Parameter Regimes in Partial Functional Panel Regression

*Dominik Liebl
University of Bonn
Institute for Financial Economics and Statistics
dliebl@uni-bonn.de
and
Fabian Walders
University of Bonn
BGSE & Institute for Financial Economics and Statistics
fwalders@uni-bonn.de

Abstract

We propose a partial functional linear regression model for panel data with time varying parameters. The parameter vector of the multivariate model component is allowed to be completely time varying while the function-valued parameter of the functional model component is assumed to change over $K < \infty$ unknown parameter regimes. We derive consistency for the suggested estimators and for our classification procedure used to detect the K unknown parameter regimes. In addition, we derive the convergence rates of our estimators under a double asymptotic where we differentiate among different asymptotic scenarios depending on the relative order of the panel dimensions n and T. The statistical model is motivated by our real data application, where we consider the so-called "idiosyncratic volatility puzzle" using high frequency data from the S&P500.

Keywords: functional data analysis, mixed data, partial functional linear regression model, classification, idiosyncratic volatility puzzle

*We want to thank Alois Kneip (University of Bonn) and Michael Vogt (University of Bonn) for fruitful discussions and valuable comments which helped to improve this research work.

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 et al. (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 that 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 the convergence rates of our estimators under a double asymptotic where we differentiate among different asymptotic scenarios depending 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) 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&P500. 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 the idiosyncratic risk component.

The remainder of this work is structured as follows. In Sections 2 and 3 we introduce the model and present the estimation procedure. Section 4 contains our main assumptions and asymptotic results. Section 5 discusses the practical choice of the tuning parameters involved. The finite sample performance of the estimators is explored in Section 6. Section 7 offers an empirical study examining regime dependent pricing of idiosyncratic risk in the US stock market. Section 8 contains a short conclusion. All proofs can be found in the online supplement supporting this article.

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 the presence of a finite dimensional random variable $z_{it} \in \mathbb{R}^P$. Indexing the cross-section units $i = 1, \ldots, n$ and time points $t = 1, \ldots, T$, our statistical model reads as

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

where μ_t is a time fixed effect[1], $\alpha_t \in L^2([0,1])$ 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 finite but potentially time heteroscedastic variances (see also our assumptions in Section 4).

The unknown function-valued parameters α_t are assumed to differ 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

$$\alpha_t(s) \equiv A_k(s) \quad \text{if} \quad t \in G_k. \tag{2}$$

The regimes $G_1, \ldots, G_K$ are mutually exclusive subsets of $\{1, \ldots, T\}$ and do not have to consist of subsequent time points t. We have in mind the situation where G_k is a collection of time points t that belong to the kth risk regime. The kth risk regime is described by the function-valued slope parameter A_k—the vector-valued slope parameters β_t describe the marginal effects of additional control variables z_{it}. The joint and the marginal distributions of X_{it}, z_{it} and ϵ_{it} are allowed to vary over the different regimes G_k.

Model (1) nests several different model 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 $t = 1, \ldots, T$.

3 Estimation

Our objective is to estimate the model parameters A_k, β_t, and the regimes $G_1, \ldots, G_K$ from realizations of the random variables $\{(y_{it}, X_{it}, z_{it}) : 1 \leq i \leq n, 1 \leq t \leq T\}$. For this, we propose a three-step estimation procedure. The first step is a pre-estimation step where model (1) is fitted to the data *separately* for each $t = 1, \ldots, T$. This pre-estimation

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

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 $G_1, \dots, G_K$. The third step is the final estimation step, in which we improve estimation of the functional parameter A_k by employing the information 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 $t = 1, \dots, T$. Estimation starts from removing the fixed effect μ_t using a classical within-transformation. For this we denote the centered variables as $y_{it}^c = y_{it} - \bar{y}_t$, $X_{ti}^c = X_{it} - \bar{X}_t$, $z_{it}^c = z_{it} - \bar{z}_{it}$, and $\epsilon_{it}^c = \epsilon_{it} - \bar{\epsilon}_{it}$, 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) is

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

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}, i.e., $\alpha_t(s) \approx \sum_{j=1}^{m_t} \langle \alpha_t, \hat{\phi}_{jt} \rangle \hat{\phi}_{jt}(s) = \sum_{j=1}^{m_t} a_{jt} \hat{\phi}_{jt}(s)$ and $X_{it}^c(s) \approx \sum_{j=1}^{m_t} \langle X_{it}^c, \hat{\phi}_{jt} \rangle \hat{\phi}_{jt}(s)$, where $\hat{\phi}_{jt}$ denotes the eigenfunction corresponding to the jth largest eigenvalue $\hat{\lambda}_{jt}$ of the empirical covariance operator $\hat{\Gamma}_t$, with

$$\begin{aligned}
(\hat{\Gamma}_t x)(u) &:= \int_0^1 \hat{K}_{X_t}(u, v) \mathrm{d}v \quad \text{for all} \quad x \in L^2([0,1]) \\
\text{and} \quad \hat{K}_{X_t}(u, v) &:= \frac{1}{n} \sum_{i=1}^n X_{it}^c(u) X_{it}^c(v).
\end{aligned}$$

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

$$\begin{aligned}
\hat{\alpha}_t &= \sum_{j=1}^{m_t} \hat{a}_{jt} \hat{\phi}_{jt}, \quad \text{with} \quad \hat{a}_{jt} = \hat{\lambda}_{jt}^{-1} \frac{1}{n} \sum_{i=1}^n \langle X_{it}^c, \hat{\phi}_{jt} \rangle (y_{it}^c - \hat{\beta}_t^\top z_{it}^c), \quad \text{and} \\
\hat{\beta}_t &= \left[\hat{\mathbf{K}}_{z,t} - \hat{\Phi}_t(\hat{\mathbf{K}}_{zX,t}) \right]^{-1} \left[\hat{\mathbf{K}}_{zy,t} - \hat{\Phi}_t \left(\hat{K}_{yX,t} \right) \right],
\end{aligned}$$

where

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

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

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

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

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 5.

Besides computing the final estimates $\hat{\beta}_t$ of β_t, this first estimation step is intended to pave the way for the classification procedure in Step 2 where we aim to detect systematic differences in the estimates $\hat{\alpha}_t$ and $\hat{\alpha}_s$ across different time points $t \neq s$. For this it is necessary to distinguish systematically large from small differences between estimators, which can be achieved by comparing these differences to an appropriate threshold. However, the estimators $\hat{\alpha}_t$ are not well suited for deriving a practically useful threshold parameter. We, therefore, compute the following transformed estimates, for which it is straightforward to derive a valid threshold parameter using distributional arguments (see Section 5):

$$\hat{\alpha}_t^{(\Delta)} := \sum_{j=1}^{\underline{m}} \frac{\hat{\lambda}_{jt}^{1/2}}{\hat{\sigma}_{\epsilon,t}} \hat{a}_{jt}\hat{\phi}_{jt}, \tag{3}$$

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 $\hat{\alpha}_t^{\Delta}$ of (3) to classify time points $t = 1\dots,T$ into regimes $G_1, \dots, G_K$. Our classification algorithm tries 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 all $x \in L^2([0,1])$.

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:

while $|S^{(k)}| > 0$ **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 $\hat{\Delta}_{ts} \le \tau_{nT}$ **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

The algorithm stops as soon as all time points t are classified into regimes and the total number $\hat{K}$ of estimated regimes $\hat{G}_1, \ldots, \hat{G}_{\hat{K}}$ can serve as a natural estimate 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 propose to augment the classification algorithm using a $K_{\max}$ parameter where $K \leq K_{\max}$, where the practical choice of $K_{\max}$ is described in Section 5. Then the algorithm stops after at most $K_{\max} - 1$ iterations and assigns all remaining time points t to the final regime $\hat{G}_{K_{\max}}$, such that $\hat{K} \leq K_{\max}$.

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

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

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

Our final estimator of A_k is obtained, as were the pre-estimators $\hat{\alpha}_t$, according to

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

where $(\tilde{\lambda}_{jk}, \tilde{\phi}_{jk})_{j \geq 1}$ denotes the eigenvalue-eigenfunction pairs of the empirical covariance operator $\tilde{\Gamma}_k$. 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 5.

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 A_k. In the following we list our theoretical assumptions, allowing us to deal with these aspects in a large sample framework.

A1 1. $\{(X_{it}, z_{it}, \epsilon_{it}) : 1 \leq i \leq n, t \in G_k\}$ are strictly stationary and independent over the index i for any $1 \leq k \leq K$. Further, the errors ϵ_{it} are centered and independent over the index $1 \leq t \leq T$ as well.

2. For every $1 \leq k \leq K$ and $1 \leq i \leq n$, the sequence $\{X_{it} : t \in G_k\}$ is L^4-m approximable in the sense of Definition 2.1 in Hörmann and Kokoszka (2010).
3. For every $1 \leq k \leq K$ and $1 \leq i \leq n$, the sequence $\{z_{it} : t \in G_k\}$ is m-dependent.
4. Suppose that $E\left[||X_{it}||_2^4\right] < \infty$, $E\left[z_{it}^4\right] < \infty$, $E\left[\epsilon_{it}^4\right] < \infty$ for any $1 \leq i \leq n$ and $1 \leq t \leq T$,
5. ϵ_{it} is independent of X_{js} and z_{js} for any $1 \leq i, j \leq n$ and $1 \leq t, s \leq T$.

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 $k = 1, \dots, K$:

1. $C_\lambda^{-1} j^{-\mu} \leq \lambda_{jk} \leq C_\lambda j^{-\mu}$ and $\lambda_{jk} - \lambda_{j+1,k} \geq C'_\lambda j^{-(\mu+1)}$, $j \geq 1$ 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. $E\left[\langle X_{it}, \phi_{jk}\rangle^4\right] \leq C_\theta \lambda_{jk}^2$ for the eigenfunction ϕ_{jk} of Γ_k corresponding to the jth eigenvalue,
3. $|\langle A_k, \phi_{jk}\rangle| \leq C_a j^{-\nu}$ uniformly over $j \geq 1$
4. $|\langle K_{z_pX,k}, \phi_{jk}\rangle| \leq C_{zX} j^{-(\mu+\nu)}$, for any $1 \leq p \leq P$, where $K_{z_pX,k} := E[X_{it} z_{p,it}]$,
5. $\sup_{1\leq t\leq T} \beta_{p,t}^2 \leq C_\beta$, for any $1 \leq p \leq P$, with $\beta_{p,t}$ the p-th coordinate in β_t and

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

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 Suppose that the random variables

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

are iid over the index i for any given $t = 1, \dots, T$ as well as stationary, ergodic and at most m-dependent over the index t within every regime G_k, $k = 1, \dots, K$. Let $E[s_{p,it}|X_{1t}, \dots, X_{nt}] = 0$ and let $[E[s_{p,it}s_{q,it}|X_{1t}, \dots, X_{nt}]]_{1\leq p,q\leq P}$ be a positive definite matrix.

A7 1. There exists some $C_\Delta > 0$ such that for any $1 \leq k \leq K$ and any $t \in G_k$

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

where $\alpha_r^{(\Delta)} := \sigma_{\epsilon,l}^{-1} \sum_{j=1}^\infty \lambda_{jl}^{1/2} \langle \alpha_r, \phi_{jl}\rangle \phi_{jl}$ 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 $1 \leq k \leq K$.

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 our estimators. The convergence rates of the period-wise estimators $\hat{\beta}_t$ and $\hat{\alpha}_t$ of Step 1 of our estimation procedure are provided in Theorem 4.1. Lemma 4.1 establishes the uniform consistency of $\hat{\beta}_t$ over all $t = 1, \dots, T$, which is a necessary property for showing the consistency of our classification procedure in Theorem 4.2. Finally, Theorem 4.3 establishes the convergence rates of our estimator $\tilde{A}_k$ of Step 3 of our estimation procedure.

Theorem 4.1 *Given Assumptions A1–A6 it holds for all* $1 \le t \le T$ *that*

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

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

Theorem 4.1 is related to the Theorems 3.1 and 3.2 in Shin (2009), though our proof deviates from that in Shin (2009) at important instances and considers the case of panel data. The above rates for $\hat{\alpha}_t$ correspond to the rates in the cross-section context of Hall and Horowitz (2007). Theorem 4.1 is necessary to show consistency of $\tilde{A}_k$; however, it is not sufficient to establish consistency of our classification algorithm. For this, we need the following uniform consistency results:

Lemma 4.1 *Under Assumptions A1–A6, it holds that*

$$\max_{1\le t\le T} ||\hat{\beta}_t - \beta_t||^2 = o(1), \quad \max_{1\le t\le T} ||\hat{\alpha}_t - \alpha_t||_2^2 = o(1),$$
$$\text{and} \quad \max_{1\le t\le T} ||\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}||_2^2 = o(1).$$

The following theorem establishes consistency of our classification procedure and is based on our results in Lemma 4.1:

Theorem 4.2 *Given Assumptions A1–A7 it holds that*

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

The statement of Theorem 4.2 is twofold. First, it says that the number of groups K is asymptotically correctly determined. Second, it says that the estimator $\hat{G}_k$ consistently estimates the regime G_k. This notion of classification consistency is sufficient to achieve the following rates of convergence for our estimators $\tilde{A}_k$ from Step 3 of our estimation procedure:

Theorem 4.3 *Given Assumptions A1–A7 it holds for all* $1 \leq k \leq \hat{K}$ *that*

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

Theorem 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 ($\nu \geq (1+\mu+\delta)/2\delta$), 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 ($\nu \leq (1+\mu+\delta)/2\delta$), 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).

5 Practical Choice of Tuning Parameters τ_{nT}, m_t, $\tilde{m}_k$ and $K_{\max}$

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 the supplement of this paper, the scaling of the estimators $\hat{\alpha}_t$ and $\hat{\alpha}_s$ as suggested in (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 α_τ-quantile of a $\chi^2_{\underline{m}}$ distribution, where α_τ is very close to one, for instance, $\alpha_\tau = 0.99$ or $\alpha_\tau = 0.999$. By scaling this quantile with $2/n$, the threshold converges to zero as n tends to infinity. The decay of the threshold τ_{nT} is sufficiently slow in order to satisfy our Assumption 7 (see Section A.7 of the supplemental paper for more details).

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

$$m_t = \arg\max_{1 \leq l < n} \hat{\lambda}_{t,l}/\hat{\lambda}_{t,l+1}, \quad 1 \leq t \leq T$$

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

[2]See also section 2.2 in Reiss et al. (2016) and references therein.

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 $[0,1]$, the L-vectors $v_t := [\hat{\alpha}_t^{(\Delta)}(s_l)]_{l=1,\dots,L}$ are calculated for $1 \le t \le T$. We suggest the maximizer

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

as an upper bound for $\hat{K}$. Here $C_j \subset \{1,\dots,T\}$ is the jth cluster formed from a k-means algorithm with c_j being the corresponding centroid and we further denote $\bar{v} := T^{-1}\sum_{t=1}^T v_t$ and $tr(\cdot)$ the trace operator.

6 Simulations

The following simulation study considers 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 $t \in G_1 = \{1,\dots,T/2\}$ and $\alpha_t = A_2$ if $t \in G_2 = \{T/2+1,\dots,T\}$, where

$$A_1(u) = \begin{cases} \sqrt{2}\sin(\pi u/2) - u^3/2 + \sqrt{18}\sin(3\pi u/2) & \text{in Scenario 1} \\ 8u - 4u^2 - 5u^3 + 2\sin(8u) & \text{in Scenario 2} \end{cases}$$
$$A_2(u) = -2u + 8u^2 + 5u^3 + 2\sin(8u) \qquad \text{in Scenarios 1 and 2.}$$

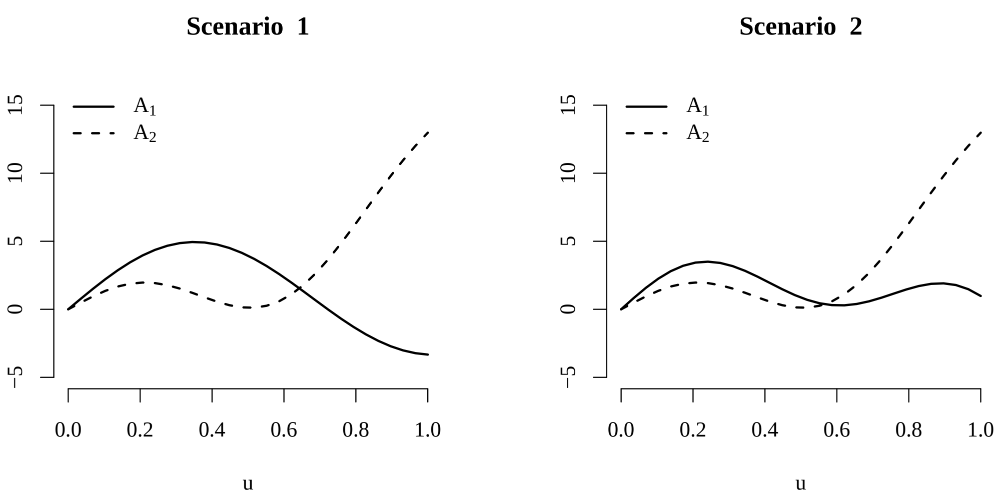

Figure 1: Regime specific parameter functions A_1 and A_2 of the two different scenarios.

The graphs of the parameter functions are shown in Figure 1. 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 $\mu_t = 5\cos(t/\pi)$. 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 $\phi_j(u) = \sqrt{2}\sin((j-1/2)\pi u)$. In order to implement the procedure we evaluate the trajectories on an equidistant grid with 30 grid points in the unit interval. For the choice of the tuning parameters, m_t, τ_{nT}, and $\tilde{m}_k$ we proceed as described in Section 5. For selecting the threshold τ_{nT} we set $\alpha = 0.99$.

$(\mathbf{n}, \mathbf{T}) = (\mathbf{50}, \mathbf{50})$	**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
$\lVert\tilde{A}_1 - A_1\rVert_2^2/\lVert A_1\rVert_2^2$	0.01	0.01	0.02	0.03	0.03	0.05	0.07	0.10	0.11	0.10
$\lVert\tilde{A}_2 - A_2\rVert_2^2/\lVert A_2\rVert_2^2$	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
$\lVert\tilde{A}_1 - A_1\rVert_2^2/\lVert A_1\rVert_2^2$	0.00	0.00	0.01	0.01	0.01	0.04	0.05	0.05	0.06	0.02
$\lVert\tilde{A}_2 - A_2\rVert_2^2/\lVert A_2\rVert_2^2$	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
$\lVert\tilde{A}_1 - A_1\rVert_2^2/\lVert A_1\rVert_2^2$	0.00	0.00	0.00	0.01	0.00	0.04	0.04	0.04	0.05	0.01
$\lVert\tilde{A}_2 - A_2\rVert_2^2/\lVert A_2\rVert_2^2$	0.01	0.02	0.03	0.03	0.04	0.01	0.02	0.02	0.02	0.02

Table 1: 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.

In order to measure the precision of classification 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. The consistency of all parameter estimates as well as the accuracy of the classification procedure is 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 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. The simulation is implemented using the statistical software package R (R Core Team, 2017) and the corresponding R-code can be obtained from the authors.

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 (see, for instance, Fu, 2009, Hou and Loh, 2016, and references therein). This finding is denoted 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 risk. Starting from theory it would thus be most reasonable to expect either no relation between idiosyncratic volatility and stock returns or a positive relation. As demonstrated in Hou and Loh (2016) this idiosyncratic volatility puzzle has, to a substantial extent, remained unsolved.

In the literature, the idiosyncratic volatility puzzle is typically considered using an aggregated monthly perspective. In contrast, we consider the cross-sectional relations between daytime returns $y_{it} \in \mathbb{R}$ of assets $i = 1, \ldots, n$ and the *non-aggregated* daily idiosyncratic volatility curves $X_{it} \in L^2([0,1])$, where the interval $[0,1]$ describes the standardized (continuous) intra-day trading time with "0" denoting the start and "1" denoting the end of the intra-day trading. Like Ang et al. (2006), who allow for time varying volatility premiums by conducting separate regressions for each considered month, we allow for daily varying volatility premiums by allowing for daily varying parameters in the following partial functional panel regression model:

$$y_{it} = \mu_t + \int_0^1 \alpha_t(s) X_{it}(s) \mathrm{d}s + \beta_t^\top z_{it} + \epsilon_{it}, \tag{4}$$

where $\mu_t \in \mathbb{R}$ is a daily fixed effect, $\alpha_t \in L^2([0,1])$ denotes the time varying parameter function describing the marginal pricing of the idiosyncratic volatility curve $X_{it} \in L^2([0,1])$ at day t, and $\beta_t \in \mathbb{R}^P$ is a time varying parameter describing the effect of additional control variables $z_{it} \in \mathbb{R}^P$. The statistical error term ϵ_{it} is a classical scalar error term with finite but potentially time heteroscedastic variances. We postulate that there are only $K < T$ different *risk 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.

Following Fu (2009), we define the dependent variable as the daytime log-returns $y_{it} := \log(P_{it}(1)/P_{it}(0))$, where $P_{it}(0)$ and $P_{it}(1)$ denote the opening and closing price of asset i at day t. As control variable $z_{it} \in \mathbb{R}$ we use the daily bid-ask spreads which serve as a proxy for liquidity risk—an important pricing-relevant factor (see Hou and Loh, 2016, and references therein).

Data-Sources. We consider high frequency data from the S&P500, where we use the $n = 377$ assets for which inta-day stock prices are recorded at 10 minute intervals over full $T = 136$ trading days between June 3, 2016, and December 15, 2016. The observed intra-day stock prices $P_{it}(s_j)$, $j = 1, \ldots, J$, are defined as the last recorded prices within standardized 10 minute intervals $[s_{j-1}, s_j]$ with $s_0 = 0$, $s_J = 1$ and equidistant interval lengths $s_j - s_{j-1} = \Delta$ such that $J\Delta = 1$, where $J = 39$. For the below described construction of the idiosyncratic volatility curves $X_{it}(.)$ we make use of the Fama-French factors. The Fama-French factors were downloaded from Kenneth French's homepage[3]; all other data were gathered from Bloomberg.

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 estimating the total volatility curve of asset i at day t via smoothing the scatter-points $(\tilde{Y}_{it,j}, s_j)$, $j = 1, \ldots, J$, where $\tilde{Y}_{it,j} := \log(\Delta^{-1} Y_{it}(s_j)^2) + q_0$ is a scaled and logarithmized version of the squared intra-day returns $Y_{it}(s_j)^2$ with $Y_{it}(s_j) := \log(P_{it}(s_j)/P_{it}(s_{j-1}))$. The constant $q_0 = 1.27$ is necessary for re-centering the involved error term (see Müller et al., 2011, for technical details). We follow their approach, but instead of using the total intra-day returns $Y_{it}(s_j)$, $j = 1, \ldots, J$, which lead to an estimate of the total volatility curve, we use only the *idiosyncratic* components $Y^*_{it}(s_j)$, $j = 1, \ldots, J$, which leads to an estimate of the idiosyncratic volatility curve $X_{it}(.)$. For computing the idiosyncratic intra-day returns $Y^*_{it}(s_j)$, we follow the usual approach and correct the total intra-day returns $Y_{it}(s_j)$ for their systematic market component by regressing them on the three Fama-French factors (see Fama and French, 1995). We do so by estimating the following functional Fama-French regression model originally proposed by Kokoszka et al. (2014):

$$Y_{it}(s_j) = b_{0,it}(s_j) + b_{1,it} \cdot M_t(s_j) + b_{2,it} \cdot S_t + b_{3,it} \cdot H_t + u_{it}(s_j), \quad j = 1, \ldots, J, \tag{5}$$

where $M_t(s_j)$ is the intra-day S&P500 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 (5) we use the least-squares estimators proposed by Kokoszka et al. (2014). The idiosyncratic intra-day returns are then defined as $Y^*_{it}(s_j) := \hat{b}_{0,it}(s_j) + \hat{u}_{it}(s_j)$, where $\hat{b}_{0,it}(s_j)$ denotes the fitted functional intercept parameter of the (i,t)th regression and $\hat{u}_{it}(s)$ are the corresponding regression residuals. Table 2 provides summary statistics for our sample. Figure 2 shows the idiosyncratic volatility curves X_{it}, along with their raw-scatter points, for the Apple stock 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 intra-day returns $Y^*_{it}(s_j)$, $j = 1, \ldots, J$, leads to "volatility curves" that have to be considered as logarithmized squared volatility curves (see Eq. (9) in Müller et al., 2011). Both transformations (taking squares and logs) are monotonic which preserves the sign of our parameter estimates for α_t. Furthermore, working with log-transformed volatility objects is generally advisable, since the original volatilities are known to be heavily skewed (see, for instance, Herskovic et al., 2016).

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

	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	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 2: Quantiles, means, and standard deviations of the considered variables.

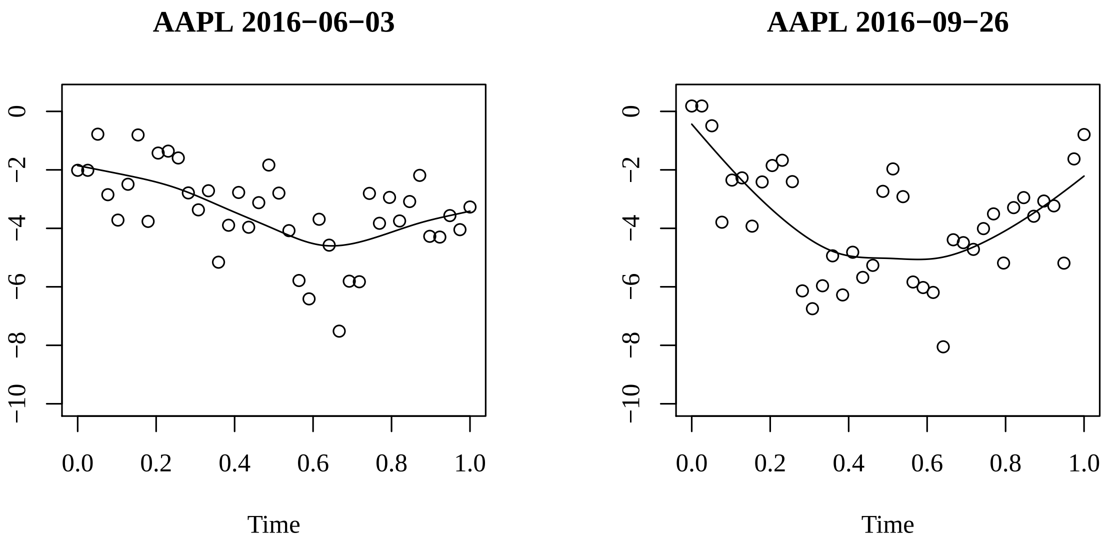

Figure 2: Idiosyncratic volatility curves X_{it} and raw scatter points for the Apple Inc. stock (AAPL) at two randomly selected trading days.

The estimation of model (4) proceeds as described in Section 3. Using our estimation 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 3). The left panel in Figure 3 shows the 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 "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.

In summary, our procedure reveals a rather complex pattern of puzzling and non-puzzling days. Hence increasing the resolution of the time scale shows a much more heterogeneous pricing of the idiosyncratic risk component than one can infer from the classical monthly perspective as, for instance, considered by Ang et al. (2006). Thus aggregating the data will misleadingly convolute the puzzling and non-puzzling pricing of idiosyncratic risk and, therefore, might contribute to the failure of current explanations of the idiosyncratic volatility puzzle (see Hou and Loh, 2016).

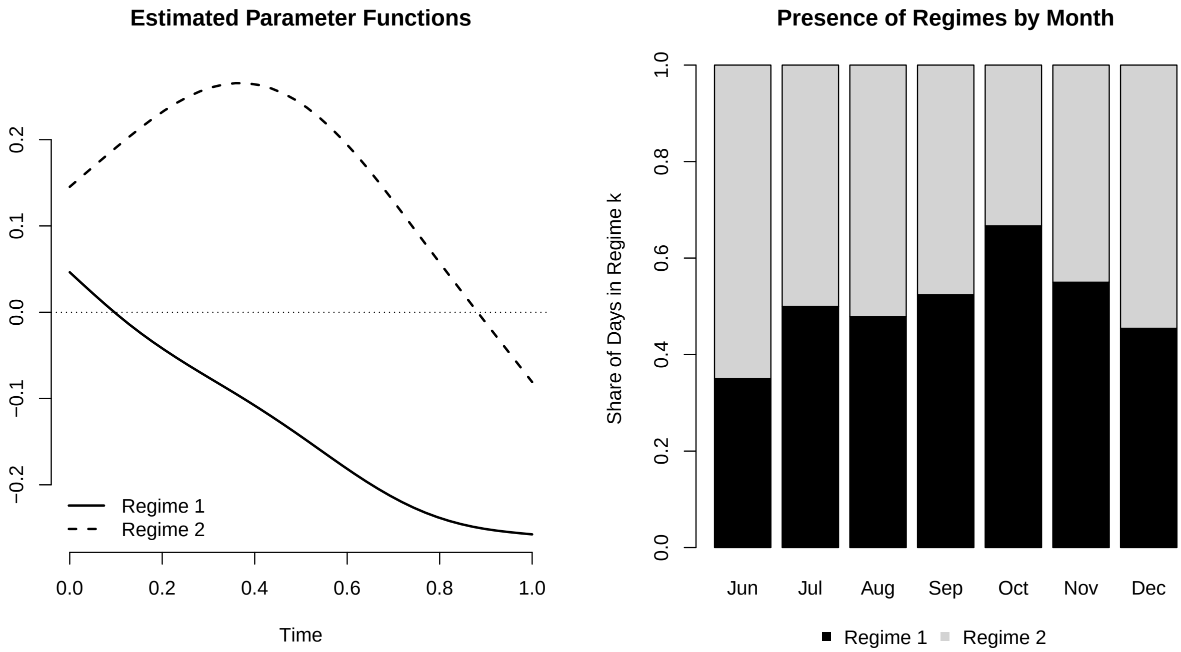

Figure 3: Estimated regime specific slope functions $\tilde{A}_1$ and $\tilde{A}_2$ (left panel) and marginal effect of idiosyncratic volatility (right panel).

8 Conclusion

In this paper 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 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 and 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 and reveals a much more heterogeneous pricing of idiosyncratic risk than suggested by the monthly analyses in the literature.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Aneiros-Pérez, G. and P. Vieu (2006). Semi-functional partial linear regression. *Statistics & Probability Letters 76*(11), 1102–1110.

Aneiros-Pérez, G. and P. Vieu (2008). Nonparametric time series prediction: A semi-functional partial linear modeling. *Journal of Multivariate Analysis 99*(5), 834–857.

Aneiros-Pérez, G. and P. Vieu (2013). Testing linearity in semi-parametric functional data analysis. *Computational Statistics 28*(2), 413–434.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *The Journal of Finance 61*(1), 259–299.

Caliński, T. and J. Harabasz (1974). A dendrite method for cluster analysis. *Communications in Statistics 3*(1), 1–27.

Ding, H., Y. Liu, W. Xu, and R. Zhang (2017). A class of functional partially linear single-index models. *Journal of Multivariate Analysis 161*, 68–82.

Fama, E. F. and K. R. French (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance 50*(1), 131–155.

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis: Theory and Practice* (1. ed.). Springer Series in Statistics. Springer.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics 91*(1), 24–37.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics 119*(2), 249–283.

Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *The Annals of Statistics 38*(3), 1845–1884.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications* (1. ed.). Springer Series in Statistics. Springer.

Horváth, L. and R. Reeder (2012). Detecting changes in functional linear models. *Journal of Multivariate Analysis 111*, 310–334.

Hou, K. and R. K. Loh (2016). Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics 121*(1), 167–194.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.

Kokoszka, P., H. Miao, and X. Zhang (2014). Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics 13*(2), 456–477.

Kong, D., K. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*(1), 147–159.

Lian, H. (2011). Functional partial linear model. *Journal of Nonparametric Statistics 23*(1), 115–128.

Lu, Y., J. Du, and Z. Sun (2014). Functional partially linear quantile regression model. *Metrika 77*(2), 317–332.

Müller, H.-G., R. Sen, and U. Stadtmüller (2011). Functional data analysis for volatility. *Journal of Econometrics 165*(2), 233–245.

Peng, Q.-Y., J.-J. Zhou, and N.-S. Tang (2016). Varying coefficient partially functional linear regression models. *Statistical Papers 57*(3), 827–841.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2. ed.). Springer Series in Statistics. Springer.

Reiss, P. T., J. Goldsmith, H. L. Shang, and R. T. Ogden (2016). Methods for scalar-on-function regression. *International Statistical Review*.

Schipper, M., J. M. Taylor, and X. Lin (2008). Generalized monotonic functional mixed models with application to modelling normal tissue complications. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 57*(2), 149–163.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference 139*(10), 3405–3418.

Shin, H. and M. H. Lee (2012). On prediction rate in partial functional linear regression. *Journal of Multivariate Analysis 103*(1), 93–106.

Su, L., Z. Shi, and P. C. Phillips (2016). Identifying latent structures in panel data. *Econometrica 84*(6), 2215–2264.

Tang, Q. and L. Cheng (2014). Partial functional linear quantile regression. *Science China Mathematics 57*(12), 2589–2608.

Vogt, M. and O. Linton (2017). Classification of non-parametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(1), 5–27.

Wang, G., X.-N. Feng, and M. Chen (2016). Functional partial linear single-index model. *Scandinavian Journal of Statistics 43*(1), 261–274.

Zhang, D., X. Lin, and M. Sowers (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 M. Chen (2012). Spline estimators for semi-functional linear model. *Statistics & Probability Letters 82*(3), 505–513.

Supplemental Paper for:

Parameter Regimes in Partial Functional Panel Regression

by Dominik Liebl and Fabian Walders

In part A of this supplement we provide formal proofs for Theorems 4.1, 4.2, and 4.3 as well as Lemma 4.1. Further, we briefly discuss the properties of the threshold τ_{nT} as suggested in Section 5. In part B, we provide additional results from our simulation study.

Throughout this supplement we use the symbols C and c to denote positive constants. Their precise meaning varies in many cases from term to term.

A Proofs

In this section we use the following notation for norms in addition to the ones introduced in the main body of the paper. Given some $f_1 \in L^2([0,1])$ and 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 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 sequence $\{(z_{it}, X_{it}) : 1 \leq i \leq n\}$ is iid with finite fourth moments for every $1 \leq t \leq T$. Moment calculations as well as the results in, e.g., Hörmann and Kokoszka (2010) imply for any $1 \leq t \leq T$ as $n \to \infty$ that

$$E\left[\left|\left|\hat{K}_{zX,t} - K_{zX,k}\right|\right|_2^2\right] = O(n^{-1}) \tag{6}$$

$$E\left[\left|\hat{K}_{z,t} - K_{z,k}\right|^2\right] = O(n^{-1}) \tag{7}$$

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

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

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 $j \geq 1$, $|\hat{\lambda}_{j,t} - \lambda_{j,k}| \leq \mathcal{D}_t$. Since $E[\mathcal{D}_t^q] = O(n^{-q/2})$ for $q = 1, 2, \ldots$ (provided sufficiently high moments exist) it holds that

$$E\left[\left|\hat{\lambda}_{j,t} - \lambda_{j,k}\right|^q\right] = O(n^{-q/2}) \;\; q = 1, 2 \ldots \tag{8}$$

for $j \geq 1$. In contrast to Shin (2009) the estimators $\hat{\beta}_t$ and $\hat{a}_{j,t}$ are calculated on the basis of empirically centered data. As, however, the arguments in Hall and Horowitz (2007) are formulated in terms of empirically centered variables as well, the argument in Shin (2009) allows to conclude immediately that for any $1 \leq t \leq T$

$$\begin{aligned} ||\hat{\Phi}_t - \Phi_k||_{H'}^2 &= \left|\left|\sum_{j=1}^{m} \frac{\langle \hat{K}_{zX,t}, \hat{\phi}_{j,t}\rangle}{\hat{\lambda}_{j,t}} \hat{\phi}_{j,t} - \sum_{j=1}^{\infty} \frac{\langle K_{zX}, \phi_j\rangle}{\lambda_j} \phi_j\right|\right|_2^2 \\ &= O_p(n^{\frac{1-2\nu}{\mu+2\nu}}), \end{aligned} \tag{9}$$

where we denote $m_t = m$ for simplicity.

A.1 Proof of Theorem 4.1

For any $1 \leq t \leq T$ the estimator $\hat{\beta}_t$ can be written as

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

with $\hat{B}_t := [\hat{K}_{z,t} - \hat{\Phi}_t(\hat{K}_{zX,t})]$. Regarding the inverse note that it follows from (6), (7), and (9) that

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

as $n \to \infty$, which certainly implies $\hat{B}_t^{-1} = B^{-1} + o_p(1)$.

As in Shin (2009), we split the term $[\hat{K}_{zy,t} - \hat{\Phi}_t(\hat{K}_{Xy,t})] = R_{0,1,t} + R_{0,2,t} + R_{0,3,t}$, while her argument allows to immediately conclude

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

However, for the remaining term we approach in a different manner:

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

with $R_{1,1,t}$ and $R_{1,2,t}$ defined as follows.

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

as a consequence of (6). The second term is defined as

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

with

$$
\begin{aligned}
R_{2,1,t} &:= |\sum_{j=m+1}^{\infty}\langle K_{zX,k}, \phi_{j,k}\rangle a_{j,k}^{*}| \\
R_{2,2,t} &:= |\sum_{j=1}^{m}\langle K_{zX,k}, \phi_{j,k}\rangle a_{j,k}^{*} - \sum_{j=1}^{m}\langle \hat{K}_{zX,t}, \hat{\phi}_{j,t}\rangle \hat{a}_{j,t}|,
\end{aligned}
$$

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

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

with

$$
R_{3,1,t} := \sum_{j=1}^{m} ||\hat{K}_{zX,t} - K_{zX,k}||_2(||\hat{\phi}_{j,t} - \phi_{j,k}||_2)||\alpha_t||_2 + a_{j,t}^{*}) = O_p(n^{-1/2}),
$$

$$
R_{3,2,t} := ||\alpha_t||_2 \sum_{j=1}^{m}\langle K_{zX,k}, \phi_{j,k}\rangle ||\hat{\phi}_{j,t} - \phi_{j,k}||_2 = O_p(n^{-1/2})
$$

and

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

[4]For $t \in G_k$, we will switch between notations $a_{j,t}^{*}$ and $a_{j,k}^{*}$ in what follows for convenience.

Bounds on $||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2$ can be obtained using the techniques in Hall and Horowitz (2007). For this purpose, define the following events:

1. $F_{1,m,t} := \{Cn^{\frac{2(1+\mu)}{\mu+2\nu}}\mathcal{D}_t^2 \leq 1/2\}$
2. $F_{2,m,t} := \{|\hat{\lambda}_{j,t} - \lambda_{l,k}|^{-2} \leq 4|\lambda_{j,k} - \lambda_{l,k}|^{-2} \leq Cn^{\frac{2(1+\mu)}{\mu+2\nu}}\}$.
3. $F_{3,m,t} := F_{1,m,t} \cap F_{2,m,t}$

Note that $\mathbb{P}(F_{1,m,t}^c) = o(1)$ as well as $\mathbb{P}(F_{2,m,t}^c) = o(1)$ due root-n consistency of the empirical covariance operator and its corresponding eigenvalues. It is shown that this property holds uniformly over $1 \leq t \leq T$ as $(n, T) \to \infty$ in the Proof of Lemma 1 below. The second event is borrowed from Hall and Horowitz (2007), while the first one allows to conclude that $\mathbb{1}(F_{1,m,t})(1 - Cn^{\frac{2(1+\mu)}{\mu+2\nu}}\mathcal{D}_t^2)^{-1} \leq 2$ for a suitably chosen constant C. This in turn implies that the bound on $||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2$ in Hall and Horowitz (2007) reads on $F_{3,m,t}$ given a suitable constant C in $F_{1,m,t}$ as follows.

$$
\begin{aligned}
&E[\mathbb{1}(F_{3,m,t})||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2] \\
&\leq 16E\left[\sum_{l:l\neq j}(\lambda_{j,k} - \lambda_{l,k})^{-2}\left[\int_0^1\int_0^1(\hat{K}_{X,t}(u,v) - K_{X,k}(u,v))\phi_{j,k}(u)\phi_{l,k}(v)\mathrm{d}u\mathrm{d}v\right]^2\right] \\
&= O(n^{-1}j^2),
\end{aligned}
$$

where $1 \leq j \leq m$. This in turn allows to conclude for the first summand in $R_{3,3,t}$,

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

due to Assumption 4. The second summand in $R_{3,3}$ behaves according to

$$
\mathbb{1}(F_{3,m,t})||K_{zX,k}||_2\sum_{j=1}^m||\hat{\phi}_{j,t} - \phi_{j,k}||_2|a_{j,t}^*| = O_p(n^{-1/2})
$$

again using Assumption 4. Since $\mathbb{P}(F_{3,m,t}^c) \leq \mathbb{P}(F_{1,m,t}^c) + \mathbb{P}(F_{2,m,t}^c) = o(1)$ as claimed before, it follows $R_{3,3,t} = O_p(n^{-1/2})$. Combining arguments implies $\hat{\beta}_t - \beta_t = O_p(n^{-1/2})$ for every $1 \leq t \leq T$, which concludes the Proof the first result in Theorem 4.1. Turning to $\hat{\alpha}_t$ note that

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

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 $O_p(n^{\frac{1-2\nu}{\mu+2\nu}})$. The remaining term can be split according to

$$\sum_{j=1}^{m}(\hat{a}_{j,t}-a^*_{j,t})^2 \leq 2\sum_{j=1}^{m}(\hat{\lambda}_{j,t}\langle\hat{K}^{\#}_{yX,t}-\hat{\beta}_t\hat{K}^{\#}_{zX,t},\hat{\phi}_{j,t}\rangle-a^*_{j,t})^2+2\sum_{j=1}^{m}(\hat{\lambda}_{j,t}\langle r_{y,t}r_{x,t}-\hat{\beta}_t r_{z,t}r_{X,t},\hat{\phi}_{j,t}\rangle)^2 \tag{10}$$

with $K^{\#}_{yX,t}:=n^{-1}\sum_{i=1}^{n}(y_{it}-E[y_{it}])(X_{it}-E[X_{it}])$, $K^{\#}_{zX,t}:=n^{-1}\sum_{i=1}^{n}(z_{it}-E[z_{it}])(X_{it}-E[X_{it}])$, $r_{x,t}:=E[X_{it}]-\bar{X}_t$, $r_{y,t}:=E[y_{it}]-\bar{y}_t$ and $r_{z,t}:=E[z_{it}]-\bar{z}_t$. Note that, $||r_{x,t}||_2, r_{y,t}$ and $r_{z,t}$ are all errors from parametric estimation problems and thus of order $n^{-1/2}$. Bounds on $\hat{\lambda}_{j,t}-\lambda_{j,k}$ as well as $||\hat{\phi}_{j,t}-\phi_{j,k}||$ are asymptotically equivalent for data centered around their arithmetic mean and data centered around their population expectation. Together with the above arguments showing $\hat{\beta}_t-\beta_t=O_p(n^{-1/2})$, it follows that the first term in (10) is asymptotically equivalent to the corresponding term in Shin (2009), implying $\sum_{j=1}^{m}(\hat{\lambda}_{j,t}\langle\hat{K}^{\#}_{yX,t}-\hat{\beta}_t\hat{K}^{\#}_{zX,t},\hat{\phi}_{j,t}\rangle-a^*_{j,t})^2=O_p(n^{\frac{1-2\nu}{\mu+2\nu}})$. Now, define the event

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

on which the second term in (10) behaves according to

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

since $\mathbb{P}(F^c_{4,m,t})=o(1)\ \forall\ t$ as $n\to\infty$ as a consequence of (8). 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 4.1. ■

A.2 Proof of Lemma 4.1

In what follows we ultimately show that the quantities $\hat{\alpha}^{\Delta}_t$ are consistent for α^{Δ}_t in the L^2 norm, uniformly over $1\leq t\leq T$. This in turn implies classification consistency as will be shown in the proof of Theorem 4.2. The remaining parts of the Lemma are shown on the way, as they are required to obtain the result concerning $\hat{\alpha}^{\Delta}_t$.

We begin 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. We also use the results in Hall and Horowitz (2007) and Hörmann and Kokoszka (2010).

- Note that since $\{X_{it}:\ t\in G_k,\ 1\leq i\leq n\}$ is stationary, $E[\mathcal{D}^2_t]$ does not vary in the time index t. Using the results in Hörmann and Kokoszka (2010) we can hence conclude

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

- Empirical variances of z_{it} and ϵ_{it} behave according to

$$\begin{aligned}&\mathbb{P}\left(\max_{1\le t\le T}|\hat{K}_{z,t} - K_{z,k}|^2 > c\right)\\ &\le K\max_{1\le k\le K}|G_k|\left[\frac{n^{-1}E\left[|(z_{it}-E[z_{it}])^2 - K_{z,k}|^2\right]}{c} + \frac{E\left[|(\bar{z}_t - E[z_{it}])^4\right]}{c}\right]\\ &= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1)\end{aligned} \tag{11}$$

and similarly

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

- The empirical covariances between regressors and error in any cross-section are denoted $\hat{K}_{X\epsilon,t} := n^{-1}\sum_{i=1}^{n}X_{it}^c\epsilon_{it}^c$ and $\hat{K}_{z\epsilon,t} := n^{-1}\sum_{i=1}^{n}z_{it}^c\epsilon_{it}^c$. 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 $\hat{K}_{\epsilon X,t}^{\#}(u) := n^{-1}\sum_{i=1}^{n}(X_{it}(u) - E[X_{it}](u))\epsilon_{it}$ as well as $r_{\epsilon,t} := \bar{\epsilon}_t$. It follows from simple moment calculations that

$$\begin{aligned}&\mathbb{P}(\max_{1\le t\le T}||\hat{K}_{X\epsilon,t}||_2^2 > c) \le \sum_{k=1}^{K}\sum_{t\in G_k}\mathbb{P}\left(||\hat{K}_{X\epsilon,t}^{\#}||_2^2 > c/4\right) + \sum_{k=1}^{K}\sum_{t\in G_k}\mathbb{P}\left(||r_{X,t}||_2^2 r_{\epsilon,t}^2 > c/4\right)\\ &\le K\max_{1\le k\le K}|G_k|\left[\frac{n^{-1}\sigma_\epsilon^2 E\left[||X_{it}-E[X_{it}]||_2^2\right]}{c} + \frac{E[(\bar{\epsilon}_t)^2]E\left[||\bar{X}_t - E[X_{it}]||_2^2\right]}{c}\right]\\ &= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1)\end{aligned} \tag{12}$$

Similar arguments can be used to show

$$\mathbb{P}(\max_{1\le t\le T}|\hat{K}_{z\epsilon,t}|^2 > c) = O(n^{\delta-1}) + O(n^{\delta-2}) = o(1). \tag{13}$$

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

$$
\begin{aligned}
&\mathbb{P}(\max_{1\le t\le T} ||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c) \\
\le& \sum_{k=1}^{K}\sum_{t\in G_k} \mathbb{P}\left(||\hat{K}^{\#}_{zX,t} - K_{zX,k}||_2^2 > c/4\right) + \sum_{k=1}^{K}\sum_{t\in G_k} \mathbb{P}\left(||r_{X,t}||_2^2 r_{z,t}^2 > c/4\right) \\
\le& K \max_{1\le k\le K} |G_k| \frac{n^{-1}E\left[||(z_{it}-E[z_{it}])(X_{it}-E[X_{it}]) - K_{zX,k}||_2^2\right]}{c} \\
&+ K \max_{1\le k\le K} |G_k| \frac{E\left[|\bar{z}_t - E[z_{it}]|^2\right] E\left[||\bar{X}_t - E[X_{it}]||_2^2\right]}{c} \\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1).
\end{aligned} \tag{14}
$$

For the following proof it will be necessary to obtain uniform consistency of $\sum_{j=1}^{m} \lambda_{j,k}^{-2} ||\phi_{j,k} - \hat{\phi}_{j,t}||_2^2$. For this purpose observe that $\hat{\lambda}_{j,t} \ge \lambda_{j,t}/2$ for $1 \le j \le m$ on on $\mathcal{F}_{4,m,t}$. Further, provided that $\mathcal{F}_{3,m,t}$ holds for any $1 \le t \le T$, observe that $E[||\hat{\phi}_{j,t} - \phi_{j,t}||_2^2] = O(n^{-1}j^2)$ uniformly on $1 \le j \le m$. Note that the results in Hall and Horowitz (2007) imply that uniformly on j and l

$$
nE\left[\left[\int_0^1\int_0^1 (\hat{K}_{X,t}(u,v) - K_{X,k})(u,v))\phi_{j,k}(u)\phi_{l,k}(v)\mathrm{d}u\mathrm{d}v\right]^2\right] \le \lambda_{j,k}\lambda_{l,k}
$$

as well as

$$
E\left[\sum_{l:l\ne j}(\lambda_{j,k}-\lambda_{l,k})^{-2}\left[\int_0^1\int_0^1 (\hat{K}_{X,t}(u,v) - K_{X,k}(u,v))\phi_{j,k}(u)\phi_{l,k}(v)\mathrm{d}u\mathrm{d}v\right]^2\right] = O(n^{-1}j^2),
$$

uniformly on $1 \le j \le m$. Further observe that

$$
\begin{aligned}
&E[\mathbb{1}(F_{3,m,t})||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2] \\
\le& 16E\left[\sum_{l:l\ne j}(\lambda_{j,k}-\lambda_{l,k})^{-2}\left[\int_0^1\int_0^1 (\hat{K}_{X,t}(u,v) - K_{X,k}(u,v))\phi_{j,k}(u)\phi_{l,k}(v)\mathrm{d}u\mathrm{d}v\right]^2\right] \\
=& O(n^{-1}j^2).
\end{aligned}
$$

In order to proceed introduce the event

- $\mathcal{F}_{5,m,t} := \{|\hat{\lambda}_{j,t} - \lambda_{j,k}| \ge \frac{1}{4}|\lambda_{j,k} - \lambda_{l,k}| : \ 1 \le j \le m\}$.

Observe for the events $\mathcal{F}_{1,m,t}, \mathcal{F}_{2,m,t}, \mathcal{F}_{4,m,t}$ and $\mathcal{F}_{5,m,t}$ the following. Note for $K\max_{1\le k\le K}|G_k|\mathbb{P}(F_{1,m,t}^c)$, that

$$
\begin{aligned}
K \max_{1\le k\le K} |G_k| \mathbb{P}(Cn^{\frac{2(1+\mu)}{\mu+2\nu}}\mathcal{D}_t^2 > 1/2) &\le K \max_{1\le k\le K} |G_k| \frac{n^{\frac{2(1+\mu)}{\mu+2\nu}} E[\mathcal{D}_t^2]}{2C} \\
&= O(n^{\frac{2+(1+\delta)\mu - 2(1-\delta)\nu}{\mu+2\nu}})
\end{aligned}
$$

for any $C > 0$. Further it holds that $K \max_{1\leq k\leq K} |G_k|\mathbb{P}(F_{1,m,t}^c) = o(1)$ as $(n,T) \to \infty$.

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

by the fact that $\mathcal{D}_t \geq |\hat{\lambda}_{j,t} - \lambda_{j,k}|$ and Assumptions 2 and 4. Beyond that it holds

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

Combining arguments leads to

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\leq t\leq T} \sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-2}||\phi_{j,k} - \hat{\phi}_{j,t}||_2^2 > c\right) \\
\leq& K \max_{1\leq k\leq K} |G_k|\frac{\sum_{j=1}^{m} \lambda_{j,k}^{-2}E[\mathbb{1}(F_{3,m,t})||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2]}{C} + K \max_{1\leq k\leq K} |G_k|(\mathbb{P}(F_{1,m,t}^c) + \mathbb{P}(F_{2,m,t}^c) + \mathbb{P}(F_{4,m,t}^c)) \\
&= O(n^{\frac{3+(1+\delta)\mu-2(1-\delta)\nu}{\mu+2\nu}}) + o(1) \\
&= o(1) \qquad (15)
\end{aligned}
$$

Note that this result 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. For later purpose, define the event $\mathcal{S}_t$ according to

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

It will be shown in a moment that also $\sum_{t=1}^{T}\mathbb{P}(\mathcal{S}_t^c)=o(1)$. However this requires some preparation since $\hat{\sigma}^2_{\epsilon,t}$ includes estimation errors from $\hat{\beta}_t$ and $\hat{\alpha}_t$. We thus begin by showing $\mathbb{P}(\max_{1\le t\le T}|\hat{\beta}_t-\beta_t|>c)=o(1)$. The estimator $\hat{\beta}_t$ makes multiple use of the operator $\hat{\Phi}_t$, which can, starting from the Riesz-Frechet representation Theorem (see 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 $R_{4,3}:=\left|\left|\sum_{j=m+1}^{\infty}\frac{\langle K_{zX,k},\phi_{j,k}\rangle}{\lambda_{j,k}}\phi_{j,k}\right|\right|^2$, 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 treated as follows. The first one behaves according to

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

The three summands behave as follows:

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

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

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

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

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

Since $\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(R_{4,1,t}>c\right)\leq\sum_{l=1}^3\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(R_{5,l,t}>c/3\right)$, it follows that $\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(R_{4,1}>c\right)=o(1)$. For $R_{4,2,t}$ note that due to Assumption 4 it holds that

$$\begin{aligned}
&K\max_{1\leq k\leq K}|G_k|\mathbb{P}\left(\left|\left|\sum_{j=1}^{m}\frac{\langle K_{zX,k},\phi_{j,k}\rangle}{\lambda_j}(\phi_{j,k}-\hat{\phi}_{j,t})\right|\right|_2^2>c\right)\\
\leq&K\max_{1\leq k\leq K}|G_k|\sum_{k=1}^{K}\sum_{t\in G_k}\mathbb{P}\left(\sum_{j=1}^{m}\frac{\langle K_{zX,k},\phi_{j,k}\rangle}{\lambda_j}\left|\left|\phi_{j,k}-\hat{\phi}_{j,t}\right|\right|_2^2\mathbb{1}(\mathcal{F}_{3,m,t})>c\right)+\sum_{k=1}^{K}\sum_{t\in G_k}\mathbb{P}(\mathcal{F}_{3,m,t}^c)\\
=&O\left(mn^{\delta-1}\right)+o(1)\\
=&o(1).
\end{aligned}$$

These results imply

$$\mathbb{P}\left(\max_{1\leq k\leq K}\max_{t\in G_k}||\hat{\Phi}_t-\Phi_k||^2_{H'}>c\right)=o(1). \tag{16}$$

Addressing the inverse in $\hat{\beta}_t$, define the event $Q_t:=\{|\hat{B}_t-B_k|\leq\frac{1}{2}B_k\}$, where $\hat{B}_t:=\hat{K}_{z,t}-\hat{\Phi}_t(\hat{K}_{zX,t})$ and its population counterpart $B_k:=K_{z,k}-\Phi_k(K_{zX,k})>0$. For this event, note that $\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(Q_t^c\right)\leq R_{6,1,t}+R_{6,2,t}$, where $R_{6,1,t}:=\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(|\hat{K}_{z,t}-K_{z,k}|^2>c\right)=o(1)$ as shown in (11).

To proceed we work again on the restated difference $\hat{\beta}_t-\beta_t$,

$$(\hat{\beta}_t-\beta_t)=\hat{B}_t^{-1}\left(R_{0,1,t}+R_{0,2,t}+R_{0,3,t}\right),$$

with $R_{0,1,t}$, $R_{0,2,t}$ and $R_{0,3,t}$ from before. Further introduce the event $Q_{t,k}:=\{|\hat{B}_t-B_k|\leq\frac{1}{2}B_k\}$. For this event, note that $\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(Q_t^c\right)\leq R_{2,1}+R_{2,2,t}$, where $R_{2,1,t}:=\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(|\hat{K}_{z,t}-K_{z,k}|^2>c\right)=o(1)$ as shown in (11). For the second term For $R_{6,2,t}$ note

$$\begin{aligned}
R_{6,2,t}:=&\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(||\hat{\Phi}_t-\Phi_k||^2_{H'}||K_{zX,k}||_2^2+(||\hat{\Phi}_t-\Phi_k||_{H'}+||\Phi_k||_{H'})^2||\hat{K}_{zX,t}-K_{zX,k}||_2^2>c\right)\\
\leq&\underbrace{\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(||\hat{\Phi}_t-\Phi_k||^2_{H'}||K_{zX,k}||_2^2>c\right)}_{=:R_{7,1,t}}+\underbrace{\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(||\hat{\Phi}_t-\Phi_k||_{H'}||^2||\hat{K}_{zX,t}-K_{zX,k}||_2^2>c\right)}_{=:R_{7,2,t}}\\
&+\underbrace{\sum_{k=1}^K\sum_{t\in G_k}\mathbb{P}\left(||\Phi_k||_{H'}||^2||\hat{K}_{zX,t}-K_{zX,k}||_2^2>c\right)}_{=:R_{7,3,t}}.
\end{aligned}$$

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

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

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 purpose consider

$$\mathbb{P}\left(\max_{1\leq t\leq T}|n^{-1}\sum_{i=1}^{n}(z_{it}^{c}-\Phi(X_{it}^{c}))\epsilon_{it}^{c}|>c\right)\leq K\max_{1\leq k\leq K}\left(\mathbb{P}(|\hat{K}_{z\epsilon,t}|>c/2)+\mathbb{P}(||\Phi_k||_{H'}|\hat{K}_{\epsilon X,t}|>c/2)\right)$$
$$=o(1)$$

due to (12) and (13). Further note for $R_{0,2,t}$

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

as a consequence of (12) and (16). For the remaining terms, argue along the same lines as in the Proof of Lemma 1:

$$\mathbb{P}(\max_{1\leq t\leq T}|R_{0,3,t}|>c):=\mathbb{P}(\max_{1\leq t\leq T}|R_{1,1,t}|>c)+\mathbb{P}(\max_{1\leq t\leq T}|R_{1,2,t}|>c)$$

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

because of (14). As before the remaining term can be bounded according to $R_{1,2,t}\leq R_{2,1,t}+R_{2,2,t}$, where the two summands are defined above. While $R_{2,1,t}=R_{2,1}=O(n^{-1/2})$ deterministically and independent of t, note for the second bound $R_{2,2,t}\leq R_{3,1,t}+R_{3,2,t}+R_{3,3,t}$ as before and further:

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

Further for $R_{3,2,t}$ note:

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

Using similar arguments, one obtains for $R_{3,3,t}$:

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

Hence $\mathbb{P}(\max_{1\leq t\leq T}(\hat{\beta}_t-\beta_t)^2>c)=o(1)$ as claimed in the Lemma. Now, turning to the estimation error in the functional parameter estimates $\hat{\alpha}_t$ observe that

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

This is because due to Assumption 2, $\sum_{j=m+1}^{\infty}{a^*_{j,k}}^2$ is a null sequence and hence arbitrarily small for sufficiently large n. The terms $R_{8,1,t}-R_{8,4,t}$ are defined and treated as follows.

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

due to Assumption 4.

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

Because of the above results and the fact that $\sum_{j=1}^m \lambda_{j,k}{}^{-2} \langle K_{zX,k}, \phi_j \rangle^2 = O(1)$ by Assumption 2, the first two summands are $o(1)$. As a consequence of (15) the third summand is $o(1)$. Together with the above results this implies $R_{8,2,t} = o(1)$. Further, with $a_{j,t} := \langle \alpha_t, \hat{\phi}_{j,t} \rangle$, define

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

by (15) and

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

due to Assumption 4. 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 $||\hat{\alpha}_t - \hat{\alpha}_s||_2^2$. As, however 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 $\mathcal{S}_t$, for which

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

where $\tilde{r}_{it} := z_{it}^c(\beta_t - \hat{\beta}_t) + \langle X_{it}^c, \alpha_t - \hat{\alpha}_t\rangle$ and $R_{9,1,t} - R_{9,3,t}$ as follows.

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

For $R_{9,2,t}$ note:

$$\begin{aligned} R_{9,2,t} := & K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left(\epsilon_{it}^c \tilde{r}_{it} \right) \right| > c \right) \\ \le & R_{10,1,t} + R_{10,2,t} + R_{10,3,t} \end{aligned}$$

with $R_{10,1,t} - R_{10,3,t}$ as follows.

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

by (13) and the above results. Further

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

by the above results and again the iid sampling in the cross-section dimension as well as independence of X_{it} and ϵ_{it}. Note for $R_{9,3,t}$ the following.

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

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

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

by (11) and the above results. Further it holds that

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

in light of the above results. Combining results yields $\sum_{k=1}^{K}\sum_{t\in G_k} \mathbb{P}\left(\mathcal{S}_t^c\right) = o(1)$. Now, finally turning to $\hat{\alpha}_t^{(\Delta)}$, for sufficiently large n

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

with

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

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

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

$$R_{12,1,t} \leq R_{13,1,t} + R_{13,2,t}$$

where, for some suitable $c > 0$,

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

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

These terms in turn behave as follows.

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

All summands are $o(1)$ using the same arguments as before. The above arguments also imply

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

Now turning to $R_{12,2,t}$ note that

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

where

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

and

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

Note for $R_{14,1,t}$:

$$R_{14,1,t} \leq R_{15,1,t} + R_{15,2,t} + R_{15,3,t} + \mathbb{P}(\mathcal{S}_t^c)$$

with

$$R_{15,1,t} := K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} |a^*_{j,t}| \hat\lambda_{j,t}^{1/2} |\sigma_{\epsilon,k} - \hat\sigma_{\epsilon,t}| > c\right)$$

$$R_{15,2,t} := K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} |a^*_{j,t}| \hat\lambda_{j,t}^{1/2} \hat\sigma_{\epsilon,t} ||\phi_{j,k} - \hat\phi_{j,t}||_2 > c\right)$$

$$R_{15,3,t} := K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} |a^*_{j,t}| |\hat\lambda_{j,t}^{1/2} - \lambda_{j,k}^{1/2}| \hat\sigma_{\epsilon,t} > c\right).$$

In order to continue, note that by the mean value Theorem

- it holds on $\mathcal{S}_t$ that $|\hat\sigma_{\epsilon,t} - \sigma_{\epsilon,k}| \le \frac{1}{2}(\sigma_{\epsilon,k}/2)^{-\frac{1}{2}} |\hat\sigma^2_{\epsilon,t} - \sigma^2_{\epsilon,k}|$ and
- it holds on $\mathcal{F}_{4,m,t}$ that $|\hat\lambda_{j,t}^{1/2} - \lambda_{j,k}^{1/2}| \le \frac{1}{2}(\lambda_{j,k}/2)^{-\frac{1}{2}} |\hat\lambda_{j,t} - \lambda_{j,k}|$.

$$\begin{aligned} R_{15,1,t} &:= K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} |a^*_{j,t}| \hat\lambda_{j,t}^{1/2} |\sigma_{\epsilon,k} - \hat\sigma_{\epsilon,t}| > c\right) \\ &\le K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} |a^*_{j,t}| \lambda_{j,k}^{1/2} |\hat\sigma^2_{\epsilon,t} - \sigma^2_{\epsilon,k}| > c\right) + K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{F}^c_{4,m,t}\right) + K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{S}^c_t\right) \\ &= o(1) \end{aligned}$$

Following the argument from before (clearly $\sum_{j=1}^m |a^*_{j,k}| \lambda_{j,k}^{1/2} = O(1)$ by Assumptions 2 and 4).

$$\begin{aligned} R_{15,2,t} &\le K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} |a^*_{j,t}| \lambda_{j,k}^{1/2} ||\phi_{j,k} - \hat\phi_{j,t}||_2 > c\right) \\ &\le K \max_{1\le k\le K} |G_k| \frac{\sum_{j=1}^m |a^*_{j,t}|^2 \lambda_{j,k} E[||\phi_{j,k} - \hat\phi_{j,t}||_2^2 \mathbb{1}(\mathcal{F}_{3,m,t})]}{c} + K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{F}^c_{3,m,t}\right) \\ &= O(mn^{\delta-1}) + o(1) = o(1) \end{aligned}$$

given δ as in Assumption 3.

$$\begin{aligned} R_{15,3,t} &\le K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} |a^*_{j,t}| \lambda_{j,k}^{-1/2} |\hat\lambda_{j,t} - \lambda_{j,k}| > c\right) \\ &\le K \max_{1\le k\le K} |G_k| \frac{\sum_{j=1}^m |a^*_{j,t}| \lambda_{j,k}^{-1/2} E[|\hat\lambda_{j,t} - \lambda_{j,k}|]}{c} + K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{F}^c_{4,m,t}\right) + K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{S}^c_t\right) \\ &= o(1) \end{aligned}$$

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

$$
\begin{aligned}
R_{14,2,t} &= K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\sum_{j=1}^{m} (a^*_{j,t} - a_{j,t})^2 \frac{\hat{\lambda}_{j,t}}{\hat{\sigma}^2_{\epsilon,t}} > c \right) \\
&\le K \max_{1\le k\le K} |G_k| \frac{\sum_{j=1}^{m} ||\alpha_t||_2^2 E[||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 \mathbb{1}(\mathcal{F}_{3,m,t})] \frac{\lambda_j^2}{\sigma_\epsilon^2}}{c} + K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{F}^c_{4,m,t}\right) \\
&+ K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{S}^c_t\right) + K \max_{1\le k\le K} |G_k| \mathbb{P}\left(\mathcal{F}^c_{3,m,t}\right) = o(1).
\end{aligned}
$$

Combining arguments implies the last part of the Lemma. ■

A.3 Proof of Theorem 4.2

Using the results presented in the previous Lemma it is possible to argue in analogy to the proof of Theorem 3.1 in Vogt and Linton (2017) to proof classification consistency in the sense of our Theorem 4.2. For this purpose consider the set $S^{(j)} = \{1, \dots, T\} \setminus \bigcup_{l<j} \hat{G}_l$ at an iteration step $1 \le j \le \hat{K} - 1$ of the algorithm described in Section 3. For a $t \in S^{(j)}$ denote the set of indexes corresponding to the ordered distances $\hat{\Delta}_{t(1)} \le \dots \le \hat{\Delta}_{t(|S^{(j)}|)}$ as $\{(1), \dots, (|S^{(j)}|)\}$. In analogy, the index set corresponding ordered population distances $\Delta_{t[1]} \le \dots \le \Delta_{t[|S^{(j)}|]}$ is denoted as $\{[1], \dots, [|S^{(j)}|]\}$, where Δ_{ts} is as in Assumption 7. Now, define an index $\hat{\kappa}$ according to $\hat{\Delta}_{t(\hat{\kappa})} < \tau_{nT} < \hat{\Delta}_{t(\hat{\kappa}+1)}$. Beyond that a population counterpart, κ, obtains as $0 = \Delta_{t[\kappa]} < \tau_{nT} < \Delta_{t[\kappa+1]}$. It holds that

$$
\begin{aligned}
\mathbb{P}\left(\{(1), \dots, (\hat{\kappa})\} \neq \{([1], \dots, [\kappa]\}\right) &\le \mathbb{P}\left(\{(1), \dots, (\kappa)\} \neq \{[1], \dots, [\kappa]\}\right) + \mathbb{P}(\hat{\kappa} \neq \kappa) \qquad (17) \\
&= o(1) + o(1) = o(1).
\end{aligned}
$$

In order to prove that the first probability on the right hand side of (17) is a null sequence, suppose that $t \in G_k$, with $1 \le k \le K$. As indicated, there are $\kappa \ge 1$ indexes in $S^{(j)}$ being elements of G_k. For the corresponding distances to reference time t it holds that $\Delta_{t[1]} = \dots = \Delta_{t[\kappa]} = 0$ by definition. The remaining distances are bounded away from zero by $0 < C_\Delta \le \Delta_{t[\kappa+1]} \le \dots \le \Delta_{t[|S^{(j)}|]}$ due to Assumption 7.
As shown in the above Lemma, $\max_{1\le t\le T} ||\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}||_2^2 = o_p(1)$ implying that $\max_{1\le s\le T} |\hat{\Delta}_{ts} - \Delta_{ts}| = o_p(1)$ which holds for any reference period t. Combining arguments allows to conclude $\max_{1\le (s)\le \kappa} \hat{\Delta}_{t(s)} = o_p(1)$ and $\min_{\kappa<(s)\le|S^{(j)}|} \hat{\Delta}_{t(s)} \ge C_\Delta + o_p(1)$ as well as $\max_{1\le [s]\le\kappa} \hat{\Delta}_{t[s]} = o_p(1)$ and $\min_{\kappa<[s]\le|S^{(j)}|} \hat{\Delta}_{t[s]} \ge C_\Delta + o_p(1)$. This immediately implies that the first probability on the right hand side of (17) 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 $n \to \infty$. 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 (17) is a null sequence. ■

A.4 Remark 1

As a consequence of Theorem 4.2, the classification error is, in what follows, asymptotically negligible. To see this note that an analogous argument as in Vogt and Linton (2017) holds in our context: let $s_1(n,T)$ and $s_2(n,T)$ be two arbitrary sequences such that $s_j(n,T) \to 0$ as $n,T \to \infty$ for $j=1,2$. Now, note that for any constants $M_1, M_2 > 0$

$$\begin{aligned}
&\mathbb{P}\left(s_1(n,T)\sum_{t\in\hat{G}_k}(\tilde{\beta}_{t,k}-\beta_t)^2 > M_1\right)\\
\leq&\mathbb{P}\left(\left\{s_1(n,T)\sum_{t\in\hat{G}_k}(\tilde{\beta}_{t,k}-\beta_t)^2 > M_1\right\}\cap\left\{\hat{G}_k = G_k\right\}\right)+\mathbb{P}\left(\left\{\hat{G}_k \neq G_k\right\}\right)\\
=&\mathbb{P}\left(s_1(n,T)\sum_{t\in G_k}(\tilde{\beta}^*_{t,k}-\beta_t)^2 > M_1\right)+o(1).
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\left(s_2(n,T)||\tilde{A}_k - A_k||_2^2 > M_2\right) &\leq \mathbb{P}\left(\left\{s_2(n,T)||\tilde{A}_k - A_k||_2^2 > M_2\right\}\cap\left\{\hat{G}_k = G_k\right\}\right)+\mathbb{P}\left(\left\{\hat{G}_k \neq G_k\right\}\right)\\
&= \mathbb{P}\left(s_2(n,T)||\tilde{A}^*_k - A_k||_2^2 > M_2\right)+o(1).
\end{aligned}$$

The quantities $\tilde{\beta}^*_{t,k}$ and $\tilde{A}^*_k$ are the estimators $\tilde{\beta}_{t,k}$ and $\tilde{A}_k$ calculated from $\{(y_{it}, X_{it}, z_{it}) : 1 \leq i \leq n,\ t \in G_k\}$, i.e. without classification errors. Note in particular that the dependence structure formulated in our assumptions does not disturb this argument.

In light of this remark, the proof of Theorem 4.3 starts from the quantities $\tilde{\beta}^*_{t,k}$ and $\tilde{A}^*_k$ rather than their contaminated counterparts.

A.5 Remark 2

Denote as $\tilde{\phi}^*_{j,k}, \tilde{\lambda}^*_{j,k}, \tilde{K}^*_{X,k}$ the estimators $\tilde{\phi}_{j,k}, \tilde{\lambda}_{j,k}, \tilde{K}_{X,k}$ from the observations $\{(z_{it}, X_{it}) : 1 \leq i \leq n,\ t \in G_k\}$). In analogy interpret $\tilde{\Phi}^*_k, \tilde{K}^*_{zX,k}$ and $\tilde{K}^*_{z,k}$ as the estimates $\tilde{\Phi}_k, \tilde{K}_{zX,k}$ and $\tilde{K}_{z,k}$ without classification error.

Note, that due to Assumption 1 for every regime G_k, the sequence $\{X_{it} :\ 1 \leq i \leq n,\ t \in G_k\}$ is L^4_m-approximable. Thus the following inequalities from Hörmann and Kokoszka (2010) hold (where for the third inequality we used our Assumption 2 already).

$$E\left[\left|\left|\tilde{K}^*_{X,k} - K_{X,k}\right|\right|_2^2\right] \leq C(n|G_k|)^{-1} \tag{18}$$

$$E\left[\left|\tilde{\lambda}^*_{j,k} - \lambda_{j,k}\right|^q\right] \leq C(n|G_k|)^{-q/2} \tag{19}$$

for $1 \leq j \leq \tilde{m}$ and $q = 1, 2, \ldots$. Further note that the sequence $\{(z_{it} X_{it}) : \ 1 \leq i \leq n, \ t \in G_k\}$ is m-dependent (after suitable relabeling). As a consequence of this,

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

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 simple moment calculations. Given Assumption 1 and 2, (18) and (19) are a consequence of Theorem 3.2 in Hörmann and Kokoszka (2010). The distances $\left|\left|\tilde{\phi}^*_{j,k} - \phi_{j,k}\right|\right|_2^q$ can be shown to converge at rate $n^{-q} j^q$ uniformly over $1 \leq j \leq \tilde{m}$ using similar arguments as in the proof of Theorem 1. Beyond that it also holds given Assumptions 1 and 2 that

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

This can be seen as follows. Suppose w.l.o.g. $G_k = \{1, 2, \ldots, |G_k|\}$. A panel regression

$$z_{it} = \langle \zeta, X_{it} \rangle + \eta_{it} \tag{20}$$

for $1 \leq t \leq |G_k|$ and $1 \leq i \leq n$ can be formulated in terms of pooled data according to

$$z_{j(i,t)} = \langle \zeta, X_{j(i,t)} \rangle + \eta_{j(i,t)} \tag{21}$$

where the re-labeling proceeds according to $1 \leq j(i,t) := (i-1)|G_k| + t \leq J_{nk} := n|G_k|$. The distance $\left|\left|\tilde{\Phi}^*_k - \Phi\right|\right|_{H'}^2$ is, as alluded to in Shin (2009), equal to the L^2 distance $||\hat{\zeta}_{J_{nk}} - \zeta||_2^2$, with $\hat{\zeta}_{J_{nk}}$ being the estimator for ζ presented in Hall and Horowitz (2007) for regression (21). Their arguments (see the proof of their Theorem 1) transfer mutatis mutandis immediately to a setup with weak dependence in the sense of L^4_m dependent regressor (which is in our setup a consequence of Assumption 1) and m-dependent errors possessing at least four moments. This can easily be shown using the set of results formulated in Hörmann and Kokoszka (2010). As the artificial estimator $\hat{\zeta}_{J_{nk}}$ is calculated from a sample of size J_{nk}, the convergence rate is

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

A.6 Proof of Theorem 2

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

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

In the proof of Lemma 4.1 it was shown already, that $\mathbb{P}\left(Q_t^c\right)=o(|G_k|^{-1})$. Regarding the remaining term, note that

$$
\begin{aligned}
\mathbb{P}(n|G_k|^{-1}\sum_{t\in G_k}R_{0,1,t}^2>c)&\leq\frac{nE[R_{0,1,t}^2]}{c}=O(1)\\
\mathbb{P}(n|G_k|^{-1}\sum_{t\in G_k}R_{0,1,t}^2>c)&\leq\frac{nE[R_{0,2,t}^2]}{c}=O(1)\\
\mathbb{P}(n|G_k|^{-1}\sum_{t\in G_k}R_{1,1,t}^2>c)&\leq\frac{nCE[||\hat{K}_{zX,t}-K_{zX}||_2^2]}{c}=O(1)
\end{aligned}
$$

as a consequence of the exogeneity of the regressors and stationarity within the regime. We argued before, that $R_{2,1}$, which does not depend on index t vanishes at a rate faster than $n^{-1/2}$. Further $\mathbb{P}(\bigcup\mathcal{F}_{3,m,t}^c)=o(1)$ as well as $\mathbb{P}(\bigcup\mathcal{F}_{4,m,t}^c)=o(1)$ as shown in the proof of Lemma 4.1. It is hence straightforward to argue

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

due to the stationarity of X_{it}, $t \in G_k$. Similarly

$$\begin{aligned}
&\mathbb{P}(n|G_k|^{-1}\sum_{t\in G_k} R_{3,2,t}^2 > c) \\
\leq &\frac{||\alpha_t||_2^2 nm \sum_{j=1}^m \langle K_{zX,k}, \phi_{j,k}\rangle^2 E\left[||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 \mathbb{1}(\mathcal{F}_{3,m,t})\right]}{c} + |G_k|\mathbb{P}(\mathcal{F}_{3,m,t}^c) \\
= &O(1) + o(1).
\end{aligned}$$

Finally, since an upper bound on $\mathbb{P}(n|G_k|^{-1}\sum_{t\in G_k} R_{3,3,t}^2 > c)$ can be obtained from

$$n||K_{zX,k}||_2^2||\alpha_t||_2^2 m \sum_{j=1}^m E\left[||\hat{\phi}_{j,t} - \phi_{j,k}||_2^4 \mathbb{1}(\mathcal{F}_{3,m,t})\right] = O(1)$$

and

$$mnE\left[\sum_{j=1}^m ||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 \mathbb{1}(\mathcal{F}_{3,m,t})(a_{j,t}^*)^2\right]$$

it follows $\mathbb{P}(\mathbb{P}(n|G_k|^{-1}\sum_{t\in G_k} R_{3,3,t}^2 > c)) = O(1)$. As the constants used in the probabilities can be chosen arbitrarily large, combining arguments it follows that $|G_k|^{-1}\sum_{t\in G_k}(\hat{\beta}_t - \beta_t)^2 = O_p(n^{-1})$ as $n, T \to \infty$.
In what follows, we use the notation listed below:

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

Now, the ultimate 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

$$||\tilde{A}_k^* - A_k||_2^2 = \left|\left|\sum_{j=1}^{\tilde{m}} \left(\tilde{a}_{j,k}^{(1)} + \tilde{a}_{j,k}^{(2)}\right) \tilde{\phi}_{j,k}^* - A_k\right|\right|_2^2 \tag{22}$$

$$\leq 2\left|\left|\sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}^{(1)}\tilde{\phi}_{j,k}^* - A_k\right|\right|_2^2 + 2\sum_{j=1}^{\tilde{m}}(\tilde{a}_{j,k}^{(2)})^2 \tag{23}$$

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 Remarks 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 (23) can be handled as follows. First, note that

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

as $E[\frac{1}{|G_k|}\sum_{t\in G_k}||X_t - \bar{\bar{X}}_k||_2^2] = O(n^{-1})$. Further

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

Finally

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

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

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

Define the event $\mathcal{F}_{5,\tilde{m},k}:=\{|\tilde{\lambda}^*_{j,k}-\lambda_{j,k}|\leq\frac{1}{2}\lambda_{j,k}:\quad 1\leq j\leq\tilde{m}\}$ and note that using similar arguments as before root-n consistency of the empirical covariance operator calculated from observations in G_k, $\mathbb{P}(\mathcal{F}_{5,\tilde{m},k})\to 1$ as $(n,T)\to\infty$. Given $\mathcal{F}_{5,\tilde{m},k}$ holds, Assumption 2 implies

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

Further it follows from the Cauchy Schwarz inequality that on $\mathcal{F}_{5,\tilde{m},k}$

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

because of the Assumptions 3-5. Further, similar arguments as before can be employed to obtain on $\mathcal{F}_{5,\tilde{m},k}$

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

implying $\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle K_{zX,k},\tilde{\phi}^*_{j,k}-\phi_{j,k}\rangle^2=O_p(1)$. Combining the above arguments yields $\sum_{j=1}^{\tilde{m}}(\tilde{a}_j^{(2)})^2=O_p(n^{-1})$. Further, if $\nu>\frac{1+\mu+\delta}{2\delta}$ 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 on the classification error the result in the Theorem follows. ■

A.7 Threshold Choice

In order to illustrate the properties of the threshold τ_{nT} as suggested in Section 5, suppose for a moment that the truncation error in regime k is negligible (i.e., $\lambda_{jk}\approx 0, j\geq\underline{m}+1$) and that the eigenvalue-eigenfunction pairs $(\lambda_{jk},\phi_{jk})_{j\geq 1}$ as well as the error variance $\sigma^2_{\epsilon,k}$ of regime k were known. In this case our estimation procedure yields variance adjusted estimators $\hat{\alpha}_t^{(\Delta^*)} = \sum_{j=1}^{\underline{m}} \sigma_{\epsilon,k}^{-1}\lambda_{jk}^{1/2}\hat{a}_{jt}\phi_{jk}$ and $\hat{\alpha}_s^{(\Delta^*)} = \sum_{j=1}^{\underline{m}} \sigma_{\epsilon,k}^{-1}\lambda_{jk}^{1/2}\hat{a}_{js}\phi_{jk}$ where the appropriately scaled difference of their jth components $(n/2)^{1/2}\sigma_{\epsilon,k}^{-1}\lambda_{jk}^{1/2}(\hat{a}_{jt}-\hat{a}_{js})$ is approximately standard normal (for large n), such that for all $t, s \in G_k$

$$\frac{n}{2}\Delta_{ts}^* := \frac{n}{2}||\hat{\alpha}_t^{(\Delta^*)} - \hat{\alpha}_s^{(\Delta^*)}||_2^2 = \sum_{j=1}^{\underline{m}} \left(\left(\frac{n}{2}\right)^{1/2} \sigma_{\epsilon,k}^{-1}\lambda_{jk}^{1/2}(\hat{a}_{jt} - \hat{a}_{js}) \right)^2 =: Q_{ts}^{\underline{m}}$$
$$\Rightarrow \Delta_{ts}^* = \frac{2}{n}Q_{ts}^{\underline{m}}, \quad \text{where (for large } n) \quad Q_{ts}^{\underline{m}} \sim \chi_{\underline{m}}^2 \quad \text{if } t \neq s \text{ and } \; Q_{ts}^{\underline{m}} = 0 \text{ if } t = s.$$

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 and holds under our setup for any $s \in G_k$ with $s \neq t$. 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 $\tau_{nT} = (2/n)F_{\underline{m}}^{-1}(\alpha_\tau)$, for some α_τ very close to one as mentioned in Section 5. Note that according to Theorem A in Inglot (2010) and our assumptions in Section 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 $n, T \to \infty$. That is, the above τ_{nT} is indeed an asymptotically valid threshold.

B Additional Simulation Results

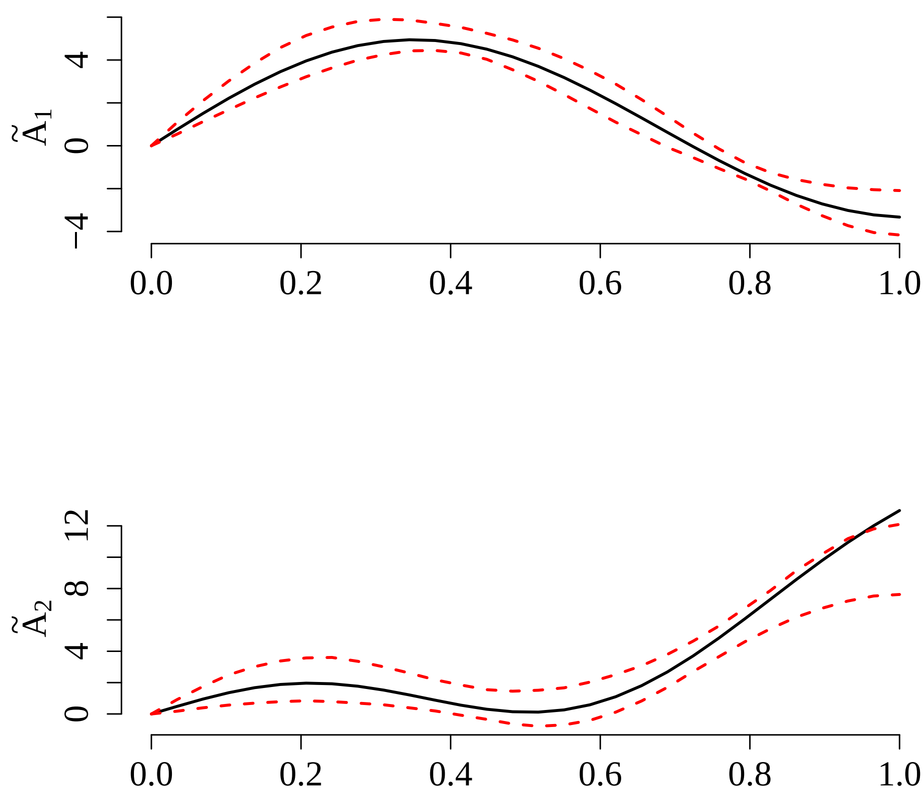

Figure 4: *Scenario 1: Estimated regime parameter functions for sample size* $(n, T) = (50, 50)$.

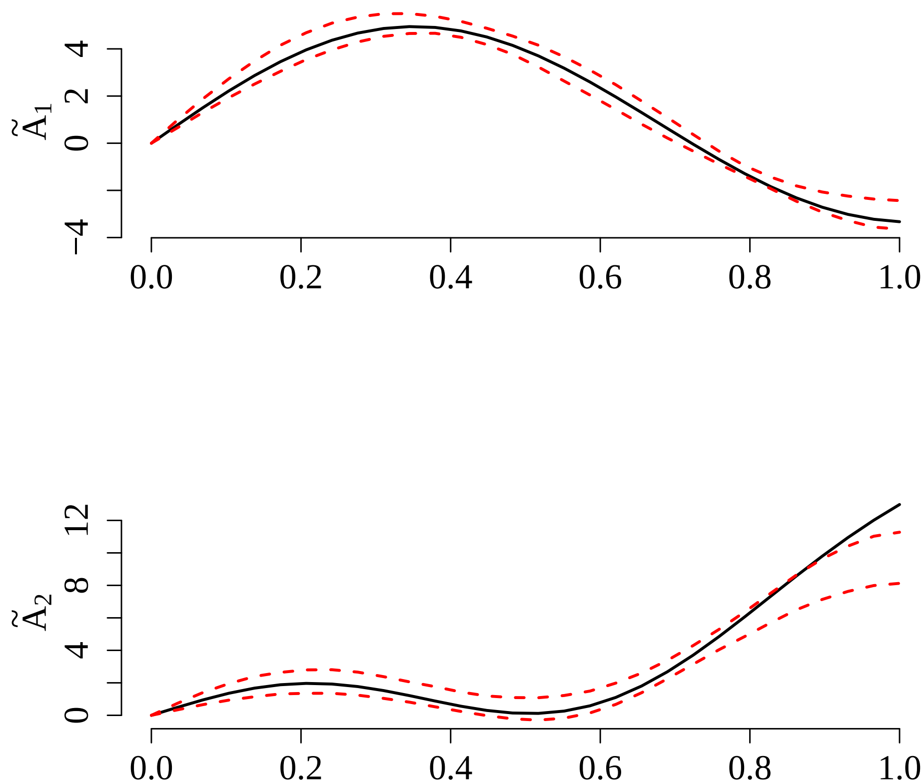

Figure 5: *Scenario 1: Estimated regime parameter functions for sample size* $(n, T) = (100, 50)$.

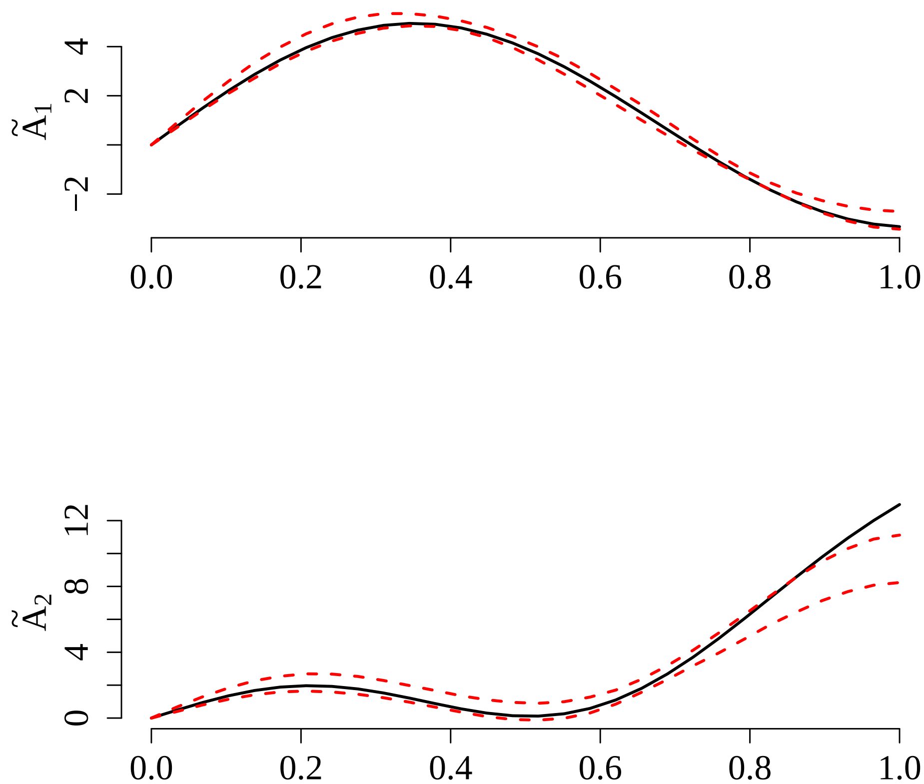

Figure 6: *Scenario 1: Estimated regime parameter functions for sample size* $(n, T) = (150, 80)$.

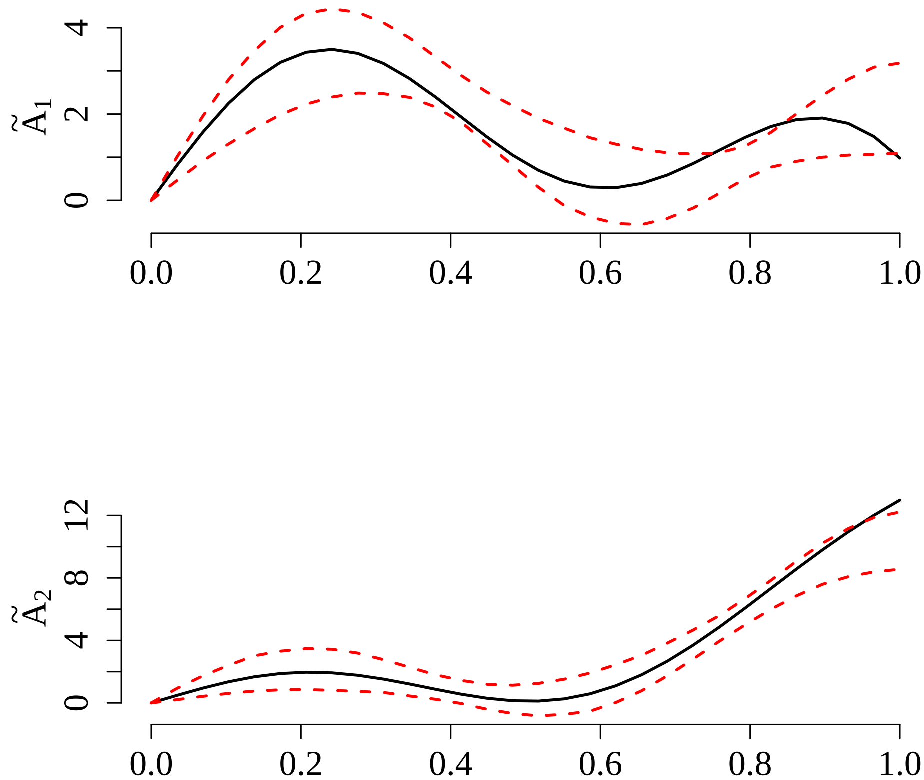

Figure 7: *Scenario 2: Estimated regime parameter functions for sample size* $(n, T) = (50, 50)$.

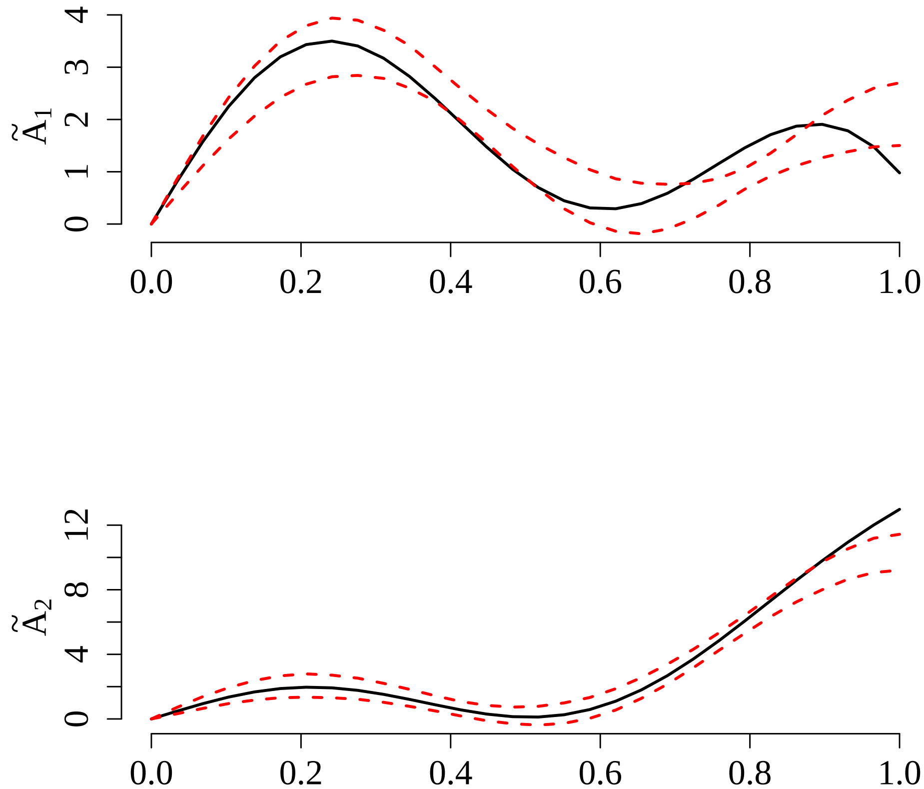

Figure 8: *Scenario 2: Estimated regime parameter functions for sample size* $(n, T) = (100, 50)$.

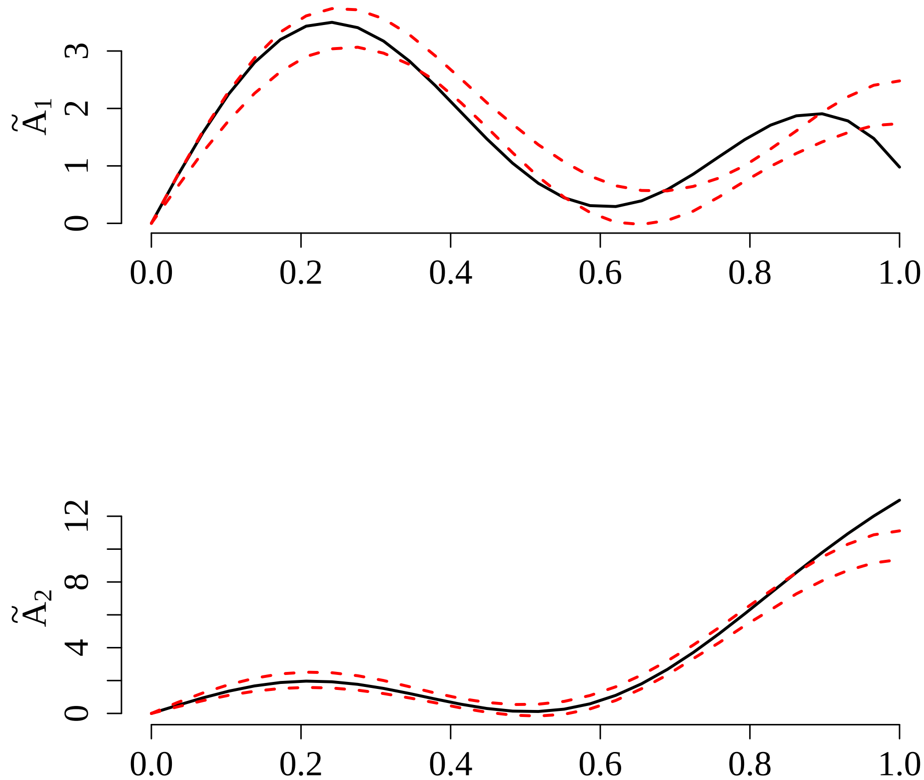

Figure 9: *Scenario 2: Estimated regime parameter functions for sample size* $(n, T) = (150, 80)$.

References

Aue, A., D. D. Norinho, and S. Hörmann (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association 110*(509), 378–392.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *The Annals of Statistics 38*(3), 1845–1884.

Inglot, T. (2010). Inequalities for quantiles of the chi-square distribution. *Probability and Mathematical Statistics 30*(2), 339–351.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference 139*(10), 3405–3418.

Vogt, M. and O. Linton (2017). Classification of non-parametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(1), 5–27.