 84 (6): 2215–2264.

Tang, Q.G. and Cheng, L.S. (2014). Partial Functional Linear Quantile Regression. *Science China Mathematics*, 57 (12): 2589–2608.

Vogt, M. and Linton, O. (2017). Classification of Non-Parametric Regression Functions in Longitudinal Data Models. *Journal of the Royal Statistical Society: Series B*, 79 (1): 5–275.

Walders, F. and Liebl, D. (2017). Parameter Regimes in Partially Functional Linear Regression for Panel Data. In: *Functional Statistics and Related Fields,* Aneiros, G., Bongiorno, E.G., Cao, R., Vieu, P. (Editors). Springer. 261–274.

Wang, G. and Feng, X.N. and Chen, M. (2016). Functional Partial Linear Single-Index Model. *Scandinavian Journal of Statistics*, 43 (1): 261–274.

Weinberger, K. Q. and Tesauro, G. (2007). Metric Learning for Kernel Regression. In: *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, 2, 612–619.

Wood, S.N. (2003). Thin Plate Regression Splines. *Journal of the Royal Statistical Society: Series B*, 65 (1): 95–114.

Wood, S.N. (2006). Generalized Additive Models: An Introduction with R. *CRC Press.*

Xiao, B., Yang, X., Xu, Y. and Zha, H. (2009). Learning Distance Metric for Regression by Semidefinite Programming with Application to Human Age Estimation. In: *Proceedings of the 17th ACM International Conference on Multimedia.*, 451–460.

Zhang, D., Lin, X. and Sowers, M.F. (2007). Two-Stage Functional Mixed Models for Evaluating the Effect of Longitudinal Covariate Profiles on a Scalar Outcome. *Biometrics*, 63 (2): 351–362.

Zhou, J. and Chen, M. (2012). Spline Estimators for Semi-Functional Linear Model. *Statistics & Probability Letters*, 82 (3): 505–513.

Zumbach, G. (2007). The Riskmetrics 2006 Methodology. *Available at SSRN: https://ssrn.com/abstract=1420185 or http://dx.doi.org/10.2139/ssrn.1420185.*