Regime Switching in Partially Functional Linear Regression for Panel Data

Fabian Walders[a] and Dominik Liebl[b]

[a]University of Bonn, BGSE & Institute for Financial Economics and Statistics, Bonn, Germany. fwalders@uni-bonn.de
[b]University of Bonn, Institute for Financial Economics and Statistics, Bonn, Germany. dliebl@uni-bonn.de

16 March 2017

Abstract

In this work we introduce a novel semiparametric functional regression model for panel data. While the parametric term is completely time varying, the functional component changes over parameter regimes in time. The regimes are latent and need to be estimated from data. Our model hence offers a data driven compromise between full homogeneity and full heterogeneity of the nonparametric terms. We develop asymptotic theory for the suggested estimators including rates of convergence as n and T diverge simultaneously. In an application, our statistical model offers a high resolution perspective on the *idiosyncratic volatility puzzle*. We undermine this perspective by economic modeling and provide a vast empirical study based on high frequency data for stocks from the S&P 500.

Keywords: Functional Data Analysis, Semiparametric Regression, Panel Data, Classification of Regression Curves

JEL Classification: C14, C23

Introduction

This work contributes a novel semiparametric regression model for longitudinal data to the literature. The suggested approach allows a scalar response to be affected by a random regressor function as well as a finite dimensional predictor in a time varying manner. The nonparametric functional term changes over time by switching between different parameter regimes. Such regimes are a latent partition of time and must be estimated from the data. The parametric component is allowed to vary over time independently from the regimes. Estimation relies on estimators from the functional data literature paired with a straightforward classification strategy. We develop asymptotic theory for the estimators as n and T diverge simultaneously. Given a set of standard assumptions, we prove classification consistency and provide rates of convergence.

Our statistical approach relates in particular to two topics in the literature: limited heterogeneity in panel data and functional data analysis. Regarding the former, the recent works of Bada et al. (2015), Vogt and Linton (2017) and Su et al. (2016) are important references. At this strand of the literature, parameters of panel data regressions vary over latent index sets, structuring the cross section or time dimension. Such sets need to be estimated from data along with the structure of the conditional expectation of the regressand. These regression frameworks are particularly appealing as they not only allow to estimate the marginal effects of a regressor but also the degree of heterogeneity in this effect over individuals or time. To this literature, we contribute a new type of regression model incorporating this type o limited parameter instabilities. The fundamental novelty of the considered regression is constituted by the joint presence of functional and vector valued regressors. This is ultimately motivated by the fact that many empirical problems ins econometrics incorporate outcomes of different types of random variables, as is e.g. made precise in our application to finance. From another perspective our model is a contribution to the functional data literature. We borrow the structure from the work of Shin (2009) who introduced a partially functional regression model in a cross section context. Our framework generalizes her model to a panel context, while incorporating the aforementioned flexible type of heterogeneity. In the functional data literature changing marginal effects of the functional regressor is, however in a time series context, also present in the model of Horváth and Reeder (2012), who develop a testing procedure to identify time heterogeneity in a fully functional regression.

Beyond the technical novelty, our model also allows to gather new empirical insights to the idiosyncratic volatility puzzle. The latter emerged from the seminal work of Ang et al. (2006) and summarizes two incompatible viewpoints. On the one hand economic theory suggests, investors should either ask for a positive premium for bearing idiosyncratic risk or should not price it at all in case they are well diversified. On the other hand many empirical studies suggest negative premiums on idiosyncratic risk. In an application we demonstrate why our model is tailor made to uncover heterogeneous dynamics of such idiosyncratic risk pricing in the presence of market frictions. Our findings allude to the presence of both, puzzling as well as non-puzzling periods in the above sense. These periods change over days as well as trading hours. Given our high resolution volatility measure we show that ultimately heterogeneity dominates puzzling.

The remainder of this work is structured as follows. In sections 2 and 3 we introduce the model and present the estimation procedure. In section 4 we develop the asymptotic theory, including the main results. The finite sample performance of the estimators is explored in section 5. Section 6 briefly discusses the practical threshold choice, while section 7 offers the application. The latter consists of a precise economic modeling approach and an empirical study. A last section 8 concludes.

Model

We suggest a regression model which is formally obtained as a partially linear regression as introduced in Shin (2009) adapted to a panel data context. Generically the task is 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 $z_{it} \in \mathbb{R}^P$. Regarding the function-valued random variables, we do not consider measurement errors for simplicity.[1] Indexing the cross section units and time periods $1 \leq i \leq N$ and $1 \leq t \leq T$ respectively, our statistical model reads as

$$y_{it} = \alpha_{0,t} + \int_0^1 \alpha_t(s) X_{it}(s) \mathrm{d}s + \beta_t^T z_{it} + \epsilon_{it}. \tag{1}$$

The unknown parameters α_t differ across different regimes $G_k \subset \{1, ..., T\}$. Within regime k, there is a square integrable parameter function $A_k \in L^2[0,1]$ governing the effect of X_{it} on y_{it}, i.e.

$$\alpha_t = A_k \ \text{ if } t \in G_k. \tag{2}$$

The regimes $G_1, \dots, G_K$ form a latent partition of the set $\{1, \dots, T\}$. This notion certainly allows a regime to be constituted not only by subsequent, i.e. adjacent periods, but also by periods being more than a single period away from one another. In this respect our model is more general than the parameter instabilities introduced in Bada et al. (2015). From a time series context view, it would rather be close to Markov-type regime switching models (see e.g. Hamilton, J. D. (2010)). In turn from a panel perspective, the parameterization in (2) can also be understood as a semiparametric time series version of the cross sectional group heterogeneity introduced in the longitudinal data setup by Su et al. (2016) and Vogt and Linton (2017).

Model (1) compromises two extreme specifications. On the one hand it might be the case that $K = 1$ and hence $G_1 = \{1, \dots, T\}$. In this situation the effect of the random function on the response is time invariant. On the other hand the model nests complete heterogeneity in another extreme situation in which $K = T$ and all Regimes G_k are singletons. This turns (1) into a collection of T different cross section models. Clearly, pure functional or pure parametric specifications are possible in case $\beta_t = 0$ or $\alpha_t = 0$ for all $1 \leq t \leq T$ respectively.

Among other interpretations, the index s might well be understood as labeling within-period time. From this perspective the notion of heterogeneity can, beyond the aforementioned one, be attributed to differences of a single A_k over different subsets of $[0,1]$. Together the regime structure and differences in function values of the parameter functions A_k over their domains allow to reflect complex forms of time heterogeneity. From this perspective the model also links to the recent literature on domain selection (see e.g. Hall and Hooker (2016)) and sparsity in functional regression. It allows to discriminate periods giving zero-weight to the regressor on subsets of the unit interval $[0,1]$. Along with the above interpretation there are also two natural notions of the degree of heterogeneity in the nonparametric parts. One is the number of regimes, K, the other one the dissimilarities between A_k and A_l.

Although we focus on explicitly modeling differences and similarities among the nonparametric terms, (1) naturally provides a notion of differences in the parametric and hence also level effects. In particular the regression is able to reflect level switching processes over time. This flexibility introduced by $\alpha_{0,t}$ also allows certain deviations from trend stationarity in y_{it} over time.

[1]See e.g. Ramsay and Silverman (2005) or Yao et al. (2005) besides others for this topic.

Estimation

The objective is to estimate the parameters A_k, β_t and $\alpha_{0,t}$ as well as the regimes $G_1, \dots, G_K$ from realizations of the random variables $\{(y_{it}, X_{it}, z_{it}):\ 1 \leq i \leq n,\ 1 \leq t \leq T\}$. We suggest to employ a multi-stage estimation procedure for this purpose. At a first stage model (1) is treated as a collection of T cross section regressions. Estimates gathered for each of these models should be reasonable already given n is sufficiently large. In particular estimates for the functional parameters should ideally reveal information about the regime memberships. Along these lines, a second stage classifies periods in regimes based on suitably scaled differences of cross section estimates. At a third stage the model is estimated within every regime, improving upon accuracy of the functional parameter estimate.

The structure of the estimation procedure is close to Vogt and Linton (2017), however differs particularly in the last step from their approach. First stage regressions can, apart from a slight modification, be estimated using the tools in Shin (2009). In what follows we outline the suggested four step procedure.

Step 1. In a first step the parameters α_t and β_t are estimated separately for each $1 \leq t \leq T$ using the estimators in Shin (2009) in the auxiliary model

$$\breve{y}_{it} = \langle \alpha_t, X_{it} \rangle + \beta_t^T z_{it} + \breve{\epsilon}_{it},$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2([0,1])$. The regressor $\breve{y}_{it} := y_{it} - \hat{\alpha}_{0,t}$ is a cross sectionally demeaned version of y_{it}, while the corresponding mean $\hat{\alpha}_{0,t} := \frac{1}{n}\sum_{i=1}^n y_{it}$ offers a natural estimate for the time specific constant $\alpha_{0,t}$. The error terms are constituted by the resulting estimation error and the original errors, i.e. $\breve{\epsilon}_{it} := \epsilon_{it} + (\alpha_{0,t} - \hat{\alpha}_{0,t})$.

The above auxiliary regression can, for a pre-specified cut-off parameter $1 \leq m < n$, be approximated by

$$\breve{y}_{it} \approx \sum_{j=1}^{m} a_{j,t} \langle X_{it}, \hat{\phi}_{j,t} \rangle + \beta_t^T z_{it} + \breve{\epsilon}_{it}, \tag{3}$$

where $a_{j,t} := \langle \alpha_t, \hat{\phi}_{j,t} \rangle$. The functions $\hat{\phi}_{j,t}$, $1 \leq j \leq n$ are the eigenfunctions of the empirical covariance operator of $\{X_{it}:\ 1 \leq i \leq n\}$, denoted $\hat{\Gamma}_t$, whose integral kernel $\hat{K}_{X,t}$ is defined below. The j-th eigenfunction is associated with the j-th largest eigenvalue of $\hat{\Gamma}_t$, $\hat{\lambda}_{j,t}$. The estimators in Shin (2009) are based on least squares estimators for $a_{j,t}$, $1 \leq j \leq m$ and β_t in (3). Note however that the estimators are in our context slightly different due to the additional estimation error in the constant. In order to obtain a closed form for the ultimate estimates, some additional notation is required:

$$\hat{K}_{z_pX,t}(s) := n^{-1}\sum_{i=1}^{n} z_{p,it}X_{it}(s), \qquad \hat{\mathbf{K}}_{zX,t}(s) := [\hat{K}_{z_1X,t}(s),\ldots,\hat{K}_{z_PX,t}(s)]^T, \qquad \hat{K}_{z_p\breve{y},t} := n^{-1}\sum_{i=1}^{n} z_{p,it}\breve{y}_{it},$$

$$\hat{\mathbf{K}}_{z\breve{y},t} := [\hat{K}_{z_1\breve{y},t},\ldots,\hat{K}_{z_P\breve{y},t}]^T, \qquad \hat{\mathbf{K}}_{z,t} := n^{-1}\sum_{i=1}^{n} z_{it}z_{it}^T, \qquad \hat{K}_{\breve{y}X,t}(s) := n^{-1}\sum_{i=1}^{n}\breve{y}_{it}X_{it}(s),$$

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

$$\hat{\Phi}_t(\hat{\mathbf{K}}_{zX,t}) := [\hat{\Phi}_{p,t}(\hat{K}_{z_qX,t})]_{1\le p\le P,\ 1\le q\le P} \quad \text{and} \quad \hat{K}_{X,t}(u,v) := n^{-1}\sum_{i=1}^{n}X_{it}(u)X_{it}(v).$$

Now, pre-estimates for α_t and β_t are obtained according to

$$\hat{\alpha}_t = \sum_{j=1}^{m}\hat{a}_{j,t}\hat{\phi}_{j,t} \quad \text{and} \quad \hat{\beta}_t = \left[\hat{\mathbf{K}}_{z,t} - \hat{\Phi}_t(\hat{\mathbf{K}}_{zX,t})\right]^{-1}\left[\hat{\mathbf{K}}_{z\breve{y},t} - \hat{\Phi}_t\left(\hat{K}_{x\breve{y},t}\right)\right],$$
$$\text{where} \quad \hat{a}_{j,t} = \hat{\lambda}_{j,t}^{-1}\frac{1}{n}\sum_{i=1}^{n}\langle X_{it},\hat{\phi}_{j,t}\rangle(\breve{y}_{it} - \hat{\beta}_t^T z_{it}) \ \ 1\le j\le m.$$

Based on these quantities the final outcome of the first stage is calculated. Precisely this is a scaled version of $\hat{\alpha}_t$, which is obtained according to

$$\hat{\alpha}_t^{(\Delta)} := \sum_{j=1}^{m}\frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{t,\epsilon}}\hat{a}_{j,t}\hat{\phi}_{j,t}, \tag{4}$$

where $\hat{\sigma}^2_{t,\epsilon} := n^{-1}\sum_{i=1}^{n}\left(\breve{y}_{it} - \langle\hat{\alpha}_t, X_{it}\rangle + \hat{\beta}_t^T z_{it}\right)^2$, the empirical residual variance from the above regression for time t.

Step 2 In the second step, the t-specific functions $\hat{\alpha}_t^{(\Delta)}$ are classified into regimes using a variant of the algorithm in Vogt and Linton (2017). For this purpose, denote the squared standard L^2-distance between $\hat{\alpha}_t^{(\Delta)}$ and $\hat{\alpha}_s^{(\Delta)}$ as $\hat{\Delta}_{ts} := ||\hat{\alpha}_t^{(\Delta)} - \hat{\alpha}_s^{(\Delta)}||_2^2$. We will use $||\cdot||_2$ to denote the L^2 norm from now on without further reference. Further, given some set S and an index $t\in S$, let $0 = \hat{\Delta}_{t(1)} \le \cdots \le \hat{\Delta}_{t(|S|)}$ denote the order statistics corresponding to differences between $\hat{\alpha}_t^{(\Delta)}$ and $\hat{\alpha}_s^{(\Delta)}$, $s\in S$. The set $\{(1),\ldots,(l)\}$ describes the original indexes in S corresponding to the first l order statistics.

In the algorithm described below we use a threshold $\tau_{nT} > 0$, determining how far $\hat{\alpha}_t^{(\Delta)}$ and $\hat{\alpha}_s^{(\Delta)}$ might be away in order to be classified in the same regime. The theoretical behavior of this tuning parameter is determined in section 4, while we suggest a practical choice in section 5. This practical choice is based on the approximate distribution of the distance $\hat{\Delta}_{ts}$. The availability of such a distributional feature is the reason of employing $\hat{\alpha}_t^{(\Delta)}$ rather than $\hat{\alpha}_t$ for classification.

Now, the aforementioned classification algorithm is initialized by setting $S^{(0)} := \{1,\ldots,T\}$. Then an index $s_0 \in S^{(0)}$ is randomly drawn and the index $(\hat{\kappa}_0)$ is determined according to $\hat{\Delta}_{t(\hat{\kappa}_0)} \le \tau_{nT} < \hat{\Delta}_{t(\hat{\kappa}_0+1)}$. Define the first regime estimate as $\hat{G}_1 := \{(1),...,(\hat{\kappa}_0)\}$. Then iterate this procedure over $j = 1, 2, \ldots$ with $S^{(j)} := \{1, \ldots, T\} \setminus \bigcup_{1 \le l \le j-1} S^{(l)}$ as long as $|S^{(j)}| > 0$. The iteration stops if $|S^{(j)}| = 0$ and the last regime estimate receives index $\hat{K}$, providing a natural estimator of K.

Step 3. Based on the estimated regimes $\hat{G}_1, \ldots, \hat{G}_{\hat{K}}$ from the previous step (i) the covariance operator of X_{it} is re-estimated and (ii) final estimates for β_t are calculated. The first point simply consists of calculating the empirical covariance operator of $\{X_{it} : 1 \le i \le n,\ t \in \hat{G}_k\}$, with integral kernel $\tilde{K}_{X,k}$ defined below for any $1 \le k \le \hat{K}$. Eigenvalue-eigenfunction-pairs of the k-th operator are denoted $(\tilde{\lambda}_{j,k}, \tilde{\phi}_{j,k})$, $1 \le j \le n|\hat{G}_k|$, and are ordered according to $\tilde{\lambda}_{1,k} \ge \cdots \ge \tilde{\lambda}_{n|\hat{G}_k|} \ge 0$. Given a truncation parameter $1 \le m \le \tilde{m} < n|\hat{G}|$ and quantities

$$\tilde{K}_{z_pX,k}(s) := \big(n|\hat{G}_k|\big)^{-1} \sum_{i=1}^{n} \sum_{t \in \hat{G}_k} z_{p,it} X_{it}(s), \quad \tilde{\mathbf{K}}_{zX,k}(s) := [\tilde{K}_{z_1X,k}(s), \ldots, \tilde{K}_{z_PX,k}(s)]^T$$

$$\tilde{K}_{z_p\breve{y},k} := \big(n|\hat{G}_k|\big)^{-1} \sum_{i=1}^{n} \sum_{t \in \hat{G}_k} z_{p,it} \breve{y}_{it}, \quad \tilde{\mathbf{K}}_{z\breve{y},k} := [\tilde{K}_{z_1\breve{y},k}, \ldots, \tilde{K}_{z_P\breve{y},k}]^T$$

$$\tilde{\mathbf{K}}_{z,k}(s) := \big(n|\hat{G}_k|\big)^{-1} \sum_{i=1}^{n} \sum_{t \in \hat{G}_k} z_{it} z_{it}^T, \quad \tilde{K}_{\breve{y}X,k}(s) := \big(n|\hat{G}_k|\big)^{-1} \sum_{i=1}^{n} \sum_{t \in \hat{G}_k} \breve{y}_{it} X_{it}(s),$$

$$\tilde{\Phi}_{p,k}(g) := \sum_{j=1}^{\tilde{m}} \frac{\langle \tilde{K}_{z_pX,k}, \tilde{\phi}_{j,k}\rangle \langle \tilde{\phi}_{j,k}, g\rangle}{\tilde{\lambda}_{j,k}} \ \text{ for } g \in L^2([0,1]), \quad \tilde{\Phi}_k(g) := [\tilde{\Phi}_{1,k}(g), \ldots, \tilde{\Phi}_{P,k}(g)]^T$$

$$\tilde{\Phi}_k(\tilde{\mathbf{K}}_{zX,k}) := [\tilde{\Phi}_{p,k}(\tilde{K}_{z_qX,k})]_{1 \le p \le P,\, 1 \le q \le P} \quad \text{and} \quad \tilde{K}_{X,k}(u,v) := \big(n|\hat{G}_k|\big)^{-1} \sum_{i=1}^{n} \sum_{t \in \hat{G}_k} X_{it}(u) X_{it}(v),$$

the final slope estimates are obtained according to

$$\tilde{\beta}_{k,t} := \left[\tilde{\mathbf{K}}_{k,z} - \tilde{\Phi}_k(\tilde{\mathbf{K}}_{zX,k})\right]^{-1} \left[\hat{\mathbf{K}}_{z\breve{y},t} - \tilde{\Phi}_k\left(\hat{K}_{\breve{y}X,t}\right)\right], \quad t \in \hat{G}_k.$$

Intuitively, any quantity in $\hat{\beta}_t$ which might be more accurately estimated by pooling observations within $\hat{G}_k$ is replaced by corresponding within-regime estimates. This makes $\tilde{\beta}_{k,t}$ in particular analytically more tractable and offers a logical extension to the first stage. Nevertheless the objects $\hat{\beta}$, might also be reported as final estimates for β_t. Finally relying on $\tilde{\beta}_{t,k}$ the response variables are for each $t \in \hat{G}_k$ transformed according to $\tilde{y}_{it} = \breve{y}_{it} - \tilde{\beta}_t^T z_{it}$, dropping the index k in $\tilde{y}_{it}$ for notational convenience.

After this step the response variable is finally *adjusted* for any parametric effects which were present in the original model. This now allows to employ the standard technique, presented in Hall and Horowitz (2007), to isolate the effect of X_{it} on y_{it} as follows.

Step 4. In a last step regime parameter functions A_k are estimated. For any $1 \le k \le \hat{K}$ the pairs $\{(\tilde{y}_{it}, X_{it}) : 1 \le i \le n,\ t \in \hat{G}_k\}$ are pooled in order to estimate A_k in an auxiliary regression

$$\tilde{y}_{it} = \langle X_{it}, A_k\rangle + \epsilon_{it}^*$$

using the procedure in Hall and Horowitz (2007). Here, the error term is composed according to $\epsilon_{it}^* := \breve{\epsilon}_{it} + (\beta_t - \tilde{\beta}_t)^T z_{it}$. The resulting estimator $\tilde{A}_k$ obtains as

$$\tilde{A}_k(s) = \sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}\tilde{\phi}_{j,k}(s), \quad \text{where} \quad \tilde{a}_{j,k} = \tilde{\lambda}_{j,k}^{-1} \left(n|\hat{G}_k|\right)^{-1} \sum_{i=1}^{n} \sum_{t\in\hat{G}_k} \langle X_{it}, \tilde{\phi}_{j,k}\rangle \tilde{y}_{it}.$$

In summary final estimates are $\hat{G}_1, ..., \hat{G}_{\hat{K}}$ for the regimes as well as $\tilde{\beta}_t$ and $\tilde{A}_k$ for the regression parameters.

Typically all steps in the above procedure are computationally convenient even for large sample sizes, due to the least squares procedures in steps 1 and 4. Step 2 involves the computation of $O(T^2)$ integrals, which will however usually be fast as these integrals are univariate. Also estimates $\tilde{\beta}_{t,k}$ are made up of simple matrix expressions and hence fast to calculate.

Asymptotic Theory

In the presented framework, there are two types of problems adding to the well understood estimation in (partially linear) functional linear regression. The first one is that there is an additional classification error contaminating estimation of A_k. Second, as the slopes and intercepts in the parametric component are t-specific they play the role of nuissance parameters in the problem of estimating A_k.

For our analysis we rely on a set of standard assumptions present in the literature[2]. The latter comprise first of all moment conditions and exogeneity assumptions listed below.

Assumption 1. Suppose that

1. $\{(\epsilon_{it}, X_{it}, z_{it}) : \ 1 \leq i \leq n, \ 1 \leq t \leq T\}$ are centered, strictly stationary and independent over the index i. In addition the ϵ_{it} are independent over the index t as well.
2. For every i, X_{it} can be represented as

$$X_{it} = f(\eta_{i1}, \ldots, \eta_{i,m-1}, 0, 0, \ldots),$$

with $\{\eta_{it} : \ 1 \leq i \leq n, \ 1 \leq t \leq T\}$ being a sequence of iid random variables with values in a measurable space $\mathcal{M}$. The function $f : \mathcal{M}^\infty \to L^2([0,1])$ is measurable.
3. For every i, the sequence $\{z_{it} : \ 1 \leq t \leq T\}$ is m-dependent.
4. Suppose that $E\left[||X_{it}||_2^4\right] < \infty$, $E\left[z_{it}^4\right] < \infty$, $E\left[\epsilon_{it}^4\right] < \infty$.
5. ϵ_{it} is independent of X_{js} and z_{js} for any $1 \leq i, j \leq n$ and $1 \leq t, s \leq T$.

These conditions postulate cross sectional independence for all variables and weak dependence, precisely m-dependence, for the regressors within a cross section unit over time. In particular it should be noted that the functional variable is for every i trivially L^4_m-approximable over time in the sense of Hörmann and Kokoszka (2010). The concept of m-dependence is among the simplest for formulating weak dependence. Nevertheless it can well serve as an indication of dependence in panel data.

[2]See in particular Hall and Horowitz (2007), Hörmann and Kokoszka (2010), Vogt and Linton (2017) as well as Shin (2009)

Beyond this set of basic sampling assumptions, a number of regularity conditions make the notion of complexity for the functional regressor as well as corresponding parameter functions A_k precise. Further a concrete relation between X_{it} and z_{it} is introduced and the magnitude of the parametric terms is limited.

Assumption 2. Suppose there exist constants $0 < C_\lambda, C'_\lambda, C_\theta, C_a, C_{zX}, C_\beta, C_\alpha < \infty$, such that

1. $C_\lambda^{-1} j^{-\mu} \leq \lambda_j \leq C_\lambda j^{-\mu}$ and $\lambda_j - \lambda_{j+1} \geq C'_\lambda j^{-(\mu+1)}$, $j \geq 1$ for the eigenvalues $\lambda_1 > \lambda_2 > \ldots$ of the covariance operator Γ of X_{it} and a $\mu > 1$,
2. $E\left[\langle X_{it}, \phi_j \rangle^4\right] < C_\theta \lambda_j^2$ for the eigenfunction ϕ_j of Γ corresponding to the j-th eigenvalue,
3. $|\langle A_k, \phi_j \rangle| \leq C_a j^{-\nu}$ for all $1 \leq k \leq K$,
4. $|\langle K_{z_p X}, \phi_j \rangle| \leq C_{zX} j^{-(\mu+\nu)}$, for any $1 \leq p \leq P$, where $K_{z_p X} := E[X_{it} z_{p,it}]$,
5. $\sup_{1 \leq t \leq T} \beta_{p,t}^2 < C_\beta$, for any $1 \leq p \leq P$, with $\beta_{p,t}$ the p-th coordinate in β_t and
6. $\sup_{1 \leq t \leq T} \alpha_{0,t}^2 < C_\alpha$.

Statements 1 and 3 in the previous Assumption 2 are traditional in the literature (see e.g. Hall and Horowitz (2007) or Kneip et al. (2016) among others) while Shin (2009) introduces (variant of) statements 2 and 3.

Assumption 2, introduces the parameters μ and ν, which play a central role in the asymptotic analysis. Regarding their interpretation, μ can be understood as a measure of simplicity of the regressor. In case μ is large, huge parts of the variation in X_{it} can be explained by the first few principal components. On the other hand μ close to 1 is synonymous with a slow decay of the eigenvalues, indicating that a higher number of basis objects are required to explain variation in X_{it} accurately. ν serves as a measure of compatibility of A_k with the regressor X_{it}. It ultimately governs how well $\sum_{j=1}^{\infty} a_{j,t}^* \langle X_i t, \phi_j \rangle$ can be approximated by a truncated version. Larger values are synonymous with a better approximation. In this sense it can also be understood as measuring the *simplicity* of the functional $\langle X_{it}, A_k \rangle$. Before turning to the relation between the quantities μ and ν, we make precise what type of asymptotics we consider and as a by-product at which rate parameters β_t and $\alpha_{0,t}$ are added as T becomes larger.

Assumption 3. Suppose that

1. $|G_k| \propto T$ and
2. $T \propto n^\delta$ for some $0 < \delta < \frac{1}{2}$.

By writing $n, T \to \infty$ it is meant, that n and T diverge simultaneously on the path specified by the second point in Assumption 3. Precisely $T/n \to 0$ as $n, T \to \infty$ on every of the set of possible paths. The introduction of such a condition is intuitive as growing T introduces additional nuissance parameters being estimated from a number n of cross section units. The parameter δ, governing the set of admissible paths, relates to the aforementioned quantities μ and ν as follows.

Assumption 4. Suppose that $\nu > \max\{r_1, r_2, r_3\}$, where $r_1 := \frac{3+(3+\delta)\mu}{2(1-\delta)}$, $r_2 := \frac{5}{2} + 2\mu$ and $r_3 := \frac{1+\delta+\mu}{2\delta}$.

Assumption 4 can be understood as setting an upper bound on the complexity of the functionals $\langle A_k, X_{it}\rangle$, in light of the interpretation of ν given before. This upper bound depends on the complexity of the regressor in the sense outlined before and the rate δ. The ambiguous role of δ reflects two diametral roles. As mentioned before it governs the rate at which nuissanse parameters are added. On the other hand estimation of A_k benefits from the increased number of observations within the regime according to Assumption 3.1. In contrast to δ, μ plays the role as in Hall and Horowitz (2007) and Shin (2009) fore example. The less complexity there is in X_{it}, the less complexity one has to expect in the effect of X_{it} on y_{it}. Related to such complexity is the choice of the threshold parameters m and $\tilde{m}$. As different amounts of smoothing are required on the first and third & fourth stage of the procedure we introduce two different rates as follows.

Assumption 5. Suppose for the truncation parameters $m = m(n)$ and $\tilde{m} = \tilde{m}(n, T)$ that $m \propto n^{\frac{1}{\mu+2\nu}}$ and $\tilde{m} \propto (n|G_k|)^{\frac{1}{\mu+2\nu}}$

While m is the standard truncation parameter as in the related literature on cross sections, $\tilde{m}$ is a logical extension to within-regime observations.

While the previous assumptions were rather related to the nonparametric part of the regression, a t-indexed sequence of regularity conditions is required for the parametric term $\beta_t^T z_{it}$. More precisely the following conditions are, for fixed t, borrowed from Shin (2009).

Assumption 6. Suppose that for any $1 \leq t \leq T$, the random variables

$$s_{p,it} := z_{p,it} - \int_0^1 X_{it}(u) \left(\sum_{j=1}^{\infty} \frac{\langle K_{p,zX}, \phi_j \rangle}{\lambda_j} \phi_j(u) \right) \mathrm{d}u$$

are iid over index i and $E[s_{p,it}|X_{1t}, \dots, X_{nt}] = 0$. Further the $P \times P$-matrix $\left[E[s_{p,it}s_{q,it}|X_{1t}, \dots, X_{nt}]\right]_{1 \leq p,q \leq P}$ is positive definite.

It should be noted that we do not require uniformity of these conditions over index t. This is because pointwise regularity conditions are sufficient to achieve pointwise rates of convergence in the first estimation step. Uniformity of convergence will only be required for classification consistency, however without having explicit rates, as will be outlined in Lemma 2 and the proof of Theorem 1. Regarding the classification consistency another central condition obtains, along the lines of Vogt and Linton (2017), as follows.

Assumption 7.

1. The threshold parameter $\tau_{nT} \to 0$ satisfies $\mathbb{P}\left(\max_{t,s \in G_k} \hat{\Delta}_{ts} \leq \tau_{nT}\right) \to 1$ for all $1 \leq k \leq K$.
2. There exists some $C_\Delta > 0$ such that for any $1 \leq k \leq K$ and any $t \in G_k$

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

where $\alpha_t^{(\Delta)} := \sigma_\epsilon^{-1} \sum_{j=1}^{\infty} \lambda_j^{1/2} \langle \alpha_t, \phi_j \rangle \phi_j$ and $\sigma_\epsilon^2 := E[\epsilon_{it}^2]$.

On the one hand assumption 7 guarantees that the parameter functions A_k are sufficiently different over different regimes k. This might be understood as a type of identification condition. On the other hand the assumption makes the asymptotic behavior of the threshold τ_{nT} precise. It ultimately ensures that differences in parameter estimates are correctly interpreted in terms of classification.

Starting from the above assumptions, we derive two major results. One states the consistency of the classfication procedure and one provides convergence rates for the estimators. We prepare the corresponding Theorems with two Lemmas. All Proofs are deferred to Appendix A.

A first Lemma is apart from the time varying constant a corollary to Theorems 3.1 and 3.2 in Shin (2009).

Lemma 1 *Given Assumptions 1,2,4,5 and 6 it holds for all* $1 \leq t \leq T$ *as* $n, T \to \infty$ *that*

$$(\hat{\alpha}_{0,t} - \alpha_{0,t})^2 = O_p\left(n^{-1}\right), \tag{5}$$

$$\left|\left|\hat{\beta}_t - \beta_t\right|\right|^2 = O_p\left(n^{-1}\right), \tag{6}$$

$$||\hat{\alpha}_t - \alpha_t||_2^2 = O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right). \tag{7}$$

The above rates for $\hat{\alpha}_t$ are the benchmark for the rates of $\hat{A}_k$. They correspond to the rates in Hall and Horowitz (2007) as proven shown by Shin (2009). To the latter work, the first point in the Lemma simply adds an explicit rate for the constant in the model.

These results however are not sufficient to guarantee the validity of the classification algorithm. This rather requires uniform consistency, which our procedure indeed achieves in the sense of the following Lemma.

Lemma 2 *Provided Assumptions 1-6, it holds as* $n, T \to \infty$ *that*

$$\max_{1\leq t\leq T} (\hat{\alpha}_{0,t} - \alpha_{0,t})^2 = o(1), \tag{8}$$

$$\max_{1\leq t\leq T} \left|\left|\hat{\beta}_t - \beta_t\right|\right|^2 = o(1), \tag{9}$$

$$\max_{1\leq t\leq T} ||\hat{\alpha}_t - \alpha_t||_2^2 = o(1), \tag{10}$$

$$\max_{1\leq t\leq T} \left|\left|\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}\right|\right|_2^2 = o(1). \tag{11}$$

Theorem 1 *Given Assumptions 1 - 7 it holds that*

$$\mathbb{P}\left(\{\hat{G}_1, ..., \hat{G}_{\hat{K}}\} \neq \{G_1, ..., G_K\}\right) = o(1) \quad \textit{as } n, T \to \infty. \tag{12}$$

The statement of the Theorem is twofold. First, it says, that the number of groups is correctly determined, second it it says, that for any regime there exists an estimate $\hat{G}_k$, which coincides with probability tending to one. This notion of classification consistency is sufficient to achieve the following rates of convergence for the final estimates.

Theorem 2 *Given Assumptions 1 - 7 it holds for all* $1 \leq k \leq \hat{K}$ *that*

$$\left|\hat{G}_k\right|^{-1} \sum_{t \in \hat{G}_k} \left(\hat{\alpha}_{0,t} - \alpha_{0,t}\right)^2 = O_p\left(n^{-1}\right) \tag{13}$$

$$\left|\hat{G}_k\right|^{-1} \sum_{t \in \hat{G}_k} \left|\left|\tilde{\beta}_{k,t} - \beta_t\right|\right|^2 = O_p\left(n^{-1}\right) \tag{14}$$

$$\left|\left|\tilde{A}_k - A_k\right|\right|_2^2 = O_p(n^{-1}) \quad \text{as } n, T \to \infty. \tag{15}$$

The first two points mean that within every regime, the average mean squared error from the parametric estimation problems vanish at a parametric rate. These terms also constitute the contamination in the estimation of A_k. An ideal estimator $\tilde{A}_k^*$, calculated without classification error, incorporated the estimation error from a standard functional linear regression and further the aforementioned one. As a consequence, it converged according to $\left|\left|\tilde{A}_k^* - A_k\right|\right|_2^2 = O_p\left(\max\left\{n^{\frac{(1+\delta)(1-2\nu)}{\mu+2\nu}}, n^{-1}\right\}\right)$, reflecting these two sources of errors. Under Assumption 4, the maximum function takes the value n^{-1}.

Regarding the convergence rate, the improvement from considering $\tilde{A}_k$ rather than $\hat{\alpha}_t$ amounts to $n^{-1}n^{\frac{2\nu-1}{\mu+2\nu}} = n^{-\frac{1+\mu}{\mu+2\nu}}$. This means, ceteris paribus, the improvement decreases as the complexity of X_{it} decreases or the the complexity of $\langle A_k, X_{it}\rangle$ decreases.

Threshold Choice

In order to derive the results presented in the previous section assumption 7.1 introduced the asymptotic behavior we requier for τ_{nT}. In this section we address the practical choice of this tuning parameter. Its value is crucial as it directly impacts the degree of heterogeneity reported by the estimation. This in turn also has an impact on the estimated parameter functions $\tilde{A}_k$. From another perspective τ finally decides how much power comes from the time series observations in the estimation of A_k.

We suggest choosing the threshold parameter τ_{nT} based on an approximate law of $\hat{\Delta}_{ts}$, for t and s belonging to the same regime. More precisely, suppose two indixes $t, s \in G_k$. Denote the first stage estimates of the j-th eigenfunction ϕ_j as $\hat{\phi}_{j,t}$ and $\hat{\phi}_{j,s}$, for $1 \leq j \leq m$. For large n these estimates should be close their population counterpart in an L^2-sense. This argument is also valid for $\hat{\lambda}_{j,t}$ and $\hat{\lambda}_{j,s}$ being close to λ_j for large n as well as $\hat{\sigma}^2_{\epsilon,t}$ and $\hat{\sigma}^2_{\epsilon,s}$ being close to σ^2_ϵ.[3] We thus suggest to approximate the distance $||\alpha_t^{(\Delta)} - \alpha_s^{(\Delta)}||_2^2$ according to

[3]Note that for all $1 \leq t \leq T$ the consistency of $\hat{\sigma}^2_{\epsilon,t}$ is a direct consequence of Lemmas 1 and 2.

$$\hat{\Delta}_{ts} = \int_0^1 \left(\sum_{j=1}^{m} \frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{\epsilon,t}} \hat{a}_{j,t} \hat{\phi}_{t,j}(u) - \sum_{j=1}^{m} \frac{\hat{\lambda}_{j,s}^{1/2}}{\hat{\sigma}_{\epsilon,s}} \hat{a}_{j,s} \hat{\phi}_{s,j}(u) \right)^2 \mathrm{d}u \tag{16}$$

$$\approx \sum_{j=1}^{m} \frac{\lambda_j}{\sigma_\epsilon^2} (\hat{a}_{j,t} - \hat{a}_{j,s})^2 \tag{17}$$

$$= \sum_{j=1}^{m} \left(\frac{\lambda_j^{1/2}}{\sigma_\epsilon} (\hat{a}_{j,t} - a_{j,t}) + \frac{\lambda_j^{1/2}}{\sigma_\epsilon} (a_{j,t} - \hat{a}_{j,s}) \right)^2, \tag{18}$$

where $a_{j,t} := \langle \alpha_t, \hat{\phi}_{j,t} \rangle = \langle A_k, \hat{\phi}_{j,t} \rangle$.

Given that $(\hat{\alpha}_{0t} - \alpha_{0t}) + (\hat{\beta}_t - \beta_t) z_{it} \approx 0$, which is reasonable in light of Lemmas 1 and 2, the law of $\hat{a}_{j,t} - a_{j,k} \approx \hat{\lambda}_j^{-1} n^{-1} \sum_{i=1}^{n} \langle X_{it}, \hat{\phi}_{j,t} \rangle \epsilon_{it}$ is reasonably close to $\mathcal{N}(0, n^{-1} \frac{\sigma_\epsilon^2}{\lambda_j})$ if n is sufficiently large. An argument of that type can also be found e.g. in Horváth and Reeder (2012).

As a consequence the law of $\frac{n^{1/2}}{2^{1/2}} \frac{\lambda_{j,t}^{1/2}}{\sigma_\epsilon} (\hat{a}_{j,t} - a_{j,k}) + \frac{n^{1/2}}{2^{1/2}} \frac{\lambda_{j,t}^{1/2}}{\sigma_\epsilon} (a_{j,k} - \hat{a}_{j,s})$ is, neglecting correlations between times t and s, close to $\mathcal{N}(0,1)$. This in turn mean that the law of the approximation introduced in (18), can thus be approximated by a scaled χ_m^2 distribution. Precisely we conclude

$$\frac{n}{2} \hat{\Delta}_{ts} \sim \chi_m^2 \quad \text{approximately.} \tag{19}$$

As was alluded to before, this finding is the reason for employing $\hat{\alpha}_t^{(\Delta)}$ rather than $\hat{\alpha}_t$ for classification. We ultimately suggest to select $\tau_{nT} = \frac{2}{n} q_\alpha(\chi_m^2)$ with $q_\alpha(\chi_m^2)$ the α-quantile of a χ_m^2. The value of α is meant to be very close to one, e.g. $\alpha = 0.99$.

This choice is heuristic in two ways. On the one hand there is an approximation error indicated in (19), which is not explicitly taken into account. On the other hand, even if the exact law of $\hat{\Delta}_{ts}$ was known, quantiles can only serve as an approximation to an ideal threshold satisfying the condition in Assumption 7.1. This is because such ideal threshold should be larger than a maximum $\hat{\Delta}_{ts}$, still given $t, s \in G_k$, which was unbounded if its law has unbounded support. Choosing an α-quantile should naturally lead to a fraction of $1 - \alpha$ misclassified periods. Noting that in practice this might still be well acceptable, we set $\tau_{nT} = \frac{2}{n} q_{0.99}(\chi_m^2)$ in the simulation as well as our application.

Simulations

In order to examine the finite sample performance of the suggested estimation procedure, we run a simulation study. Beyond varying sample sizes, the study is also designed to explore the behavior of the estimates in for more or less distinguishable regimes.

In order to create comparability to the literature we use a modified version of the data generating process from Shin (2009) adapted to our setup. We decided to simulate model (1) in two scenarios. In both scenarios there are $K = 2$ parameter regimes. In the first regime the distance between regime specific functions is substantially greater than in the second scenario. We set $\alpha_t = A_1$ if $t \leq T/2$ and $\alpha_t = A_2$ if $t > T/2$. In both scenarios $A_2 = -s2 + 8s^2 + 5s^3 + 2\sin(8s)$. For the parameter function A_1 we choose

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

as depicted in Figure 1. Numerically the difference $||A_1 - A_2||_2^2$ amounts to approximately 40.1 in the first and 6.6 in the second scenario.

We simulate the regressor z and the error according to $z_{it} \sim \mathcal{N}(0,1)$, $\epsilon_{it} \sim \mathcal{N}(0,1)$. Further parameterizations are (in both scenarios) $\beta_t = 5\sin(t/\pi)$ for the slopes and $\hat{\alpha}_{0,t} = 5\cos(t/\pi)$ for the intercepts. The trajectories X_{it} is simulated according to a Brownian motion, however without dependencies over index t. To do so, we draw scores $\theta_{it,j} := \langle X_{it}, \phi_j \rangle$ independently from a $\mathcal{N}\left(0, [(j-1/2)\pi]^{-2}\right)$-distribution for $1 \leq j \leq 20$. Corresponding eigenfunctions are $\phi_j(s) = \sqrt{2}\sin((j-1/2)\pi s), 1 \leq j \leq 20$ and the trajectories are obtained as $X_{it}(s) = \sum_{j=1}^{\infty} \theta_{it,j}\phi_j(s)$. The trajectories are evaluated on an equidistant grid made up of 30 points in the unit interval.

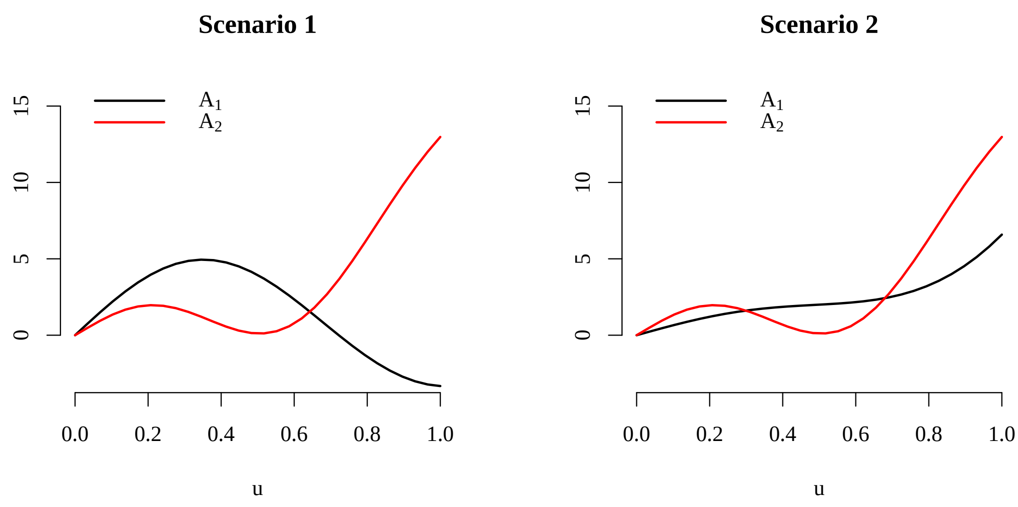

Figure 1: *Regime parameter functions in the different scenarios.*

We explore 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. In the estimation the tuning parameters were set to $m = 4$ and $\tilde{m} = 6$.

Regarding classification the only interest is in recovering A_1 and A_2. Hence we only consider two of the estimated regimes, say w.l.o.g. $\hat{G}_1$, $\hat{G}_2$. G_1 is chosen to be the regime estimate having the largest intersection with the set $\{1, \ldots, T/2\}$ and G_2 is chosen to be the regime estimate having the largest intersection with the set $\{T/2+1, \ldots, T\}$. In order to measure the precision of classification we calculate a classification error as the number of incorrectly classified periods t divided by T. Simulation results are reported in table 1.

In both scenarios estimation of parameters as well as classification benefits as desired from increasing n as well as increasing T. Classification is, for medium and large sample sizes, quite accurate. The algorithm delivers as expected a better performance in the first scenario but also in the second

$(\mathbf{n}, \mathbf{T}) = (\mathbf{50}, \mathbf{50})$

Scenario	**1**					**2**				
	$q_{0.25}$	$q_{0.5}$	$\bar{x}$	$q_{0.75}$	s	$q_{0.25}$	$q_{0.5}$	$\bar{x}$	$q_{0.75}$	s
$T^{-1}\sum_{t=1}^{T}(\tilde{\beta}_t - \beta_t)^2$	1.57	2.36	2.88	3.54	1.95	2.02	3.12	4.20	5.15	3.55
$T^{-1}\sum_{t=1}^{T}(\hat{\alpha}_{0,t} - \alpha_{0,t})^2$	0.28	0.33	0.33	0.38	0.08	0.30	0.35	0.36	0.41	0.08
Classification Error	0.10	0.16	0.16	0.22	0.08	0.20	0.26	0.27	0.32	0.09
$\|\|\tilde{A}_1 - A_1\|\|_2^2$	0.89	1.61	2.38	2.95	2.62	1.23	2.28	3.32	4.18	3.76
$\|\|\tilde{A}_2 - A_2\|\|_2^2$	1.66	3.00	4.67	5.40	5.44	2.28	3.94	5.35	6.49	6.57

$(\mathbf{n}, \mathbf{T}) = (\mathbf{100}, \mathbf{50})$

Scenario	**1**					**2**				
	$q_{0.25}$	$q_{0.5}$	$\bar{x}$	$q_{0.75}$	s	$q_{0.25}$	$q_{0.5}$	$\bar{x}$	$q_{0.75}$	s
$T^{-1}\sum_{t=1}^{T}(\tilde{\beta}_t - \beta_t)^2$	0.58	0.84	1.11	1.30	0.90	0.76	1.14	1.49	1.82	1.10
$T^{-1}\sum_{t=1}^{T}(\hat{\alpha}_{0,t} - \alpha_{0,t})^2$	0.14	0.17	0.17	0.19	0.04	0.15	0.18	0.18	0.20	0.04
Classification. Error	0.10	0.14	0.16	0.22	0.08	0.14	0.22	0.23	0.30	0.11
$\|\|\tilde{A}_1 - A_1\|\|_2^2$	0.31	0.51	0.65	0.83	0.51	0.41	0.72	0.96	1.23	0.85
$\|\|\tilde{A}_2 - A_2\|\|_2^2$	0.72	1.20	1.59	1.92	1.51	0.91	1.49	1.86	2.31	1.55

$(\mathbf{n}, \mathbf{T}) = (\mathbf{150}, \mathbf{80})$

Scenario	**1**					**2**				
	$q_{0.25}$	$q_{0.5}$	$\bar{x}$	$q_{0.75}$	s	$q_{0.25}$	$q_{0.5}$	$\bar{x}$	$q_{0.75}$	s
$T^{-1}\sum_{t=1}^{T}(\tilde{\beta}_t - \beta_t)^2$	0.46	0.70	1.00	1.23	0.90	0.61	0.97	1.31	1.74	1.05
$T^{-1}\sum_{t=1}^{T}(\hat{\alpha}_{0,t} - \alpha_{0,t})^2$	0.10	0.11	0.11	0.13	0.02	0.10	0.12	0.12	0.13	0.02
Classification. Error	0.09	0.14	0.15	0.20	0.08	0.11	0.18	0.19	0.25	0.10
$\|\|\tilde{A}_1 - A_1\|\|_2^2$	0.13	0.20	0.26	0.32	0.21	0.21	0.35	0.44	0.55	0.35
$\|\|\tilde{A}_2 - A_2\|\|_2^2$	0.51	0.74	0.90	1.09	0.67	0.52	0.76	0.90	1.11	0.57

Table 1: **Estimation errors.** The quantities $q_{0.25}$, $q_{0.5}$, $q_{0.75}$,denote the 25% 50% and 75% quantiles of the empirical distribution over Monte Carlo samples. $\bar{x}$ and s denote corresponding arithmetic mean and standard deviation.

scenario classification is fairly accurate. In the case of a small sample size $(n, T) = (50, 50)$ the classification error is still on a reasonable level, though with up to 27% on average substantial. This underlines that, in lines with the provided theory, the estimation procedure is particularly useful in a $n/T >> 1$ setup.

Estimation of $\alpha_{0,t}$ seems to be quite accurate as well and improves with increasing sample size as expected. The estimation of the β_ts seems to deliver noisier results than the remaining quantities. Nevertheless the error obviously decreases as n increases, while the classification error seems to contribute only little to the overall error.

Estimation of regime specific parameter functions is as expected more precise in the first scenario. Averages of squared L^2-losses differ up to almost one unit over scenarios for A_1 and the small sample size $(50, 50)$. However for larger sample sizes the errors all go down as desired and also the magnitude of the difference between scenarios shrinks substantially.

Figures 5-10 in Appendix B show pointwise confidence intervals of the estimated parameter functions of both regimes. The figures illustrate the tightening of such confidence bands when the sample size increases. However towards the left and right boundaries of the domain our procedure seems to deliver implausible estimates. At the left boundary the bands' widths all shrink to zero, while estimates seem to suffer from a bias at the right boundary. In the region beyond 0.9 pointwise confidence intervals even do not include the true parameter functions in some cases. Finally this problematic behavior seems to be of local nature. Estimates in the interior appear to be quite accurate.

Regime Dependent Pricing of Idiosyncratic Risk

Emerging from the influential work of Ang et al. (2006) a considerable number of studies confirmed a negative relation between idiosyncratic volatility and (cross sectional) stock returns. This finding is puzzling as asset pricing theory suggests two opposite views. 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 compensating for bearing risk. Starting from theory it would be most reasonable to expect a negative relation between idiosyncratic volatility and stock returns. As demonstrated in Hou and Loh (2016) this puzzle has to a substantial extend remained unsolved.

The statistical framework presented in the previous sections offers a tailor-made solution to get a full detail perspective on this puzzle. This is because the our regression model has at least two important advantages over the majority of existing approaches.

1. Our model allows to incorporate a high resolution volatility object and allows to isolate its effect in the presence of additional pricing relevant variables.
2. The suggested regression model allows to get a more differentiated view of risk pricing. Heterogeneity appears not only across regimes, which allow to separate puzzling from non-puzzling periods.[4] It is also inherently present for any i, t pair: the functions $\tilde{A}_k$ might be positive for some and negative for other time intervals.

We prepare confronting our approach with data by reviewing and adapting a theoretical pricing model. Such model serves as the economic counterpart to our regression model. The pricing model also offers an advice how to gather a high resolution volatility object from discrete price data sampled at high frequencies. For this purpose we adapt the work of Müller et al. (2011) to our context. We subsequently present our empirical study.

Risk Pricing in Continuous Time & Volatility-Curve Estimation

As a central building block the price dynamics should be related to idiosyncratic volatility in continuous time. The stochastic volatility framework of Barndorff-Nielsen and Shephard (2002)

[4]See e.g. Anderson (2011) for the notion of time varying risk premiums.

offers an excellent starting point for this purpose. The key aspects of their work are the specification of the (i) asset price and (ii) volatility dynamics. We address both such aspects in our context as well, however the second one in an empirical environment.

For our purpose, consider a market made up of n assets, labeled $1 \leq i \leq n$. The market is open at periods $1 \leq t \leq T$ and trading time is standardized to the unit interval $[0,1]$ in each period. Denote the price of asset i in period t at time $s \in [0,1]$ as $P_{it}(s)$. We suggest to model the corresponding log price dynamics according to

$$\mathrm{d}\log\left(P_{it}(s)\right) = \mu_{it}(s)\mathrm{d}s + \alpha_t(s)x_{it}^2(s)\mathrm{d}s + x_{it}(s)\mathrm{d}W_{it,s}. \tag{20}$$

The quantities μ and W_s denote a stochastic drift and a Wiener process. The object $x_{it}(s)$ is the stochastic instantaneous volatility and $\alpha_t(s)$ an nstantaneous marginal risk premium. Such premium is assumed to be non-stochastic but time varying.

(20) can be viewed as a multiperiod-multiasset version of the log-price process in Barndorff-Nielsen and Shephard (2002), who do not incorporate variation of $\alpha_t(s)$ over index s. In the context of risk pricing, we believe that period dependence might be of central relevance as we will argue in the empirical study.

The dynamics in (20) lead to a natural expression for period log-returns. These returns are obtained according to

$$y_{it} := \log\left(\frac{P_{it}(1)}{P_{it}(0)}\right) = \gamma_{it} + \int_0^1 \alpha_t(v)x_{it}^2(v)\mathrm{d}v + \varepsilon_{it}. \tag{21}$$

with $\gamma_{it} := \int_0^1 \mu_{it}(v)\mathrm{d}v$, and $\varepsilon_{it} := \int_0^1 x_{it}(v)\mathrm{d}W_{it,v}$. In (21) the aggregate $\int_0^1 \alpha_t(v)x_{it}^2(v)\mathrm{d}v$ is interpreted as the absolute idiosyncratic risk premium, being one constituent of the period return. Beyond that we define the associated *marginal risk premium* the be the object $\int_0^1 \alpha_t(v)\mathrm{d}v$.

In order to make the pricing model empirically relevant, two more steps are required. The first one relates to the construction of actually latent curves x_{it} from densely sampled price data. The second step relates to the introduction of further pricing relevant variables.

Regarding the former aspect, Müller et al. (2011) offer a straightforward solution. In a closely related framework they introduce a procedure to estimate curves $\log(x_{it}^2)$ from prices P_{it}, sampled on an equidistant grid $[0, \Delta, 2\Delta, \ldots, 1]$. At the core of this procedure is an approximation

$$\log\left(Y_{it,\Delta}(s)^2\right) + c_0 \approx \log(x_{it}^2(s)) + e_s, \tag{22}$$

with e_s being an error term, $c_0 \approx 1.27$ and $Y_\Delta(s) := \Delta^{-\frac{1}{2}}\log\left(\frac{P(s+\Delta)}{P(s)}\right)$. Under suitable assumptions, this approximation remains valid in our environment as we argue along with more details in Appendix C. Based on these discrete values $\log\left(Y_{it,\Delta}(s)^2\right)$ log-volatility curves, can be reconstructed using suitable techniques. We address this curve construction from a practical viewpoint in the empirical study. Given there is a suitable estimate of the curve, say X_{it}, available, one should note that this object might be employed in (21) instead of x_{it}, introducing an error constituted by a time specific constant and a remainder term from a Tailor expansion, as we argue in the Appendix.

The pricing mechanism in (21) is well suited to theoretically isolate risk pricing. Classical asset pricing based on fundamentals but also recent empirical research however, highlighted the importance of additional factors. For this purpose, we postulate that the aggregate drift collects such effects according to $\gamma_{it} = \alpha_{0,t} + \beta_t^T z_{it}$ borrowing the notation from the previous sections. The constant $\alpha_{0,t}$ might also collect nonrandom terms emerging from the approximation of $\int_0^1 \alpha_t(v) x_{it}^2(v) \mathrm{d}v$ by $\int_0^1 \alpha_t(v) X_{it}(v) \mathrm{d}v$.

Taking together the two indicated aspects leads to an empirically relevant pricing model[5]

$$y_{it} = \alpha_{0,t} + \beta_t^T z_{it} + \int_0^1 \alpha_t(v) X_{it}(v) \mathrm{d}v + \epsilon_{it}. \tag{23}$$

Model (23) might be confronted with data on y, X and z right away. However, adding the regime structure we introduced in the previous sections offers a valuable extension: it allows to uncover the degree of heterogeneity in the pricing mechanism. E.g. it enables the researcher to perform a clear cut separation of times, in which risk pricing matters at all from periods in which there it is not priced. Addressing the idiosyncratic volatility puzzle a meta structure on the α_t provides a much more detailed understanding of the timing of puzzling periods.

The Idiosyncratic Risk Puzzle: Evidence from the US Stock Market

Starting from the above pricing framework, our empirical study addresses the idiosyncratic volatility puzzle for the US stock market. The implementation we suggest is particularly influenced by two aspects in the literature. First, as emphasized e.g. in Fu (2009) a contemporaneous relation between idiosyncratic volatility, from now just *IVOL*, and stock returns is a theoretical artifact. This is because IVOL is a priori unknown to investors. It is more reasonable to assume that pricing is based on conditional expectations about IVOL in the previous period. In the model framework this is just to interpret x_{it} as a conditional expectation. In our empirical study we roughly follow Fu (2009) and model such expectations explicitly. Beyond that a second important point is that the meta study of Hou and Loh (2016) alludes to the potential value of controlling for market frictions. We thus include investors conditional expectations about market frictions in our model. More concrete we choose z_{it} to be the average bid-ask spread in a period t.

Investors conditional expectations are straight forward incorporated according to

$$y_{it} = \int_0^1 \alpha_t(s) E_{t-1}[X_{it}](s) \mathrm{d}s + \beta_t E_{t-1}[z_{it}](s) + \epsilon_{it}. \tag{24}$$

[5]Note that the errors on (21) and (23) do not coincide due to the approximations in the risk premium.

The quantities $E_{t-1}[X_{it}]$ and $E_{t-1}[z_{it}]$ are expectations given the information in $t-1$. We model these expectations using the autoregressive models

$$X_{it}(u) = \int_0^1 Q_i^{(X)}(u,v)X_{i,t-1}(u)\mathrm{d}u + e_1(u) \tag{25}$$

$$\text{and} \quad z_{it}(u) = Q_i^{(z)} z_{i,t-1} + e_2. \tag{26}$$

In order to estimate the above models, we consider data for $n = 377$ S&P 500 constituents recorded over $T = 136$ full trading days between $2016-06-03$ and $2016-12-15$. Beyond asset prices, which are sampled at a frequency of 10 minutes, we employ a daily bid-ask, z_{it}, spread as we alluded to before. Using a straightforward estimation procedure based on the methodology outlined above, we are able to construct daily log-volatility trajectories for each i, t combination. These IVOL-trajectories are employed as regressors X_{it}, while y_{it} is daytime return at day t. For analyzing the regimes we also use a daily average of the VIX-index in the period of interest as well as daily values of the economic policy uncertainty index in the US (EPU) of Baker et al. (2016). All information but the EPU is extracted from Bloomberg. Table 3 in Appendix D, provides summary statistics for the available sample.

The functional autoregressions in (25) are estimated, for time series data on each stock i separately, using the implementation in the `far` package in `R`. The dimension of the subspace in which the operators $Q_i^{(X)}$ are determined by cross validation. This yields asset specific model complexities. The one-trading-day ahead forecasts from these models are used as estimates of the expectations $E_{t-1}[X_{it}]$. In analogy, one-step-ahead forecasts gathered from (26) are used to estimate $E_{t-1}[z_{it}]$. Parameter estimates are obtained again for each i separately from the corresponding time series using Gaussian Maximum Likelihood estimators. In what follows we use $\tilde{E}_{t-1}[X_{it}]$ and $\tilde{E}_{t-1}[z_{it}]$ to denote the one-step-ahead forecasts.

In order to come from the discrete IVOL values to the curves X_{it}, we proceed as follows. First we rescale intraday time to the interval $[0, 1]$, with zero being 09:30 a.m. and 1 being 03:40 p.m.. The curves X_{it} are constructed on this interval using a simple Nadaraya-Watson estimator applied all observations also this including after 03:40 p.m. in order to avoid boundary problems, which are present at the right end of the interval.[6]

The estimation of the regression model (24) proceeds as indicated in sections 3 and 5. The regressors, $\tilde{E}_{t-1}[X_{it}]$ and $\tilde{E}_{t-1}[z_{it}]$ are for each t centered around their cross section means. A summary of the ultimate regression results are presented in table 4 in the Appendix. The procedure partitions the set of 135 trading days[7] into 14 regimes. The regime specific estimates $\tilde{A}_k$ are depicted in Figure 2.

Beyond the estimated curves we calculate a marginal effect according to $\int_0^1 \tilde{A}_k(u)\mathrm{d}u$ and an absolute effect, i.e. an idiosyncratic risk premium, according to $\int_0^1 \tilde{A}_k(u)\tilde{E}_{t-1}[X_{it}]\mathrm{d}u$. In order to facilitate the access to differences of such 14 regimes we sort the regimes according to their marginal effect into three groups. This is reflected by the color scheme in the plots.

[6]By this we aim to prevent problematic behavior of the estimates as identified in the simulation study.
[7]One day is lost due to the forecasting of (f)AR(1) models.

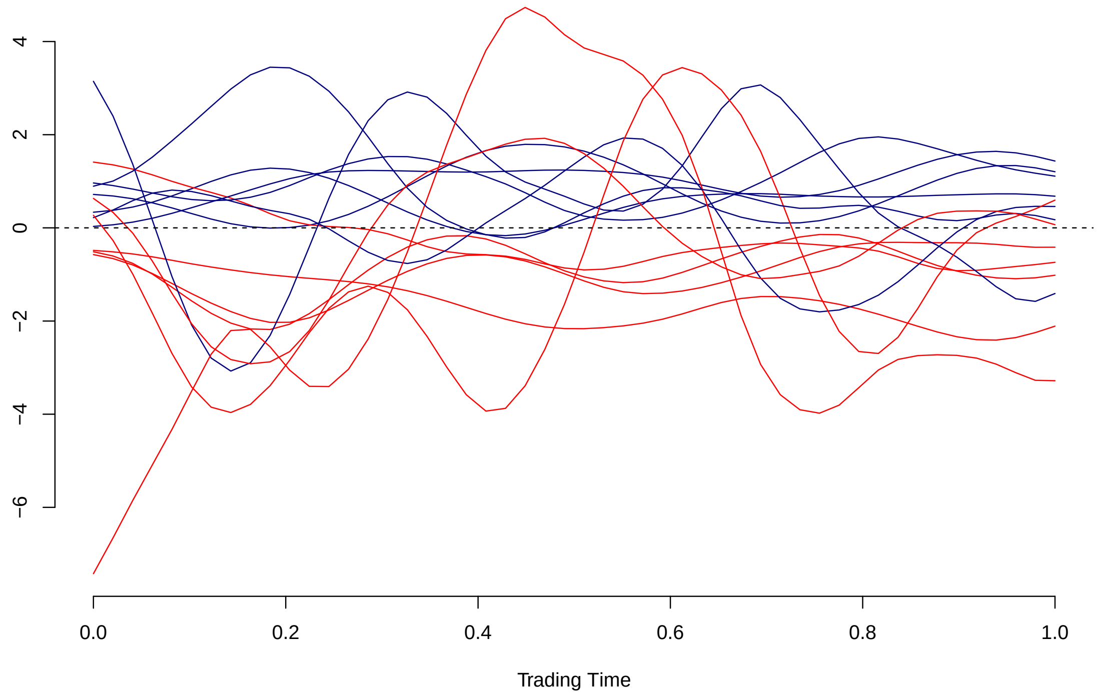

Figure 2: *Estimated Regime Specific Curves $\tilde{A}_k$. Blue curves have positive values of $\int_0^1 \tilde{A}_k(u)\mathrm{d}u$, red curves negative values.*

As reported in table 4 in Appendix D, there are five regimes with 12 or more trading days and in total 7 Regimes with more than a single day. Interestingly the slope coefficients $\tilde{\beta}_{t,k}$ are in tendency smaller for groups constituted by a single trading day. From this perspective regimes also distinguish days at which the pricing of trading days is more important than at other days.

The marginal effects $\int_0^1 \tilde{A}_k(u)\mathrm{d}u$ can be split in two groups. For one set this effect is negative, for the other one positive. Due to centering of the functional regressors the absolute effects are by construction close to zero. The amount of dispersion of their distribution is depicted in the lower panel of figure 3 and remains comparably stable over time. In tendency regimes having a smaller marginal effect show a higher dispersion in risk premiums: separation in regimes thus distinguishes periods of high pricing heterogeneity from periods of homogeneous pricing of idiosyncratic risk.

An obvious question is, whether other risk factors are informative for the pricing of idiosyncratic risk. As the covered time span captures the *BREXIT* vote (23-06-2016) as well as the US presidential elections (08-11-2016) at least two important political events might influence the pricing. We thus consider the daily economic policy uncertainty index for the US (EPU) as introduced by Baker et al. (2016) to measure uncertainty in the relevant political dimension. Beyond that we employ a daily average of the Volatility Index, VIX, as a proxy for market uncertainty. In order to see, whether these factors are in any sense informative for the the occurrence of identified risk regimes we (i) calculate correlations and (ii) employ a support vector machine[8] (SVM). Cross correlation between

[8]See e.g. Chapter 12 in Hastie et al. (2009)

k/k	1	2	3	5	6	7	8
1	0.07	0.07	0.13	0.53	0.13	0.00	0.07
2	0.14	0.46	0.11	0.04	0.14	0.04	0.04
3	0.06	0.61	0.06	0.06	0.11	0.00	0.00
5	0.08	0.42	0.25	0.00	0.25	0.00	0.00
6	0.10	0.30	0.25	0.05	0.10	0.05	0.00
7	0.00	0.33	0.00	0.00	0.67	0.00	0.00
8	0.00	1.00	0.00	0.00	0.00	0.00	0.00

Table 2: **Empirical Transition Frequencies: Seven Largest Regimes.** The table reports the first seven columns&rows of the matrix of transition frequencies.

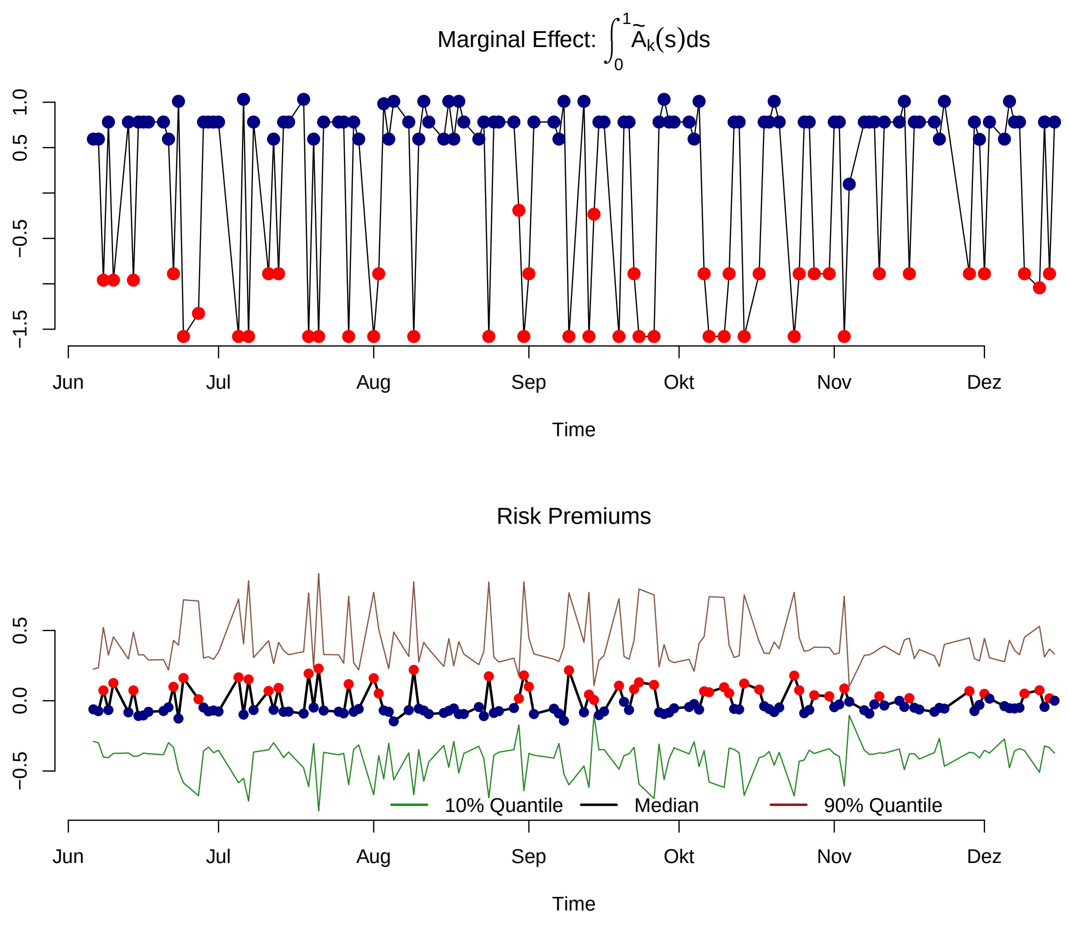

Figure 3: *Upper Panel: Values of* $\int_0^1 \tilde{A}_k(u)\mathrm{d}u$ *over time. Lower Panel: Quantiles of empirical distribution of* $\int_0^1 \tilde{A}_k(u)\tilde{E}_{t-1}[X_{it}]\mathrm{d}u$ *within a regime over time. Blue dots indicate positive values of marginal effects, red dots negative values.*

the EPU or VIX and marginal effects or median risk premiums are overall negligible as can be seen in figure 11 in Appendix D. In order to assess the relation between the EPU / VIX and the regime structure from a different point of view we run the following classification strategy. First, we sort the regimes according to their norm in two groups: one for which the marginal effect as defined before is positive and another one for which it is negative. Then we run a SVM with using VIX and EPU as inputs and the binary group indicator as output[9]. The classification, i.e. training, error

[9]We used the default implementation in R's `kernlab` package. See Zeileis et al. (2004) for details.

amounts to approximately 28% with a number of 116 support vectors. The classification hence appears to be fairly poor. Together with our linear correlation analysis, this finding supports the view that there is only little predictive content for the regimes present in VIX and EPU.

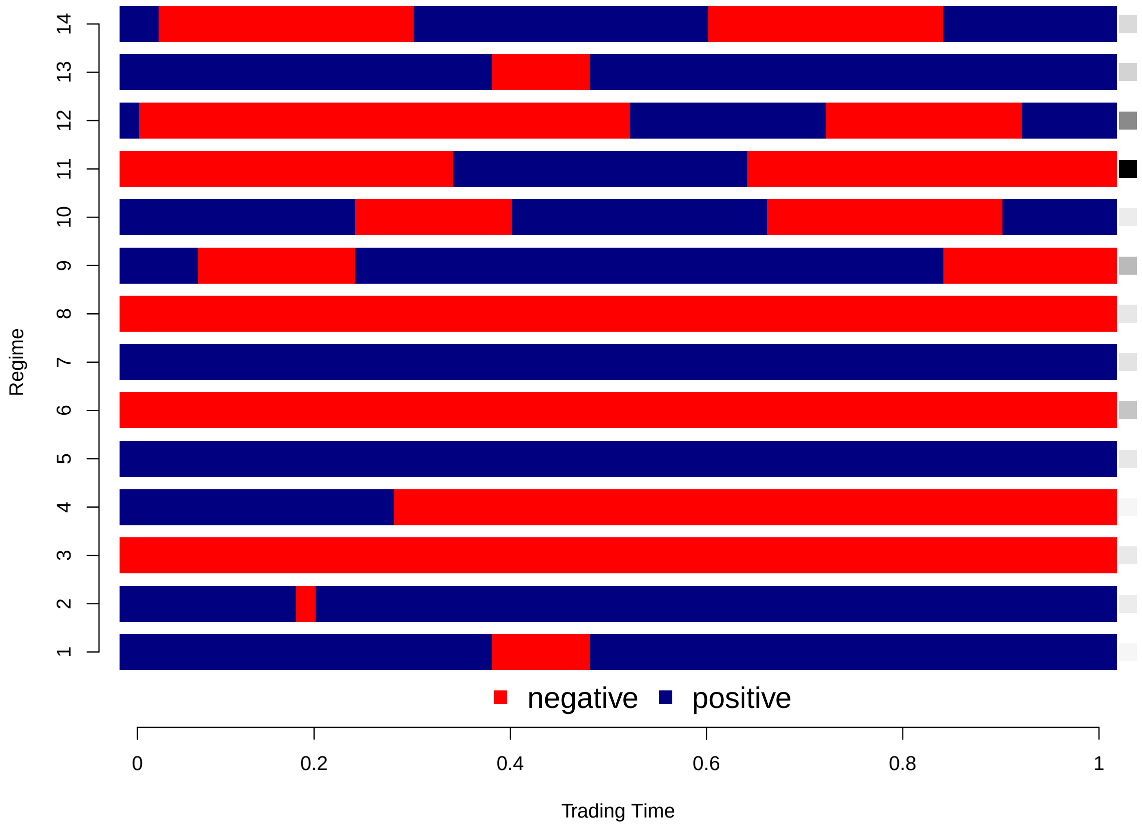

Figure 4: *Figure shows in which regions of the domain $[0,1]$ estimated parameter functions $\tilde{A}_k$ take positive or negative values. The transparanya of the squares at the right margin are proportional to $||\tilde{A}_k||_2^2$: less transparency means larger values.*

The idiosyncratic volatility puzzle is formulated in terms of correlations between scalar objects. Our high resolution volatility estimate is functional and thus offers a richer relation to the scalar returns. The marginal effect as we have calculated it before is an overall measure of the direction of the relation. In order to provide a more detailed view of such directions figure 4 indicates the directions depending on the trading time.

For 5 regimes the function $\tilde{A}_k$ has the same sign over the whole domain. For the remaining regime functions, including the largest regime, the sign changes at least once. It appears as if there is no striking pattern at which subdomain positive or negative values are dominating. For the majority of regimes, there are regions within a day, for which risk is priced in the way theory would suggest, however there also puzzling regions. The functional nature of our volatility estimate as well as the distinction between regimes thus allows to uncover the transition between times of puzzling and theory conform risk pricing. However, controlling for market frictions we find a comparably small impact of IVOL on the returns, which pushes the finding towards the theory.

Conclusion

In this paper we present a novel regression framework, allowing to examine regime specific effects of a random function on a scalar response in the presence of a parametric nuissance term. The suggested estimation procedure is designed for a longitudinal data context. We prove consistency of the estimators and provide rates of convergence. In summary our framework offers a highly flexible and data driven way of assessing heterogeneity in large panels.

The economic relevance of our semiparametric model is underlined by an application to idiosyncratic risk pricing in the presence of frictions. In an application we show that the model is absolutely compatible with a theoretical continuous time pricing model. Confronted with high frequency stock level data from the US, we find that the idiosyncratic volatility puzzle is not a global but rather a temporary phenomenon. The dominance of heterogeneity in the pricing, within as well as across periods, seems to dominate the mechanism.

Our approach might be extended in multiple directions for further research. We believe that relations to the literature on domain selection[10] might be particularly valuable. Another interesting direction concerns the estimation procedure: different approaches, like e.g. the one in Su et al. (2016), are quite promising. Beyond that adjustments to our procedure might be useful as well. For example a less heuristic choice of the threshold, e.g. by a Bootstrap, could improve the performance of the classification algorithm.

[10]See e.g. Hall and Hooker (2016)

Appendix

In part A of the appendix we provide proofs for Lemmas 1 and 2 as well as Theorems 1 and 2 as well as remarks concerning the theoretical aspects of our work. Part B is concerned with details of the simulation study while part C offers results and details of the empirical study.

A. Proofs

Throughout this section we use the symbols C and c to denote constants. Their precise meaning varies, in most parts, from term to term.

Beyond that 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 norm as $|||F_2||| := ||f_2||_2$, where in this case $|| \cdot ||_2$ is the usual L^2-norm in $L^2([0,1] \times [0,1])$.

For the sake of readability we will proof the Lemmas and Theorems for $P = 1$, while the generalization to $P > 1$ is straightforward and does not add any 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 $1 \leq t \leq T$. Adding Assumption 2 to that guarantees that Lemmas A.1-A.5 in Shin (2009) hold for our corresponding cross section estimates as $n \to \infty$. As a consequence of their results and the results in Hall and Hosseini-Nasab (2006) it holds for any $1 \leq t \leq T$ that[11]

$$E\left[\left|\left|\hat{K}_{zX,t} - K_{zX,t}\right|\right|_2^2\right] = O(n^{-1}) \tag{27}$$

$$E\left[\left|\hat{K}_{z,t} - K_{z,t}\right|^2\right] = O(n^{-1}) \tag{28}$$

$$\left|\left|\hat{\Phi}_t - \Phi\right|\right|_{H'}^2 = O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right) \tag{29}$$

as well as

$$E\left[\left|\left|\hat{K}_{X,t} - K_{X,t}\right|\right|_2^2\right] \leq Cn^{-1} \tag{30}$$

$$E\left[\left|\hat{\lambda}_{j,t} - \lambda_j\right|^q\right] \leq Cn^{-q/2} \tag{31}$$

$$E\left[\left|\left|\hat{\phi}_{j,t} - \phi_j\right|\right|_2^q\right] \leq Cn^{-q/2} j^{q(1+\mu)}, \tag{32}$$

for $1 \leq j \leq m$ and $q = 1, 2, \ldots$. Beyond that simple moment calculations yield

[11] As a consequence of the stationarity postulated in Assumption 1, none of the six expectations listed below depends on t.

$$|\bar{z}_t - E[z_{it}]|^2 = |\bar{z}_t|^2 = O_p(n^{-1}) \tag{33}$$
$$||\bar{X}_t - E[X_{it}]||_2^2 = ||\bar{X}_t||_2^2 = O_p(n^{-1}), \tag{34}$$

where $\bar{X}_t := n^{-1}\sum_{i=1}^n X_{it}$. The central result

$$||\hat{\Phi}_t - \Phi||_{H'}^2 = O_p(n^{\frac{1-2\nu}{\mu+2\nu}}) = o_p(1) \tag{35}$$

from Shin (2009) transfers to our setup for any $1 \leq t \leq T$, as for every t the cross section of random variables is independent over index i.

Proof of Lemma 1

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

$$\begin{aligned}\hat{\beta}_t &= [\hat{K}_{z,t} - \hat{\Phi}_t(\hat{K}_{zx,t})]^{-1}[\hat{K}_{z\breve{y},t} - \hat{\Phi}_t(\hat{K}_{x\breve{y},t})] \\ &= \hat{\beta}_t^{(S09)} + \hat{B}_t^{-1} r_{\beta,t}.\end{aligned}$$

with $r_{\beta,t} := [(\hat{\alpha}_{0,t} - \alpha_{0,t})(\bar{z}_t - \hat{\Phi}_t(\bar{X}_t))]$ and $\hat{\beta}_t^{(S09)}$ being the cross section estimator for β_t in model (1) with $\alpha_{0,t} = 0$ presented in Shin (2009). From this work it is further immediate that

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

as $n \to \infty$, which certainly implies $\hat{B}_t^{-1} = O_p(1)$ since $B > 0$ by regularity Assumption 6. It thus remains to show that $r_{\beta,t}$ is at least of order $n^{-1/2}$. To do so note that due to independence of the random variables in the cross section it holds that

$$\begin{aligned}
E\left[(\hat{\alpha}_{0,t} - \alpha_{0,t})^4\right] &= n^{-4}E\left[\left(\sum_{i=1}^n y_{it} - E[y_{it}]\right)^4\right] && (36)\\
&\leq n^{-4}2E\left[\left(\sum_{i=1}^n (y_{it} - E[y_{it}])^2\right)^2\right] + n^{-4}2E\left[\left(\sum_{i=1}^n \sum_{j\neq i}(y_{it} - E[y_{it}])(y_{jt} - E[y_{jt}])\right)^2\right] && (37)\\
&\leq 2n^{-2}E[(y_{it} - E[y_{it}])^4] + 2n(n-1)E[(y_{it} - E[y_{it}])^2]^2 && (38)\\
&= O(n^{-2}) && (39)
\end{aligned}$$

which follows from straightforward moment calculations using the fact that

$$E\left[(y_{it}-E[y_{it}])^4\right] = 27\left[E[\langle X_{it},A_k\rangle^4]+E[z_{it}^4]\beta_t^4+E[\epsilon_{it}^4]\right]$$
$$\leq 27\left(E[||X_{it}||_2^4]||A_k||_2^4+E[z_{it}^4]C_\beta^2+E[\epsilon_{it}^4]\right)<\infty.$$

This implies the first statement of the Lemma. Since $\bar{z}_t = o_p(1)$ and $||\bar{X}_t||_2 = o_p(1)$ as argued before, it holds that $r_{\beta,t} = o_p(n^{-1/2})$, because $|\hat{\Phi}_t(\bar{X}_t)| \leq ||\hat{\Phi}_t||_{H'} \cdot ||\bar{X}_t||$ and the boundedness of Φ.[12] Together with Theorem 3.1 in Shin (2009) the second result of the Lemma follows.

Turning to the third claim formulated in the Lemma, the basis coefficient estimates $\hat{a}_{j,t}$ might be written as $\hat{a}_{j,t} = \hat{a}_{j,t}^{(S09)} - r_{a,jt}$ where

$$r_{a,jt} := \hat{\lambda}_{j,t}\frac{1}{n}\sum_{i=1}^n \langle X_{it},\hat{\phi}_{j,t}\rangle\left(z_{it}\hat{B}_t^{-1}r_{\beta,t}+(\hat{\alpha}_{0,t}-\alpha_{0,t})\right).$$

Hence for the estimator $\hat{\alpha}_t$ it holds that

$$||\hat{\alpha}_t-\alpha_t||_2^2 \leq 2||\hat{\alpha}_t^{(S09)}-\alpha_t||_2^2+2\sum_{j=1}^m r_{a,jt}^2.$$

Define the event

$$\mathcal{F}_{m,t} := \left\{\left|\hat{\lambda}_{j,t}-\lambda_j\right| \leq (2C_\lambda)^{-1}j^{-\mu}: \quad 1\leq j\leq m\right\} \tag{40}$$

and note that due to consistency of the eigenvalue estimators $\mathbb{P}(\mathcal{F}_{t,m}) \to 1$ as $n\to\infty$ for every $1\leq t\leq T$ as a consequence of the results in the proof of Lemma 2. Now, provided $\mathcal{F}_{t,m}$ holds,

$$\begin{aligned}
\sum_{j=1}^m \hat{\lambda}_{j,t}^{-2}\left[\frac{1}{n}\sum_{i=1}^n\langle X_{it},\hat{\phi}_{j,t}\rangle\left(z_{it}\hat{B}_t^{-1}r_{\beta,t}\right)\right]^2 \leq& 4\sum_{j=1}^m\lambda_j^{-2}\left[\frac{1}{n}\sum_{i=1}^n\langle X_{it},\hat{\phi}_{j,t}\rangle\left(z_{it}\hat{B}_t^{-1}r_{\beta,t}\right)\right]^2\\
=&4\sum_{j=1}^m\lambda_j^{-2}\left[\langle K_{zX},\phi_j\rangle+\langle\hat{K}_{zX,t}-K_{zX},\hat{\phi}_{j,t}\rangle+\langle K_{zX},\hat{\phi}_{j,t}-\phi_j\rangle\right]^2\hat{B}_t^{-2}r_{\beta,t}^2\\
\leq&12\hat{B}_t^{-2}r_{\beta,t}^2C\sum_{j=1}^m j^{2\mu}(Cj^{-2(\mu+\nu)}+||\hat{K}_{zX,t}-K_{zX}||_2^2\\
&+||K_{zX}||_2^2\cdot||\hat{\phi}_{j,t}-\phi_j||_2^2)
\end{aligned}$$

Due to (32) and the fact that $m^{\frac{1+2\mu}{\mu+2\nu}}n^{-1}=o(1)$, it holds that

[12]See Shin (2009)). We will use the fact that Φ is bounded in what follows without further reference.

$$E\left[\sum_{j=1}^{m} j^{2\mu}||\hat{\phi}_{j,t}-\phi_j||_2^2\right] = O\left(m^{1+2\mu+2(1+\mu)}n^{-1}\right)$$
$$= O\left(n^{\frac{3(1+\mu)-2\nu}{\mu+2\nu}}\right) = o(1).$$

Combining the observation $\hat{B}_t^{-2}r_{\beta,t}^2 = o_p(n^{-1})$ and (27) it follows

$$\sum_{j=1}^{m}\hat{\lambda}_{j,t}^2\left[\frac{1}{n}\sum_{i=1}^{n}\langle X_{it},\hat{\phi}_{j,t}\rangle\left(z_{it}\hat{B}_t^{-1}r_{\beta,t}\right)\right]^2 = o_p(n^{-1}).$$

Together with Theorem 3.2 in Shin (2009) result 3. in the Lemma follows. ■

Proof of Lemma 2

In what follows we ultimately show that the quantities $\hat{\alpha}_t^{\Delta}$ are consistent for α_t^{Δ} 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 1. The remaining parts of the Lemma are shown on the way, as they are required to obtain the result concerning $\hat{\alpha}_t^{\Delta}$.

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. We also use the results in Hall and Horowitz (2007) and Kneip et al. (2016) and, as the fundamental tool, the Markov and Chebyshev inequalities to obtain the following.

Denote $\Delta^{(\Gamma,t)} := |||\hat{\Gamma}_t - \Gamma||| := \int_0^1\int_0^1\left(\hat{K}_{X,t}(u,v) - E[X_{it}(u)X_{it}(v)]\right)^2 \mathrm{d}u\mathrm{d}v^{1/2}$, $\hat{K}_{X\epsilon} := n^{-1}\sum_{i=1}^{n}X_{it}\epsilon_{it}$ and $\hat{K}_{z\epsilon} := n^{-1}\sum_{i=1}^{n}X_{it}\epsilon_{it}$.

$$\sum_{t=1}^{T} \mathbb{P}\left(\Delta^{(\Gamma,t)^2} > c\right) \leq T \frac{E\left[\Delta^{(\Gamma,t)^2}\right]}{c} = O(n^{\delta-1}) = o(1) \tag{41}$$

$$\sum_{t=1}^{T} \mathbb{P}\left(||\hat{K}_{zX,t} - K_{zX}||_2^2 > c\right) \leq T \frac{n^{-1}E\left[||z_{it}X_{it} - E(z_{it}X_{it})||_2^2\right]}{c} = O(n^{\delta-1}) = o(1) \tag{42}$$

$$\sum_{t=1}^{T} \mathbb{P}\left(|\hat{K}_{z,t} - K_z|^2 > c\right) \leq T \frac{n^{-1}E\left[|z_{it}^2 - E(z_{it}^2)|^2\right]}{c} = O(n^{\delta-1}) = o(1) \tag{43}$$

$$\sum_{t=1}^{T} \mathbb{P}\left(||\hat{K}_{X\epsilon,t}||_2^2 > c\right) \leq T \frac{n^{-1}\sigma_\epsilon^2 E\left[||X_{it}||_2^2\right]}{c} = O(n^{\delta-1}) = o(1) \tag{44}$$

$$\sum_{t=1}^{T} \mathbb{P}\left(|\hat{K}_{z\epsilon,t}|^2 > c\right) \leq T \frac{n^{-1}\sigma_\epsilon^2 E\left[z_{it}^2\right]}{c} = O(n^{\delta-1}) = o(1) \tag{45}$$

$$\sum_{t=1}^{T} \mathbb{P}\left(|n^{-1}\sum_{i=1}^{n} \epsilon_{it} - \sigma_\epsilon^2| > c\right) \leq T \frac{n^{-1}E\left[|\epsilon_{it}^2 - \sigma_\epsilon^2|^2\right]}{c} = O(n^{\delta-1}) = o(1) \tag{46}$$

$$\sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \lambda_j^{-2}||\phi_j - \hat{\phi}_{j,t}||_2^2 > c\right) \leq T \frac{n^{-1}C\sum_{j=1}^{m} \lambda_j^{-2} j^{2(1+\mu)} n^{-1}}{c} = O(m^{3+4\mu}n^{\delta-1}) = o(1). \tag{47}$$

Note that the result in (47) is due Assumption 4. In particular (47) implies $\sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} ||\phi_j - \hat{\phi}_{j,t}||_2^2 > c\right) = o(1)$.

Beyond these facts, we show that $\sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c\right)$ is a null sequence, which will be used multiple times later on. Denoting by $\cdot^c$ the complement and employing the Cauchy-Schwartz inequality, the results in Hall and Horowitz (2007) and the references therein, stationarity of the functional regressor in the time dimension as well as Assumption 4 it holds that:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c\right) &\leq T^{-1} \sum_{t=1}^{T} \mathbb{P}\left(\sup_{1\leq j\leq m} |\hat{\lambda}_{j,t} - \lambda_j| > \frac{1}{2}\lambda_m\right) \\
&\leq \sum_{t=1}^{T} \mathbb{P}\left(\Delta^{(\Gamma,t)} > \frac{1}{2}\lambda_m\right) \\
&\leq \sum_{t=1}^{T} \frac{4E\left[(\Delta^{(\Gamma,t)})^2\right]}{\lambda_m^2} \\
&= O(n^{\delta} n^{\frac{\mu-2\nu}{\mu+2\nu}}) = o(1).
\end{aligned}$$

For later purpose, define the in addition to $\mathcal{F}_{m,t}$ the event $\mathcal{S}_t$ according to

$$\mathcal{S}_t := \left\{\left|\hat{\sigma}_{\epsilon,t}^2 - \sigma\epsilon^2\right| \leq \frac{1}{2}\sigma^2 : \quad 1 \leq j \leq m\right\}. \tag{48}$$

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 $\hat{\alpha}_{0,t}$, $\hat{\beta}_t$ and $\hat{\alpha}_t$. $\hat{\alpha}_{0,t}$ can be handled in a straightforward manner due to (39).

We thus begin by showing $\mathbb{P}(\max_{1\leq t\leq T}|\hat{\beta}_t - \beta_t| > c) = o(1)$. 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

$$||\hat{\Phi}_t - \Phi||^2_{H'} = 3R_{0,1} + 3R_{0,2} + 3R_{0,3}.$$

The last summand is defined as $R_{0,3} := \left|\left|\sum_{j=m+1}^{\infty}\frac{\langle K_{zX},\phi_j\rangle}{\lambda_j}\phi_j\right|\right|^2$. It is independent of t and $o(1)$ because the truncation parameter diverges at infinity and hence $R_{0,3}$ is arbitrarily small for n large enough. The remaining summands are defined and treated as follows. The first one behaves according to

$$\begin{aligned}
R_{0,1} :=& \left|\left|\sum_{j=1}^{m}\left(\frac{\langle\hat{K}_{zX,t},\hat{\phi}_{j,t}\rangle}{\hat{\lambda}_{j,t}} - \frac{\langle\hat{K}_{zX,t},\hat{\phi}_{j,t}\rangle}{\hat{\lambda}_{j,t}}\right)\hat{\phi}_{j,t}\right|\right|^2 \\
\leq& 2\sum_{j=1}^{m}(\hat{\lambda}_{j,t}\lambda_j)^{-2}\left[\langle\lambda_j\hat{K}_{zX,t} - \hat{\lambda}_{j,t}K_{zX},\phi_j\rangle + \langle\lambda_j\hat{K}_{zX,t},(\hat{\phi}_{j,t} - \phi_j)\rangle\right]^2 \\
\leq& 4\sum_{j=1}^{m}(\hat{\lambda}_{j,t}\lambda_j)^{-2}\left[\langle K_{zX},\phi_j\rangle^2(\lambda_j - \hat{\lambda}_{j,t})^2 + \langle\hat{K}_{zX,t} - K_{zX},\phi_j\rangle^2\lambda_j^2\right] \\
&+ 2\sum_{j=1}^{m}(\hat{\lambda}_{j,t})^{-2}\langle\hat{K}_{zX,t},(\hat{\phi}_{j,t} - \phi_j)\rangle^2 \\
\leq& \underbrace{4\sum_{j=1}^{m}(\hat{\lambda}_{j,t}\lambda_j)^{-2}\langle K_{zX},\phi_j\rangle^2(\lambda_j - \hat{\lambda}_{j,t})^2}_{=:R_{1,1}} + \underbrace{4\sum_{j=1}^{m}||\hat{K}_{zX,t} - K_{zX}||_2^2\hat{\lambda}_{j,t}^{-2}}_{=:R_{1,2}} \\
&+ \underbrace{2\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\hat{K}_{zX,t}||_2^2||\hat{\phi}_{j,t} - \phi_j||_2^2}_{=:R_{1,3}}.
\end{aligned}$$

The three summands behave as follows.

Ad $R_{1,1}$:

$$
\begin{aligned}
&\sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} (\hat{\lambda}_{j,t}\lambda_j)^{-2} \langle K_{zX}, \phi_j\rangle^2 (\lambda_j - \hat{\lambda}_{j,t})^2 > c\right) \\
\leq &\sum_{t=1}^{T} \mathbb{P}\left(4\sum_{j=1}^{m} (\lambda_j)^{-4} \langle K_{zX}, \phi_j\rangle^2 (\lambda_j - \hat{\lambda}_{j,t})^2 > c\right) + \sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c\right) \\
\leq &\frac{\sum_{j=1}^{m} \lambda_j^{-4} \langle K_{zX}, \phi_j\rangle^2 C n^{-1}}{c/4} + \sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c\right) \\
= &O(n^{\delta} n^{-1}) + o(1) = o(1).
\end{aligned}
$$

Ad $R_{1,2}$:

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

Ad $R_{1,3}$:

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

Since $\sum_{t=1}^{T} \mathbb{P}\left(R_{0,1} > c\right) \leq \sum_{l=1}^{3} \sum_{t=1}^{T} \mathbb{P}\left(R_{1,l} > c\right) +, \sum_{t=1}^{T} \mathbb{P}\left(R_{0,1} > c\right) = o(1)$ follows.

For $R_{0,2}$ note that due to Assumption 4 it holds that

$$
\begin{aligned}
&\sum_{t=1}^{T}\mathbb{P}\left(\left|\left|\sum_{j=1}^{m}\frac{\langle K_{zX},\phi_j\rangle}{\lambda_j}(\phi_j-\hat{\phi}_{j,t})\right|\right|_2^2>c\right)\\
\leq&\sum_{t=1}^{T}\frac{Cm\sum_{j=1}^{m}j^{4(1+\mu)-2\nu}n^{-2}}{c^2}\\
=&O\left(mn^{\delta-2}\right)\\
=&o(1).
\end{aligned}
$$

This set of results implies $\mathbb{P}\left(\max_{1\leq t\leq T}||\hat{\Phi}_t-\Phi||_{H'}^2>c\right)=o(1)$.

Addressing the inverse in $\hat{\beta}_t$, define the event $Q_t:=\{|\hat{B}_t-B|\leq\frac{1}{2}B\}$, where $\hat{B}_t:=\hat{K}_{z,t}-\hat{\Phi}_t(\hat{K}_{zX,t})$ and its population counterpart $B:=K_z-\Phi(K_{zX})>0$. For this event, note that $\sum_{t=1}^{T}\mathbb{P}\left(Q_t^c\right)\leq R_{2,1}+R_{2,2}$, where $R_{2,1}:=\sum_{t=1}^{T}\mathbb{P}\left(|\hat{K}_{z,t}-K_z|^2>c\right)=o(1)$ as shown in (43).

For $R_{2,2}$ note

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

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

$$
\begin{aligned}
R_{3,2}\leq&\sum_{t=1}^{T}\mathbb{P}\left(||\hat{\Phi}_t-\Phi||_{H'}||^2||\hat{K}_{zX,t}-K_{zX}||_2^2>c\right)\\
\leq&\sum_{t=1}^{T}\mathbb{P}\left(||\hat{\Phi}_t-\Phi||_{H'}^2>c\right)+\sum_{t=1}^{T}\mathbb{P}\left(||\hat{K}_{zX,t}-K_{zX}||_2^2>c\right)\\
=&\,o(1)
\end{aligned}
$$

using the same arguments as before. Now restate the difference $\hat{\beta}_t-\beta_t$ as

$$
(\hat{\beta}_t-\beta_t)=\hat{B}_t^{-1}\left(\bar{z}_t(\alpha_{0,t}-\hat{\alpha}_{0,t})+n^{-1}\sum_{i=1}^{n}\hat{V}_{it}+\hat{K}_{z\epsilon,t}-\hat{\Phi}_t(\hat{K}_{X\epsilon,t})-\hat{\Phi}_t(\bar{X}_t)(\alpha_{0,t}-\hat{\alpha}_{0,t})\right),
$$

where $\hat{V}_{it} := (z_{it} - \hat{\Phi}_t(X_{it}))\langle X_{it}, \alpha_t \rangle$ which will later on correspond to $V_{it} := (z_{it} - \Phi(X_{it}))\langle X_{it}, \alpha_t \rangle$. For later purpose we work with the squared difference

$$(\hat{\beta}_t - \beta_t)^2 \leq 5\hat{B}_t^{-2} \left(\bar{z}_t^2(\alpha_{0,t} - \hat{\alpha}_{0,t})^2 + \left(n^{-1} \sum_{i=1}^{n} \hat{V}_{it} \right)^2 + \hat{K}_{z\epsilon,t}^2 + ||\hat{\Phi}_t||_{H'}^2 ||\hat{K}_{X\epsilon,t}||_2^2 + ||\hat{\Phi}_t||_{H'}^2 ||\bar{X}_t||_2^2 (\alpha_{0,t} - \hat{\alpha}_{0,t})^2 \right)$$

And thus

$$\mathbb{P}(\max_{1 \leq t \leq T} (\hat{\beta}_t - \beta_t)^2 > c) \leq R_{4,1} + R_{4,2} + R_{4,3} + R_{4,4} + R_{4,5} + \sum_{t=1}^{T} \mathbb{P}(Q_t^c).$$

While $\sum_{t=1}^{T} \mathbb{P}(Q_t^c) = o(1)$ as shown before the remaining terms can be handled as follows.

$$R_{4,1} := \sum_{t=1}^{T} \mathbb{P}\left(B^{-2}\bar{z}_t^2(\alpha_{0,t} - \hat{\alpha}_{0,t})^2 > c\right) \leq \sum_{t=1}^{T} \frac{E\left[\bar{z}_t^4\right]^{\frac{1}{2}} E\left[(\alpha_{0,t} - \hat{\alpha}_{0,t})^4\right]^{\frac{1}{2}}}{cB^2} = o(1)$$

because of the same arguments as in the Proof of Lemma 1. Note at this point, it also holds, along the same line of reasoning, that

$$\sum_{t=1}^{T} \mathbb{P}\left((\alpha_{0,t} - \hat{\alpha}_{0,t})^2 > c\right) \leq \sum_{t=1}^{T} \frac{E\left[(\alpha_{0,t} - \hat{\alpha}_{0,t})^2\right]}{c} = O(n^{1-\delta}) = o(1),$$

proving the first claim of the Lemma. Now, turning to back to the terms in $\mathbb{P}(\max_{1 \leq t \leq T} (\hat{\beta}_t - \beta_t)^2 > c)$, observe for $R_{4,2}$ that

$$R_{4,2} := \sum_{t=1}^{T} \mathbb{P}\left(\left(n^{-1} \sum_{i=1}^{n} V_{it} \right)^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(\left(n^{-1} \sum_{i=1}^{n} (\hat{\Phi}_t(X_{it}) - \Phi(X_{it}))\langle X_{it}, \alpha_t \rangle \right)^2 > c \right).$$

As is shown in the Proof of Theorem 2, $\{V_{it} :\ 1 \leq i \leq n\}$ is a sequence of uncorrelated, zero mean random variables with finite and, due to stationarity time invariant upper bound for the second moment. Hence conclude that

$$
\begin{aligned}
R_{4,2} \leq & \sum_{t=1}^{T} \frac{E[V_{it}^2]n^{-1}}{c} + \sum_{t=1}^{T} \mathbb{P}\left(\left(\Delta^{(\Gamma)^2} ||\alpha_t||_2^2 + |||\Gamma|||^2 ||\alpha_t||_2^2 \right) ||\hat{\Phi}_t - \Phi||_{H'}^2 > c \right) \\
\leq & \sum_{t=1}^{T} \frac{E[V_{it}^2]n^{-1}}{c} + \sum_{t=1}^{T} \mathbb{P}\left(\Delta^{(\Gamma)^2} ||\alpha_t||_2^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi||_{H'}^2 > c \right) \\
& + \sum_{t=1}^{T} \mathbb{P}\left(|||\Gamma|||^2 ||\alpha_t||_2^2 ||\hat{\Phi}_t - \Phi||_{H'}^2 > c \right)
\end{aligned}
$$

Since $\sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi||_{H'}^2 > c \right) = o(1)$ for arbitrary c, the last two summands on the right hand side are $o(1)$. The first summand on the right hand side is $O(n^{1-\delta}) = o(1)$. For the remaining summand note that

$$
\sum_{t=1}^{T} \mathbb{P}\left(\Delta^{(\Gamma,t)^2} ||\alpha_t||_2^2 > c \right) \leq \sum_{t=1}^{T} \frac{E\left[\Delta^{(\Gamma,t)^2} \right]}{c} = O(n^{1-\delta}) = o(1)
$$

(see e.g. Hall and Horowitz (2007) and references therein) and thus $R_{4,2} = o(1)$.

Now, let $\hat{K}_{z\epsilon,t} := n^{-1} \sum_{i=1}^{n} z_{it} \epsilon_{it}$ and observe for $R_{4,3}$ that

$$
R_{4,3} := \sum_{t=1}^{T} \mathbb{P}\left(\hat{K}_{z\epsilon,t}^2 > c \right) \leq \sum_{t=1}^{T} \frac{E[\hat{K}_{z\epsilon,t}^2]}{c} = O(n^{\delta-1}) = o(1).
$$

For $R_{4,4}$ note

$$
\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t||_{H'}^2 ||\hat{K}_{X\epsilon,t}||_2^2 > c \right) \leq & \sum_{t=1}^{T} \mathbb{P}\left(||\Phi||_{H'}^2 ||\hat{K}_{X\epsilon,t}||_2^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi||_{H'}^2 ||\hat{K}_{X\epsilon,t}||_2^2 > c \right) \\
\leq & \sum_{t=1}^{T} \mathbb{P}\left(||\Phi||_{H'}^2 ||\hat{K}_{X\epsilon,t}||_2^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi||_{H'}^2 > c \right) \\
& + \sum_{t=1}^{T} \mathbb{P}\left(||\hat{K}_{X\epsilon,t}||_2^2 > c \right) \\
\leq & \sum_{t=1}^{T} \frac{||\Phi||_{H'}^2 E\left[||\hat{K}_{X\epsilon,t}||_2^2 \right]}{c} + \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi||_{H'}^2 > c \right) \\
& + \sum_{t=1}^{T} \frac{E\left[||\hat{K}_{X\epsilon,t}||_2^2 \right]}{c} \\
= & O(n^{\delta-1}) + o(1) + O(n^{\delta-1}) = o(1).
\end{aligned}
$$

Beyond that $R_{4,5}$ behaves according to

$$\begin{aligned}R_{4,5} &:= \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi||_{H'}^2 ||\bar{X}_t||_2^2 (\alpha_{0,t} - \hat{\alpha}_{0,t})^2 > c\right) \\ &\leq \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi||_{H'}^2 > c\right) + \sum_{t=1}^{T} \frac{E\left[||\bar{X}_t||_2^4\right]^{\frac{1}{2}} E\left[(\alpha_{0,t} - \hat{\alpha}_{0,t})^4\right]^{\frac{1}{2}}}{c} \\ &= o(1)\end{aligned}$$

combining the results from above. 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

$$\sum_{t=1}^{T} \mathbb{P}\left(||\hat{\alpha}_t - \alpha_t||_2^2 > c\right) \leq R_{5,1} + R_{5,2} + R_{5,3} + R_{5,4}.$$

This is because due to Assumption 2, $\sum_{j=m+1}^{\infty} a_j^{*2}$ is a null sequence and hence arbitrarily small for sufficiently large n. The terms $R_{5,1} - R_{5,4}$ are defined and treated as follows.

$$\begin{aligned}R_{5,1} &:= \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-2} \left(n^{-1} \sum_{i=1}^{n} \langle X_{it}, \hat{\phi}_{j,t} \rangle \epsilon_{it}\right)^2 > c\right) \\ &\leq \sum_{t=1}^{T} \frac{\sum_{j=1}^{m} \lambda_j^{-2} E\left[||\hat{K}_{X\epsilon,t}||_2^2\right]}{c} + \sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c\right) \\ &= O(n^{\frac{1+(1+\delta)\mu - 2(1-\delta)\nu}{\mu+2\nu}}) + o(1) = o(1)\end{aligned}$$

due to Assumption 4.

$$\begin{aligned}R_{5,2} :=& \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-2} \left(n^{-1} \sum_{i=1}^{n} \langle X_{it}, \hat{\phi}_{j,t} \rangle z_{it}\right)^2 (\hat{\beta}_t - \beta_t)^2 > c\right) \\ &\leq \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \lambda_j{}^{-2} \langle \hat{K}_{zX,t}, \hat{\phi}_{j,t} \rangle^2 (\hat{\beta}_t - \beta_t)^2 > c\right) + \sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c\right) \\ \leq& \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \lambda_j{}^{-2} \langle K_{zX}, \phi_j \rangle^2 (\hat{\beta}_t - \beta_t)^2 > c\right) + 2 \sum_{t=1}^{T} \mathbb{P}\left((\hat{\beta}_t - \beta_t)^2 > c\right) \\ &+ \sum_{t=1}^{T} \mathbb{P}\left(||K_{zX}||_2^2 \sum_{j=1}^{m} ||\phi_j - \hat{\phi}_{j,t}||_2^2 > c\right) + \sum_{t=1}^{T} \mathbb{P}\left(||K_{zX} - \hat{K}_{zX,t}||_2^2 > c\right) + \sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c\right).\end{aligned}$$

Because of the above results and the fact that $\sum_{j=1}^{m} \lambda_j{}^{-2} \langle K_{zX}, \phi_j \rangle^2 = O(1)$ by Assumption 2, the first two summands are $o(1)$. As a consequence of (47) the third summand is $o(1)$. Together with (42) and (40) this implies $R_{5,2} = o(1)$. Further define

$$
\begin{aligned}
R_{5,3} :=& \sum_{t=1}^{T} \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} (a_j^* - a_j)\hat{\phi}_{j,t} \right|\right|_2^2 > c \right) \\
\leq& \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} ||\alpha_t||_2^2 \, ||\phi_j - \hat{\phi}_{j,t}||_2^2 > c \right) = o(1)
\end{aligned}
$$

by (47) and

$$
\begin{aligned}
R_{5,4} :=& \sum_{t=1}^{T} \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} a_j^* (\hat{\phi}_{j,t} - \phi_j) \right|\right|_2^2 > c \right) \\
\leq& \sum_{t=1}^{T} \mathbb{P}\left(m \sum_{j=1}^{m} a_j^{*2} \, ||\hat{\phi}_{j,t} - \phi_j||_2^2 > c \right) \\
\leq& \sum_{t=1}^{T} \frac{Cm \sum_{j=1}^{m} a_j^{*2} E\left[||\phi_j - \hat{\phi}_{j,t}||_2^2 \right]}{c} \\
=& O\left(mn^{\delta-1} \right) = o(1).
\end{aligned}
$$

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 third 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

$$
\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}\left(\mathcal{S}_t^c\right) \leq& T^{-1} \sum_{t=1}^{T} \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left(\epsilon_{it}^2 - \sigma_\epsilon^2 + 2\epsilon_{it}\tilde{r}_{it} + \tilde{r}_{it}^2 \right) \right| > \frac{1}{2}\sigma_\epsilon^2 \right) \\
\leq& R_{6,1} + R_{6,2} + R_{6,3}
\end{aligned}
$$

where $\tilde{r}_{it} := (\alpha_{0,t} - \hat{\alpha}_{0,t}) + z_{it}(\beta_t - \hat{\beta}_t) + \langle X_{it}, \alpha_t - \hat{\alpha}_t \rangle$ and $R_{6,1} - R_{6,3}$ as follows.

Ad $R_{6,1}$:

$$
\begin{aligned}
R_{6,1} :=& \sum_{t=1}^{T} \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left(\epsilon_{it}^2 - \sigma_\epsilon^2 \right) \right| > c \right) \\
\leq& \sum_{t=1}^{T} \frac{n^{-1} E\left[\left(\epsilon_{it}^2 - \sigma_\epsilon^2 \right)^2 \right]}{c} = O(n^{\delta-1}) = o(1).
\end{aligned}
$$

Ad $R_{6,2}$:

$$R_{6,2} := \sum_{t=1}^{T} \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} (\epsilon_{it} \tilde{r}_{it}) \right| > c \right)$$
$$\leq R_{7,1} + R_{7,2} + R_{7,3}$$

with $R_{7,1} - R_{7,3}$ as follows.

$$R_{7,1} := \sum_{t=1}^{T} \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \epsilon_{it} (\alpha_{0,t} - \hat{\alpha}_{0,t}) \right| > c \right)$$
$$\leq \sum_{t=1}^{T} \frac{n^{-1} \sigma_\epsilon E\left[(\alpha_{0,t} - \hat{\alpha}_{0,t})^2 \right]^{\frac{1}{2}}}{c} = O(n^{\delta-2}) = o(1)$$

$$R_{7,2} := \sum_{t=1}^{T} \mathbb{P}\left(\left| (\beta_t - \hat{\beta}_t) n^{-1} \sum_{i=1}^{n} \epsilon_{it} z_{it} \right| > c \right)$$
$$\leq o(1) + \sum_{t=1}^{T} \mathbb{P}\left((\beta_t - \hat{\beta}_t)^2 > c \right) = o(1)$$

by (45) and the above results. Further

$$R_{7,3} := \sum_{t=1}^{T} \mathbb{P}\left(\left| \langle \hat{K}_{X\epsilon,t}, \alpha_t - \hat{\alpha}_t \rangle \right| > c \right)$$
$$\leq \sum_{t=1}^{T} \frac{n^{-1} \sigma_\epsilon^2 E\left[||X_{it}||_2^2 \right]}{c} + \sum_{t=1}^{T} \mathbb{P}\left(||\alpha_t - \hat{\alpha}_t||_2^2 > c \right) = o(1)$$

by the above results and again the iid sampling in the cross section dimension as well as independence of X_{it} and ϵ_{it}.

Ad $R_{6,3}$:

$$
\begin{aligned}
R_{6,3} :=& \sum_{t=1}^{T} \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^{n} \left(\tilde{r}_{it}^{2}\right) \right| > c \right) \\
\leq& \underbrace{\sum_{t=1}^{T} \mathbb{P}\left((\hat{\alpha}_{0,t} - \alpha_{0,t})^2 > c \right)}_{=:R_{8,1}} + \underbrace{\sum_{t=1}^{T} \mathbb{P}\left(\hat{K}_{z,t}(\hat{\beta}_t - \beta_t)^2 > c \right)}_{=:R_{8,2}} \\
&+ \underbrace{\sum_{t=1}^{T} \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} ||X_{it}||_2^2 ||\hat{\alpha}_t - \alpha_t||_2^2 > c \right)}_{=:R_{8,3}}
\end{aligned}
$$

with $R_{8,1} = o(1)$ as shown be fore and $R_{8,2} - R_{8,3}$ to be treated as follows.

$$
\begin{aligned}
R_{8,2} :=& \sum_{t=1}^{T} \mathbb{P}\left(\hat{K}_{z,t}(\hat{\beta}_t - \beta_t)^2 > c \right) \\
\leq& \sum_{t=1}^{T} \mathbb{P}\left(|\hat{K}_{z,t} - K_z|(\hat{\beta}_t - \beta_t)^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(K_z(\hat{\beta}_t - \beta_t)^2 > c \right) \\
\leq& \sum_{t=1}^{T} \mathbb{P}\left((\hat{K}_{z,t} - K_z) > c \right) + \sum_{t=1}^{T} \mathbb{P}\left((\hat{\beta}_t - \beta_t)^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(K_z(\hat{\beta}_t - \beta_t)^2 > c \right) = o(1)
\end{aligned}
$$

by (43) and the above results. Further it holds that

$$
\begin{aligned}
R_{8,3} :=& \sum_{t=1}^{T} \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} ||X_{it}||_2^2 ||\hat{\alpha}_t - \alpha_t||_2^2 > c \right) \\
\leq& \sum_{t=1}^{T} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} ||X_{it}||_2^2 - E\left[||X_{it}||_2^2 \right] \right| ||\hat{\alpha}_t - \alpha_t||_2^2 > c \right) \\
&+ \sum_{t=1}^{T} \mathbb{P}\left(E\left[||X_{it}||_2^2 \right] ||\hat{\alpha}_t - \alpha_t||_2^2 > c \right) \\
\leq& \sum_{t=1}^{T} \frac{n^{-1} E\left[\left| ||X_{it}||_2^2 - E\left[||X_{it}||_2^2 \right] \right|^2 \right]}{c} + \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\alpha}_t - \alpha_t||_2^2 > c \right) \\
&+ \sum_{t=1}^{T} \mathbb{P}\left(E\left[||X_{it}||_2^2 \right] ||\hat{\alpha}_t - \alpha_t||_2^2 > c \right) \\
=& O(n^{\delta - 1}) + o(1) + o(1) = o(1)
\end{aligned}
$$

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

$$\sum_{t=1}^{T} \mathbb{P}\left(||\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}||_2^2 > c \right) \leq R_{9,1} + R_{9,2}$$

with

$$R_{9,1} := \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} (\hat{a}_{j,t} - a_j)^2 \frac{\hat{\lambda}_{j,t}}{\hat{\sigma}_{\epsilon,t}^2} > c \right)$$

$$\text{and} \quad R_{9,2} := \sum_{t=1}^{T} \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} \left(\frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{\epsilon,t}} \hat{\phi}_{j,t} a_j - \frac{\lambda_j^{1/2}}{\sigma_\epsilon} \phi_j a_j^* \right) \right|\right|_2^2 > c \right).$$

$R_{9,1}$ can be decomposed according to

$$R_{9,1} \leq R_{10,1} + R_{10,2} + R_{10,3}$$

where, for some suitable $c > 0$,

$$R_{10,1} := \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-1} \langle \hat{K}_{zX,t}, \hat{\phi}_{j,t} \rangle^2 (\beta_t - \hat{\beta}_t)^2 > c \right)$$

$$R_{10,2} := \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-1} \langle \bar{X}_t, \hat{\phi}_{j,t} \rangle^2 (\alpha_{0,t} - \hat{\alpha}_{0,t}) > c \right)$$

$$R_{10,3} := \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \hat{\lambda}_{j,t}^{-1} \langle \hat{K}_{X\epsilon,t}, \hat{\phi}_{j,t} \rangle^2 > c \right)$$

These terms in turn behave as follows.

$$\begin{aligned} R_{10,1} \leq & \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \lambda_j^{-1} \langle K_{zX}, \phi_j \rangle^2 (\beta_t - \hat{\beta}_t)^2 > c \right) + 2 \sum_{t=1}^{T} \mathbb{P}\left((\beta_t - \hat{\beta}_t)^2 > c \right) \\ & + \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \lambda_j^{-1} ||\hat{K}_{zX,t} - K_{zX}||_2^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} \lambda_j^{-1} ||\hat{\phi}_{j,t} - \phi_j||_2^2 > c \right) + \sum_{t=1}^{T} \mathbb{P}\left(\mathcal{F}_{m,t}^c \right) \\ & = o(1). \end{aligned}$$

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

$$R_{10,2} \leq \sum_{t=1}^{T} \frac{\sum_{j=1}^{m} \lambda_j^{-1} E\left[||\bar{X}_t||_2^4\right]^{1/2} E\left[|\hat{\alpha}_{0,t} - \alpha_{0,t}|^4\right]^{1/2}}{c} = o(1)$$

and

$$R_{10,3} \leq \sum_{t=1}^{T} \frac{\sum_{j=1}^{m} \lambda_j^{-1} E\left[||\hat{K}_{X\epsilon,t}||_2^2\right]}{c} = o(1)$$

Now turning to $R_{9,2}$ note that

$$R_{9,2} \leq R_{11,1} + R_{11,2}$$

where

$$R_{11,1} := \sum_{t=1}^{T} \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} \left(\frac{\hat{\lambda}_{j,t}^{1/2} \sigma_\epsilon}{\sigma_\epsilon \hat{\sigma}_{\epsilon,t}} \hat{\phi}_{j,t} - \frac{\lambda_j^{1/2} \hat{\sigma}_{\epsilon,t}}{\sigma_\epsilon \hat{\sigma}_{\epsilon,t}} \phi_j \right) a_j^* \right|\right|_2^2 > c \right)$$

and

$$R_{11,2} = \sum_{t=1}^{T} \mathbb{P}\left(\left|\left| \sum_{j=1}^{m} (a_j^* - a_j) \frac{\hat{\lambda}_{j,t}^{1/2}}{\hat{\sigma}_{\epsilon,t}} \hat{\phi}_j \right|\right|_2^2 > c \right).$$

Note for $R_{11,1}$:

$$R_{11,1} \leq R_{12,1} + R_{12,2} + R_{12,3} + \mathbb{P}(\mathcal{S}_t^c)$$

with

$$R_{12,1} := \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} |a_j^*| \hat{\lambda}_{j,t}^{1/2} |\sigma_\epsilon - \hat{\sigma}_{\epsilon,t}| > c \right)$$
$$R_{12,2} := \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} |a_j^*| \hat{\lambda}_{j,t}^{1/2} \hat{\sigma}_{\epsilon,t} ||\phi_j - \hat{\phi}_{j,t}||_2 > c \right)$$
$$R_{12,3} := \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} |a_j^*| |\hat{\lambda}_{j,t}^{1/2} - \lambda_j^{1/2}| \hat{\sigma}_{\epsilon,t} > c \right)$$

By the mean value theorem for any $1 \leq t \leq T$ there exists a number $g_t^{(\sigma)} \in (0,1)$ such that $|\sigma_\epsilon - \hat{\sigma}_{\epsilon,t}| = \frac{1}{2}\left[g_t^{(\sigma)}\hat{\sigma}^2_{\epsilon,t} + (1-g_t^{(\sigma)})\sigma^2_\epsilon\right]^{-\frac{1}{2}} |\hat{\sigma}^2_{\epsilon,t} - \sigma^2_\epsilon|$. On $\mathcal{S}_t$ it thus holds that $|\sigma_\epsilon - \hat{\sigma}_{\epsilon,t}| \leq \frac{1}{2}\sigma_\epsilon^{-1}|\hat{\sigma}^2_{\epsilon,t} - \sigma^2_\epsilon|$ and we can conclude

$$\begin{aligned}
R_{12,1} &:= \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} |a_j^*|\hat{\lambda}_{j,t}^{1/2}|\sigma_\epsilon - \hat{\sigma}_{\epsilon,t}| > c\right) \\
&\leq \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} |a_j^*|\lambda_j^{1/2}|\hat{\sigma}^2_{\epsilon,t} - \sigma^2_\epsilon| > c\right) + \sum_{t=1}^{T}\mathbb{P}\left(\mathcal{F}^c_{m,t}\right) + \sum_{t=1}^{T}\mathbb{P}\left(\mathcal{S}^c_t\right) \\
&= o(1)
\end{aligned}$$

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

$$\begin{aligned}
R_{12,2} &\leq \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} |a_j^*|\lambda_j^{1/2}||\phi_j - \hat{\phi}_{j,t}||_2 > c\right) \\
&\leq \sum_{t=1}^{T} \frac{\sum_{j=1}^m |a_j^*|\lambda_j^{1/2}E[||\phi_j - \hat{\phi}_{j,t}||_2]}{c} \\
&= O(n^{\delta-1/2}) = o(1)
\end{aligned}$$

given δ as in Assumption 3.

Again employing the mean value theorem, for any $1 \leq t \leq T$ and $1 \leq j \leq m$ there exists some number $g_{j,t}^{(\lambda)} \in (0,1)$ such that $|\sigma_\epsilon - \hat{\sigma}_{\epsilon,t}| = \frac{1}{2}\left[g_{j,t}^{(\lambda)}\hat{\lambda}_{j,t} + (1-g_{j,t}^{(\lambda)})\lambda_j\right]^{-\frac{1}{2}} |\hat{\lambda}_{j,t} - \lambda_j|$. On $\mathcal{F}_{m,t}$ it thus holds that $|\lambda_j^{1/2} - \hat{\lambda}_{j,t}^{1/2}| \leq \frac{1}{2}\lambda_j^{-1/2}|\hat{\lambda}_{j,t} - \lambda_j|$. It follows that

$$\begin{aligned}
R_{12,3} &\leq \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} |a_j^*|\lambda_j^{-1/2}|\hat{\lambda}_{j,t} - \lambda_j| > c\right) \\
&\leq \sum_{t=1}^{T} \frac{\sum_{j=1}^m |a_j^*|\lambda_j^{-1/2}E[|\hat{\lambda}_{j,t} - \lambda_j|]}{c} + \sum_{t=1}^{T}\mathbb{P}\left(\mathcal{F}^c_{m,t}\right) + \sum_{t=1}^{T}\mathbb{P}\left(\mathcal{S}^c_t\right) = o(1)
\end{aligned}$$

as was argued before.

It remains to show that $R_{11,2} = o(1)$. For this purpose note

$$\begin{aligned}
R_{11,2} &= \sum_{t=1}^{T} \mathbb{P}\left(\sum_{j=1}^{m} (a_j^* - a_j)^2 \frac{\hat{\lambda}_{j,t}}{\hat{\sigma}^2_{\epsilon,t}} > c\right) \\
&\leq \sum_{t=1}^{T} \frac{\sum_{j=1}^m ||\alpha_t||_2^2 E[||\hat{\phi}_{j,t} - \phi_j||_2^2]\frac{\lambda_j^2}{\sigma_\epsilon^2}}{c} + \sum_{t=1}^{T}\mathbb{P}\left(\mathcal{F}^c_{m,t}\right) + \sum_{t=1}^{T}\mathbb{P}\left(\mathcal{S}^c_t\right) = o(1).
\end{aligned}$$

Combining arguments implies the last part of the Lemma. ■

Proof of Theorem 1.

Consider the set $S^{(j)}$ at an iteration j as described in step 2 of the estimation procedure. For a $t \in S$ denote the set of indexes corresponding to the ordered distances $\hat{\Delta}_{t(1)} \leq \cdots \leq \hat{\Delta}_{t(|S|)}$ as $\{(1), \ldots, (|S|)\}$. In analogy, the index set corresponding ordered population distances $\Delta_{t[1]} \leq \cdots \leq \Delta_{t[|S|]}$ is denoted as $\{[1], \ldots, [|S|]\}$, where Δ_{ts} is as in Assumption 7. With $\hat{\kappa}$ defined in the estimation procedure it holds that

$$\mathbb{P}\left(\{(1), \ldots, (\hat{\kappa})\} \neq \{([1], \ldots, [\kappa]\}\right) \leq \mathbb{P}\left(\{(1), \ldots, (\kappa)\} \neq \{[1], \ldots, [\kappa]\}\right) + \mathbb{P}(\hat{\kappa} \neq \kappa) \tag{49}$$

$$= o(1) + o(1) = o(1). \tag{50}$$

In order to prove that the first probability on the right hand side of (49) is a null sequence, suppose that $t \in G$, with G being some regime. Let there be, beyond t, κ indexes in S being elements of G as well. Then $\Delta_{t[1]} = \cdots = \Delta_{t[\kappa]} = 0$ and $0 < C_\Delta \leq \Delta_{t[\kappa+1]} \leq \cdots \leq \Delta_{t[|S|]}$ due to Assumption 7.

As shown before, $||\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}||_2^2 = o_p(1)$ and thus the distances converge as well, i.e. $\hat{\Delta}_{ts} \xrightarrow{\mathbb{P}} \Delta_{ts}$ for any $t, s \in S$. As this convergence is uniform over t it also holds that $\max_{1 \leq s \leq \kappa} \hat{\Delta}_{t(s)} = o_p(1)$ and $\min_{\kappa < s \leq |S|} \hat{\Delta}_{t(s)} \geq C_\Delta + o_p(1)$ as well as $\max_{1 \leq s \leq \kappa} \hat{\Delta}_{t[s]} = o_p(1)$ and $\min_{\kappa < s \leq |S|} \hat{\Delta}_{t[s]} \geq C_\Delta + o_p(1)$.

This immediately implies that the first probability on the right hand side of (49) 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 $n \to \infty$, from which the claim in the Theorem follows. ■

Remark

In light of Theorem 1, the classification error is, in what follows, asymptotically negligible. To see that 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 $M_1, M_2 > 0$

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

and

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

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

In light of this remark, the proof of Theorem 2 starts from the quantities $\tilde{\beta}^*_{t,k}$ and $\tilde{A}^*_k$ rather than their contaminated counterparts.

Remark

Denote as $\tilde{\phi}^*_{j,k}, \tilde{\lambda}^*_{j,k}, \tilde{K}^*_{X,k}$ the estimators $\tilde{\phi}_j, \tilde{\lambda}_j, \tilde{K}_X$ from the observations $\{(z_{it}, X_{it}) : 1 \le i \le n,\ t \in G_k\}$). 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 $\{X_{it} : 1 \le i \le n,\ t \in G_k\}$ is L^4_m-approximable. Thus the following three inequalities from Hörmann and Kokoszka (2010) hold (where for the third inequality we used our Assumption 2 already).

$$E\left[||\tilde{K}^*_{X,k} - K_X||_2^2\right] \le C(n|G_k|)^{-1} \tag{51}$$

$$E\left[\left|\tilde{\lambda}^*_{j,k} - \lambda_j\right|^q\right] \le C(n|G_k|)^{-q/2} \tag{52}$$

$$E\left[\left|\left|\tilde{\phi}^*_{j,k} - \phi_j\right|\right|_2^q\right] \le C(n|G_k|)^{-q/2} j^{q(1+\mu)} \tag{53}$$

for $1 \le j \le \tilde{m}$ and $q = 1, 2, \ldots$.

Further note that the sequence $\{(z_{it}, X_{it}) : 1 \le i \le n,\ t \in G_k\}$ is m-dependent (after suitable relabeling). As a consequence of this,

$$E\left[||\tilde{K}^*_{zX,k} - K_{zX}||_2^2\right] = O((n|G_k|)^{-1}) \tag{54}$$

$$E\left[|\tilde{K}^*_{z,k} - K_z|^2\right] = O((n|G_k|)^{-1}), \tag{55}$$

which can be shown by simple moment calculations. Given Assumption 1, (51)- (53) are a consequence of Theorem 3.2 in Hörmann and Kokoszka (2010) and Assumption 2. Beyond that it also holds given Assumptions 1 and 2 that

$$||\tilde{\Phi}^*_k - \Phi||^2_{H'} = O_p\left((n|G_k|)^{\frac{1-2\nu}{\mu+2\nu}}\right). \tag{56}$$

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

$$z_{it} = \langle \zeta, X_{it} \rangle + \eta_{it} \tag{57}$$

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

$$z_{j(i,t)} = \langle \zeta, X_{j(i,t)} \rangle + \eta_{j(i,t)} \tag{58}$$

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 (58). Their arguments (see the proof of their Theorem 1) transfer mutatis mutandis immediately to a setup with weak dependence in the sense of L_m^4 dependent regressor (which is in our setup a consequence of Assumption 1) and m-dependent errors possessing fourth 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 as claimed in (56), i.e.

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

Proof of Theorem 2

Building on Lemma 1 and the above remarks, we will begin by showing

$$(|G_k|)^{-1} \sum_{t \in G_k} (\tilde{\beta}_{t,k}^* - \beta_t)^2 = O_p\left(n^{-1}\right) \tag{59}$$

as $n, T \to \infty$ in what follows. To do so, define

$$\tilde{B}_k^* := \tilde{K}_{z,k}^* - \tilde{\Phi}_k^*\left(\tilde{K}_{zX,k}^*\right) \qquad \text{and} \qquad \breve{B}_{t,k}^* := \hat{K}_{z,t} - \tilde{\Phi}_k^*\left(\hat{K}_{zX,t}\right)$$

as well as

$$\hat{K}_{zy^{(c)},t} := n^{-1}\sum_{i=1}^{n} z_{it}(y_{it} - \alpha_{0,t}) \qquad \text{and} \qquad \hat{K}_{Xy^{(c)},t}(s) := n^{-1}\sum_{i=1}^{n} X_{it}(s)(y_{it} - \alpha_{0,t}).$$

Now, note that $\hat{K}_{z\breve{y},t} := \hat{K}_{zy^{(c)},t} + \bar{z}_t(\alpha_{0,t} - \hat{\alpha}_{0,t})$ and $\hat{K}_{X\breve{y},t}(s) := \hat{K}_{Xy^{(c)},t}(s) + \bar{X}_t(s)(\alpha_{0,t} - \hat{\alpha}_{0,t})$. Given this, it is possible to restate the $t-th$ summand in (59) according to

$$\begin{aligned}
(\tilde{\beta}_{t,k}^* - \beta_t) &= (\tilde{B}_k^*)^{-1}(\hat{K}_{zy^{(c)},t} - \tilde{\Phi}_k^*(\hat{K}_{Xy^{(c)},t})) + (\tilde{B}_k^*)^{-1} r_{\beta,t}^* \\
&= (\tilde{B}_k^*)^{-1}\left((\breve{B}_{t,k}^* - \tilde{B}_k^*)\beta_t + n^{-1}\sum_{i=1}^{n}(z_{it} - \tilde{\Phi}_k^*(X_{it}))(\langle X_{it}, A_k\rangle) + \epsilon_{it})\right) + (\tilde{B}_k^*)^{-1} r_{\beta,t}^*
\end{aligned}$$

where $r^*_{\beta,t} := (\bar{z}_t - \tilde{\Phi}^*_k(\bar{X}_t))(\alpha_{0,t} - \hat{\alpha}_{0,t})$. It follows from the Cauchy-Schwarz inequality

$$\begin{aligned}\frac{1}{|G_k|}\sum_{t\in G_k}\left(\tilde{\beta}^*_{t,k} - \beta_t\right)^2 \le &(\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}(\breve{B}^*_{t,k} - \tilde{B}^*_k)^2\beta_t^2 \\ &+ (\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^n (z_{it} - \tilde{\Phi}^*_k(X_{it}))(\langle X_{it}, A_k\rangle + \epsilon_{it})\right)^2 + (\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}(r^*_{\beta,t})^2\end{aligned}$$

For the first to summands it holds due to Assumption 2

$$\begin{aligned}&(\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}\left((\breve{B}^*_{t,k} - \tilde{B}^*_k)^2\beta_t^2\right) + (\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^n (z_{it} - \tilde{\Phi}^*_k(X_{it}))(\langle X_{it}, A_k\rangle + \epsilon_{it})\right)^2 \\ &\le C(\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}(\breve{B}^*_{t,k} - \tilde{B}^*_k)^2 + (\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^n (z_{it} - \tilde{\Phi}^*_k(X_{it}))(\langle X_{it}, A_k\rangle + \epsilon_{it})\right)^2 \\ &\le C(\tilde{B}^*_k)^{-2}\frac{3}{|G_k|}\sum_{t\in G_k}(\breve{B}^*_{t,k} - \tilde{B}^*_k)^2 + (\tilde{B}^*_k)^{-2}\frac{9}{|G_k|}\sum_{t\in G_k}\sum_{l=1}^3 A^2_{l,nT}\end{aligned}$$

where

$$A_{1,nt} := n^{-1}\sum_{i=1}^n (z_{it} - \tilde{\Phi}^*_k(X_{it}))\langle X_{it}, A_k\rangle \tag{60}$$

$$A_{2,nt} := n^{-1}\sum_{i=1}^n (\Phi(X_{it}) - \tilde{\Phi}^*_k(X_{it}))\epsilon_{it} \tag{61}$$

$$A_{3,nt} := n^{-1}\sum_{i=1}^n (z_{it} - \Phi(X_{it}))\epsilon_{it}. \tag{62}$$

Due to Assumptions 3 and 4, $(n|G_k|)^{\frac{1-2\nu}{\mu+2\nu}} = o(n^{-1})$ and hence it follows from (54)-(56) that

$$\begin{aligned}(\tilde{B}^*_k - B)^2 &\le 2|\tilde{K}^*_{z,k} - K_z|^2 + 2|\tilde{\Phi}^*_k(\tilde{K}^*_{zX,k}) - \Phi(K_{zX})|^2 \\ &\le 2|\tilde{K}^*_{z,k} - K_z|^2 + 4||\tilde{\Phi}^*_k - \Phi||^2_{H'}||K_{zX}||^2 + 8(||\Phi||^2_{H'} + ||\tilde{\Phi}^*_k - \Phi||^2_{H'})|\tilde{K}^*_{zX,k} - K_{zX}|^2 = O_p(n^{-1}).\end{aligned}$$

Assumption 1 and 4 together with (56) imply

$$\begin{aligned}\frac{1}{|G_k|}\sum_{t\in G_k}(\breve{B}^*_{t,k} - B)^2 \le &2\frac{1}{|G_k|}\sum_{t\in G_k}|\hat{K}_{z,t} - K_z|^2 + 4||\tilde{\Phi}^*_k - \Phi||^2_{H'}|K_z|^2 \\ &+ 8(||\Phi||^2_{H'} + ||\tilde{\Phi}^*_k - \Phi||^2_{H'})\frac{1}{|G_k|}\sum_{t\in G_k}|\hat{K}_{z,t} - K_z|^2 = O_p(n^{-1})\end{aligned}$$

as $(n, T) \to \infty$. Combining the above observations yields

$$\frac{1}{|G_k|}\sum_{t\in G_k}(\breve{B}^*_{t,k}-\tilde{B}^*_k)^2 \leq \frac{2}{|G_k|}\sum_{t\in G_k}(\breve{B}^*_{t,k}-B)^2+2(B-\tilde{B}^*_k)^2$$
$$= O_p(n^{-1})$$

as $(n,T)\to\infty$.

In the following we will show $\frac{1}{|G_k|}\sum_{t\in G_k}A^2_{l,nT}=O_p(n^{-1})$ as $(n,T)\to\infty$, $l=1,2,3$.

$$\frac{1}{|G_k|}\sum_{t\in G_k}A^2_{1,nt} := \frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}(z_{it}-\tilde{\Phi}^*_k(X_{it}))\langle X_{it},A_k\rangle\right)^2$$
$$\leq \frac{2}{|G_k|}\sum_{t\in G_k}\left(\left(n^{-1}\sum_{i=1}^{n}(z_{it}-\Phi(X_{it}))\langle X_{it},A_k\rangle\right)^2+\left(n^{-1}\sum_{i=1}^{n}(\Phi(X_{it})-\tilde{\Phi}^*_k(X_{it}))\langle X_{it},A_k\rangle\right)^2\right)$$

As was alluded to in the proof of Lemma 1, consider the random variable $V_{it}:=(z_{it}-\Phi(X_{it}))\langle X_{it},A_k\rangle$, where we suppress the dependence on k for a moment for readability. Note for its first moment that

$$\begin{aligned}
E\left[V_{it}\right] &= E\left[(z_{it}-\Phi(X_{it}))\langle X_{it},A_k\rangle\right]\\
&= \langle K_{zX},A_k\rangle-\sum_{j=1}^{\infty}\frac{\langle K_{zX},\phi_j\rangle}{\lambda_j}E\left[\langle X_{it},\phi_j\rangle\left(\sum_{l=1}^{\infty}\langle X_{it},\phi_l\rangle\langle\phi_l,A_k\rangle\right)\right]\\
&= \sum_{j=1}^{\infty}\langle K_{zX},\phi_j\rangle\langle\phi_j,A_k\rangle-\sum_{j=1}^{\infty}\frac{\langle K_{zX},\phi_j\rangle}{\lambda_j}\lambda_j\langle\phi_j,A_k\rangle\\
&= 0
\end{aligned}$$

because

$$E\left[\langle X_{it},\phi_j\rangle\langle X_{it},\phi_l\rangle\right]=\begin{cases}\lambda_j & \text{if } l=j\\ 0 & \text{if } l\neq j.\end{cases}$$

It follows from the independence of X_{it} and z_{it} over index i that

$$E\left[V_{it}V_{jt}\right]=\begin{cases}E\left[V_{it}^2\right]<\infty & \text{if } i=j\\ 0 & \text{if } i\neq j.\end{cases}$$

Note for the second moment of V_{it} that application of the Cauchy Schwarz inequality implies

$$E\left[V_{it}^2\right] \leq 2\left(E[z_{it}^2]E[||X_{it}||_2^2]||A_k||_2^2 + ||\Phi||_{H'}^2 E[||X_{it}||_2^4]||A_k||_2^2\right) < \infty$$

due to Assumption 1. Combining these arguments leads to

$$E\left[\frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}(z_{it}-\Phi(X_{it}))\langle X_{it}, A_k\rangle\right)^2\right] = n^{-1}E\left[V_{it}^2\right] = O(n^{-1}).$$

For the remaining term in $\frac{1}{|G_k|}\sum_{t\in G_k} A_{1,nt}^2$ note

$$\frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}(\Phi(X_{it})-\tilde{\Phi}_k^*(X_{it}))\langle X_{it}, A_k\rangle\right)^2 \leq ||\tilde{\Phi}_k^*-\Phi||_{H'}^2\frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}||X_{it}||_2\langle X_{it}, A_k\rangle\right)^2.$$

It follows from Assumption 1 that

$$E\left[\frac{2}{|G_k|}\sum_{t\in G_k}(n^{-1}\sum_{i=1}^{n}||X_{it}||_2\langle X_{it}, A_k\rangle)^2\right] = O(1)$$

which leads in combination with (56) to the observation that

$$\frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}(\Phi(X_{it})-\tilde{\Phi}_k^*(X_{it}))\langle X_{it}, A_k\rangle\right)^2 = O_p(n^{-1}).$$

Thus $\frac{1}{|G_k|}\sum_{t\in G_k} A_{1,nt}^2 = O_p(n^{-1})$.

The term (61) can be bounded by

$$\begin{aligned}\frac{1}{|G_k|}\sum_{t\in G_k} A_{2,nt}^2 &= \frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}(\tilde{\Phi}_k^*(X_{it})-\Phi(X_{it}))\epsilon_{it}\right)^2\\ &\leq ||\tilde{\Phi}_k^*-\Phi||_{H'}\frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}||X_{it}||_2\epsilon_{it}\right)^2 = O_p(n^{-2})\end{aligned}$$

by exogeneity of X_{it} and the sampling scheme in Assumption 1. The expression in (62) can be treated according to

$$
\begin{aligned}
\frac{1}{|G_k|}\sum_{t\in G_k} A_{3,nt}^2 &= \frac{1}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}(z_{it}-\Phi(X_{it}))\epsilon_{it}\right)^2 \\
&\leq \frac{2}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n} z_{it}\epsilon_{it}\right)^2 + \frac{2}{|G_k|}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}||\Phi||_{H'}||X_{it}||_2\epsilon_{it}\right)^2 = O_p(n^{-1})
\end{aligned}
$$

which is a consequence of simple moment calculations.

Since $E\left[(\hat{\alpha}_{0,t}-\alpha_{0,t})^4\right] = O(n^{-2})$ as shown above, the stationarity of X_{it} and z_{it},

$$
\frac{1}{|G_k|}\sum_{t\in G_k} E\left[(\hat{\alpha}_{0,t}-\alpha_{0,t})^2\bar{z}_t^2\right] \leq E\left[(\hat{\alpha}_{0,t}-\alpha_{0,t})^4\right]^{\frac{1}{2}} E\left[\bar{z}_t^4\right]^{\frac{1}{2}} = O(n^{-2})
$$

as well as

$$
\frac{1}{|G_k|}\sum_{t\in G_k} E\left[(\hat{\alpha}_{0,t}-\alpha_{0,t})^2||\bar{X}_t||_2^2\right] \leq E\left[(\hat{\alpha}_{0,t}-\alpha_{0,t})^4\right]^{\frac{1}{2}} E\left[||\bar{X}_t||_2^4\right]^{\frac{1}{2}} = O(n^{-1})
$$

it follows from the Cauchy-Schwarz inequality and the arguments employed before that

$$
\begin{aligned}
\frac{1}{|G_k|}\sum_{t\in G_k}(r_{\beta,t}^*)^2 &\leq \frac{2}{|G_k|}\sum_{t\in G_k}(\hat{\alpha}_{0,t}-\alpha_{0,t})^2\bar{z}_t^2 + ||\tilde{\Phi}_k^*||_{H'}^2\frac{2}{|G_k|}\sum_{t\in G_k}(\hat{\alpha}_{0,t}-\alpha_{0,t})^2||\bar{X}_t^2||_2^2 \\
&= O_p(n^{-1}).
\end{aligned}
$$

Together with our Remark on the classification error, part 2. of the Theorem follows. Further note, that since, $E\left[(\hat{\alpha}_{0,t}-\alpha_{0,t})^4\right]$ does not depend on t due to Assumption 1, part 1., of the Theorem follows along with the above arguments.

Regarding the third part of the Theorem, denote the basis coefficients of $\tilde{A}_k^*$, $\tilde{a}_{j,k}^* = \tilde{a}_{j,k}^{(1)} + \tilde{a}_{j,k}^{(2)}$, where

$$
\begin{aligned}
\tilde{a}_j^{(1)} &:= (\tilde{\lambda}_{j,k}^*)^{-1}\frac{1}{n|G_k|}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\tilde{\phi}_{j,k}^*\rangle\left(\epsilon_{it}+\langle A_k,\phi_j\rangle\right) \\
\tilde{a}_j^{(2)} &:= (\tilde{\lambda}_{j,k}^*)^{-1}\frac{1}{n|G_k|}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\tilde{\phi}_{j,k}^*\rangle\left(z_{it}(\tilde{\beta}_{t,k}^*-\beta_t)+(\hat{\alpha}_{0,t}-\alpha_{0,t})\right).
\end{aligned}
$$

The upper bound

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

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 Remark and Assumptions 1-5, it holds that $\left|\left| \sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}^{(1)} \tilde{\phi}_{j,k}^* - A_k \right|\right|_2^2 = O_p\left(n^{\frac{(1+\delta)(1-\mu-2\nu)}{\mu+2\nu}} \right)$. The remaining sum can be treated according to

$$
\begin{aligned}
\sum_{j=1}^{\tilde{m}} (\tilde{a}_{j,k}^{(2)})^2 \leq & 2 \left(|G_k|^{-1} \sum_{t \in G_k} (\tilde{\beta}_{t,k}^* - \beta_t)^2 \right) \left(|G_k|^{-1} \sum_{t \in G_k} \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(n^{-1} \sum_{i=1}^{n} z_{it} \langle X_{it}, \tilde{\phi}_{j,k}^* \rangle \right)^2 \right) \\
& + 2 \left(|G_k|^{-1} \sum_{t \in G_k} (\hat{\alpha}_{0,t} - \alpha_{0,t})^2 \right) \left(|G_k|^{-1} \sum_{t \in G_k} \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}_{j,k}^*)^{-2} \left(n^{-1} \sum_{i=1}^{n} \langle X_{it}, \tilde{\phi}_{j,k}^* \rangle \right)^2 \right).
\end{aligned}
$$

We begin examining the first summand on the right hand side. Note that $n^{-1} \sum_{i=1}^{n} z_{it} \langle X_{it}, \tilde{\phi}_{j,k}^* \rangle = \langle \hat{K}_{zX,t}, \tilde{\phi}_{j,k}^* \rangle$ and thus

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

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

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

Clearly $\mathbb{1}(\mathcal{F}_{\tilde{m}})\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle K_{Xz},\phi_j\rangle^2 = O_p(1)$ is thus sufficient for $\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle K_{Xz},\phi_j\rangle^2 = O_p(1)$.[13] Further it follows from the Cauchy Schwarz inequality that

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

because of the Assumptions 3-5. From this we conclude using an analogous argument as before $|G_k|^{-1}\sum_{t\in G_k}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle \hat{K}_{Xz,t}-K_{Xz},\tilde{\phi}^*_{j,k}\rangle^2 = o_p(1)$.

Further, due to (53) and Assumption 4, it holds that

$$\mathbb{1}(\mathcal{F}_{\tilde{m}})\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle K_{Xz},\tilde{\phi}^*_{j,k}-\phi_j\rangle^2 \leq ||K_{zX}||_2^2\sum_{j=1}^{\tilde{m}}||\tilde{\phi}^*_{j,k}-\phi_j||_2^2\lambda_j^{-2} = o_p(1)$$

implying $\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle K_{Xz},\tilde{\phi}^*_{j,k}-\phi_j\rangle^2 = O_p(1)$.

For the remaining term note that $|G_k|^{-1}\sum_{t\in G_k}(\hat{\alpha}_{0,t}-\alpha_{0,t})^2 = O_p(n^{-1})$ as shown before. From the fact that $\{\langle X_{it},\tilde{\phi}^*_{j,k}\rangle:\quad 1\leq i\leq n,\ t\in G_k\}$ are empirically uncorrelated it follows

$$\begin{aligned}
\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}|G_k|^{-1}\sum_{t\in G_k}\left(n^{-1}\sum_{i=1}^{n}\langle X_{it},\tilde{\phi}^*_{j,k}\rangle\right)^2 =& n^{-1}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}|nG_k|^{-1}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\tilde{\phi}^*_{j,k}\rangle^2\\
\leq& \frac{2}{n}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}|nG_k|^{-1}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\phi_j\rangle^2\\
&+\frac{2}{n}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}|nG_k|^{-1}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\tilde{\phi}^*_{j,k}-\phi_j\rangle^2.
\end{aligned}$$

For the first summand on the right hand side note that

[13]In analogy to the arguments used in the Proof of Lemma 2, it holds for some constant $c>0$

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle K_{Xz},\phi_j\rangle^2 > c\right) &\leq \mathbb{P}\left(\left\{\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}^*_{j,k})^{-2}\langle K_{Xz},\phi_j\rangle^2 > c\right\}\cap\{\mathcal{F}_{\tilde{m}}\}\right)+\mathbb{P}(\mathcal{F}^c_{\tilde{m}})\\
&\leq \mathbb{P}\left(\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}\langle K_{Xz},\phi_j\rangle^2 > c/4\right)+\mathbb{P}(\mathcal{F}^c_{\tilde{m}})\\
&= \mathbb{P}\left(\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}\langle K_{Xz},\phi_j\rangle^2 > c/4\right)+o(1).
\end{aligned}$$

We use this type of argument again without further reference.

$$
\begin{aligned}
E\left[\mathbb{1}(\mathcal{F}_{\tilde{m}})\frac{2}{n}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^{*})^{-2}|nG_k|^{-1}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\phi_j\rangle^2\right] &\leq \frac{8}{n}\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}E\left[\langle X_{it},\phi_j\rangle^2\right] \\
&= \frac{8}{n}\sum_{j=1}^{\tilde{m}}\lambda_j^{-1} \\
&= O\left(n^{\frac{(1+\delta)(1+\mu)-\mu-2\nu}{\mu+2\nu}}\right) \\
&= o(1)
\end{aligned}
$$

by Assumption 4. Hence using the same arguments as before it follows

$$
\frac{2}{n}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^{*})^{-2}|nG_k|^{-1}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\phi_j\rangle^2 = o_p(1).
$$

For the second summand observe that

$$
\begin{aligned}
E\left[\mathbb{1}(\mathcal{F}_{\tilde{m}})\frac{2}{n}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^{*})^{-2}|nG_k|^{-1}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\tilde{\phi}_{j,k}^{*}-\phi_j\rangle^2\right] &\leq \frac{8}{n}\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}E\left[\langle X_{it},\tilde{\phi}_{j,k}^{*}-\phi_j\rangle^2\right] \\
&\leq \frac{8}{n}\sum_{j=1}^{\tilde{m}}\lambda_j^{-2}E\left[||X_{it}||_2^4\right]^{\frac{1}{2}}E\left[||\tilde{\phi}_{j,k}^{*}-\phi_j||_2^4\right]^{\frac{1}{2}} \\
&= O\left(n^{\frac{(1+\delta)(3+4\mu)-(1+\delta)(\mu+2\nu)}{\mu+2\nu}}\right) \\
&= o(1)
\end{aligned}
$$

by Assumption 4. Using once more the arguments from before it can be concluded

$$
\frac{2}{n}\sum_{j=1}^{\tilde{m}}(\tilde{\lambda}_{j,k}^{*})^{-2}|nG_k|^{-1}\sum_{t\in G_k}\sum_{i=1}^{n}\langle X_{it},\tilde{\phi}_{j,k}^{*}-\phi_j\rangle^2 = o_p(1).
$$

Collecting the above arguments it thus follows $\sum_{j=1}^{\tilde{m}}(\tilde{a}_j^{(2)})^2 = O_p(n^{-1})$ and hence

$$
||\tilde{A}_k - A_k||_2^2 = O_p(\max\left\{n^{\frac{(1+\delta)(1-2\nu)}{\mu+2\nu}}, n^{-1}\right\}) = O_p(n^{-1})
$$

as $(n,T)\to\infty$ using Lemma 2. Together with our Remark on the classification error the desired result follows. ■

B. Further Simulation Results

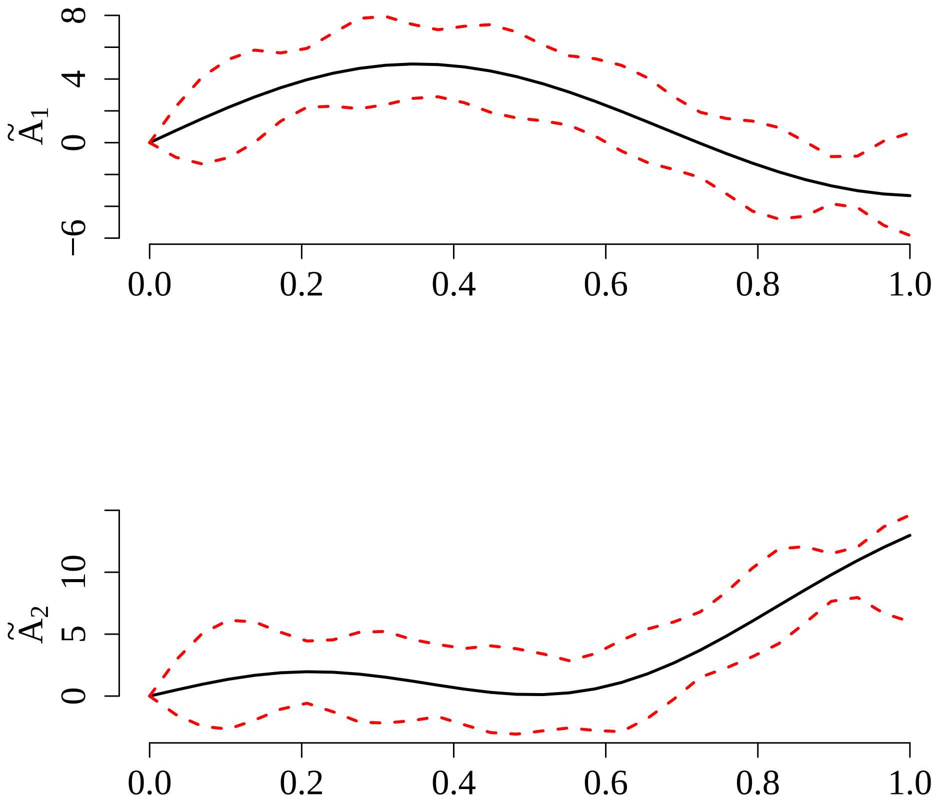

Figure 5: *Scenario 1: Estimated regime parameter functions for sample size* $(n, T) = (50, 50)$.

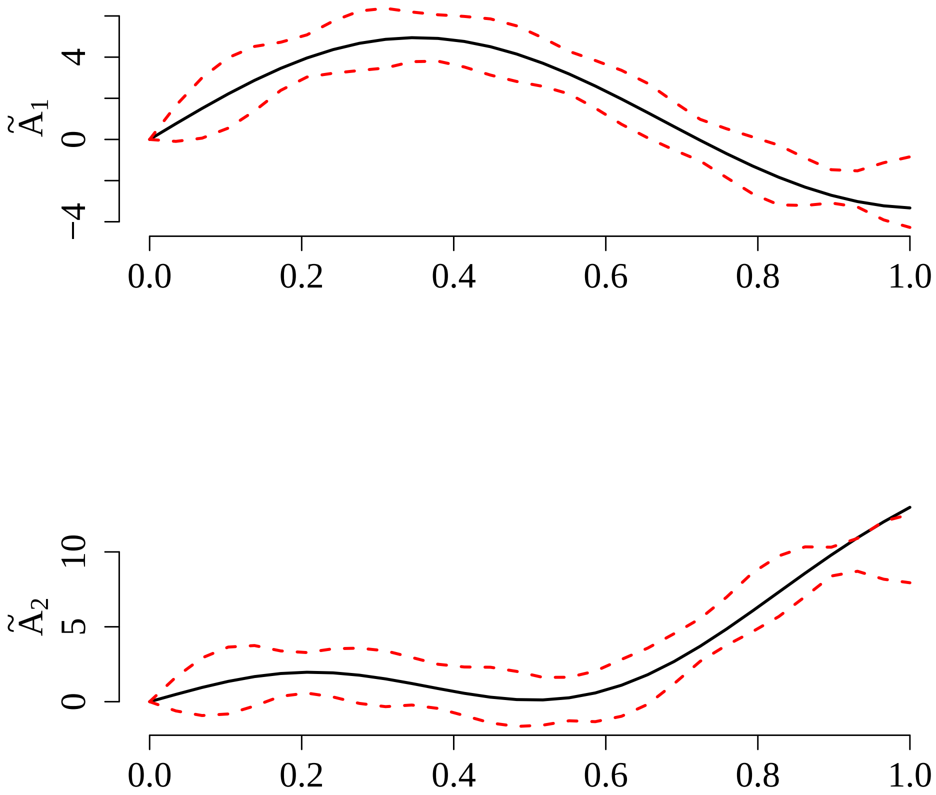

Figure 6: *Scenario 1: Estimated regime parameter functions for sample size* $(n, T) = (100, 50)$.

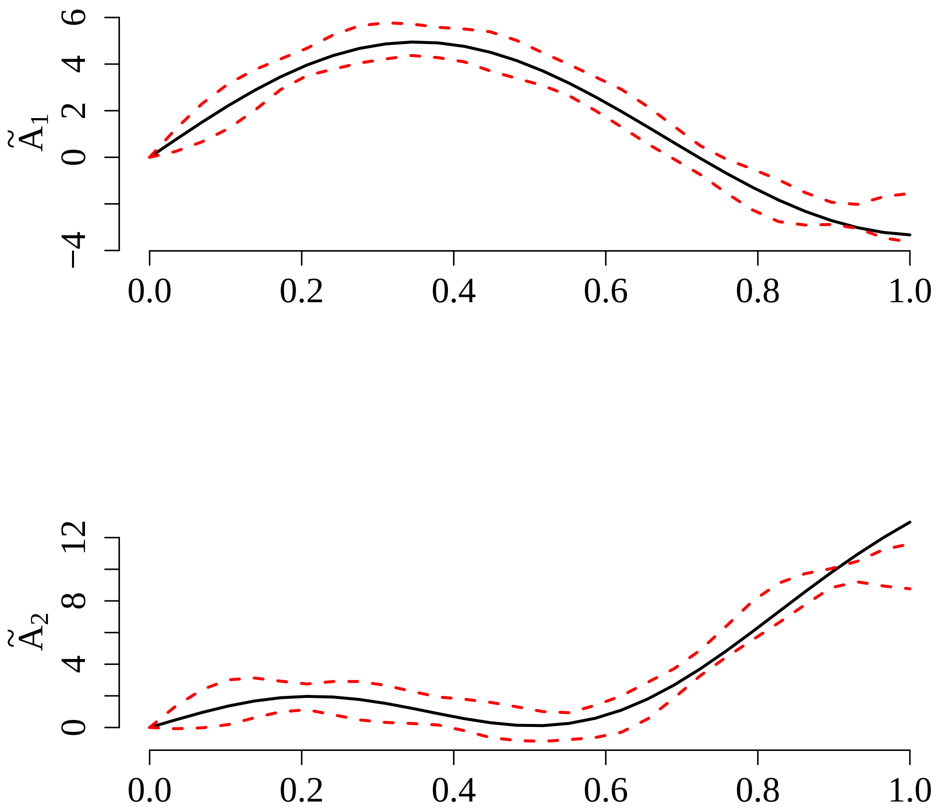

Figure 7: *Scenario 1: Estimated regime parameter functions for sample size* $(n, T) = (150, 80)$.

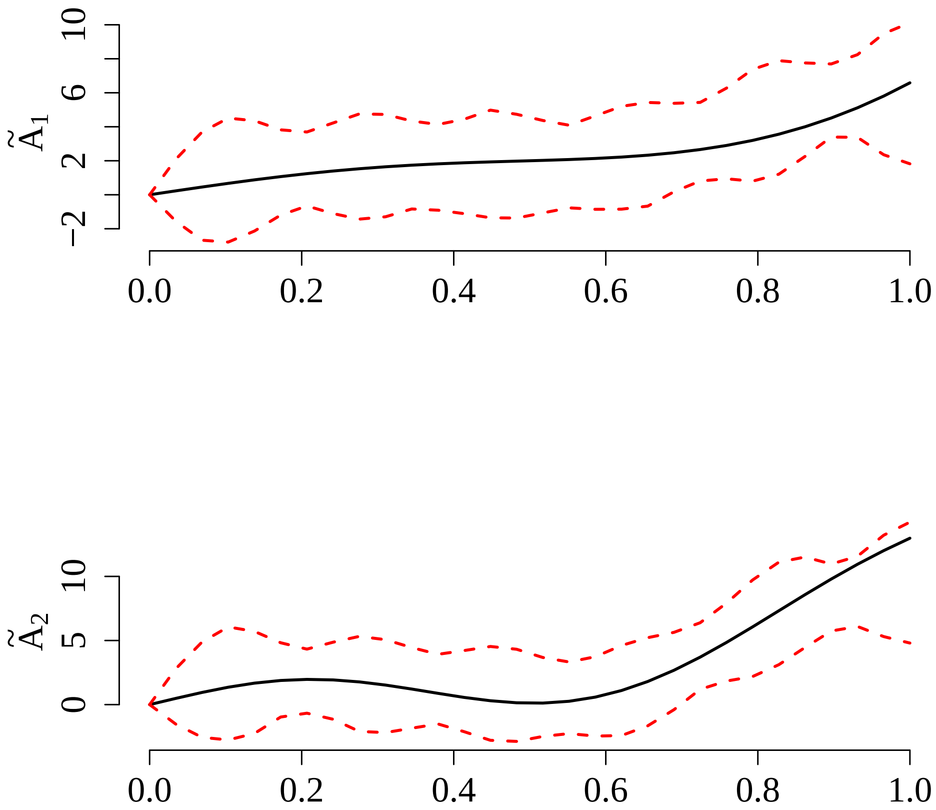

Figure 8: *Scenario 2: Estimated regime parameter functions for sample size* $(n, T) = (50, 50)$.

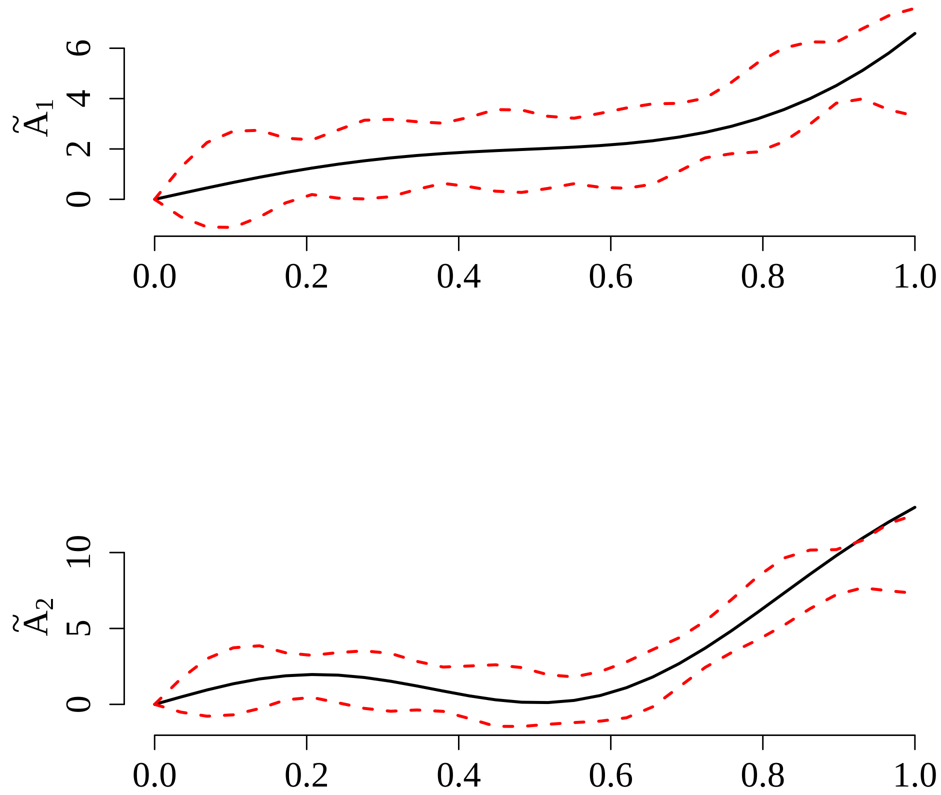

Figure 9: *Scenario 2: Estimated regime parameter functions for sample size* $(n, T) = (100, 50)$.

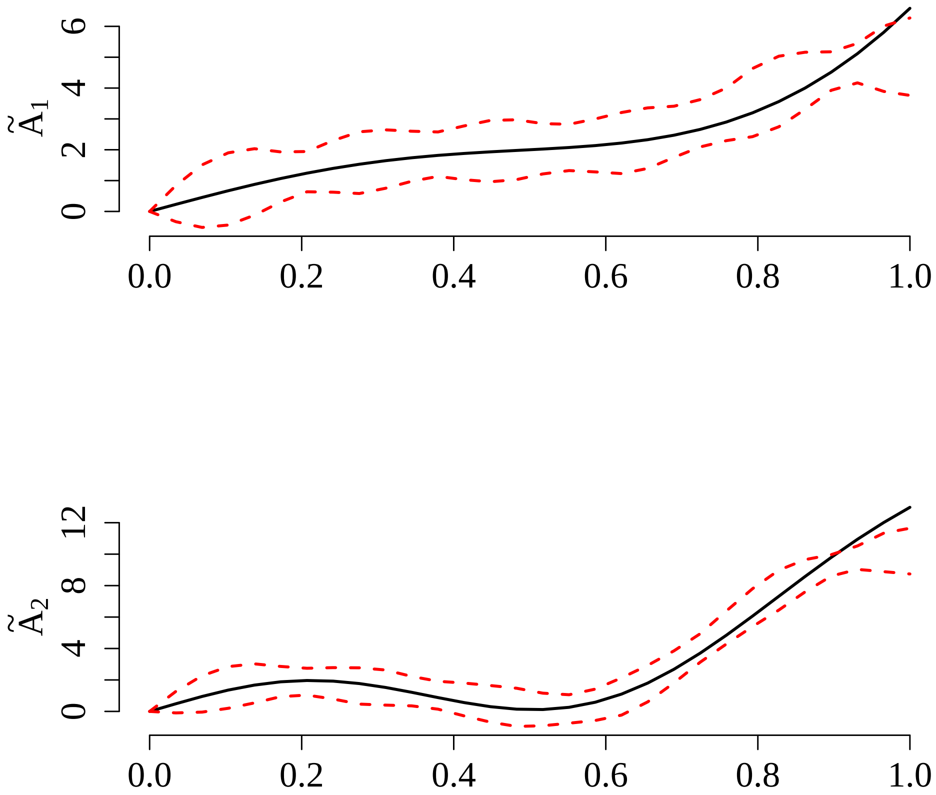

Figure 10: *Scenario 2: Estimated regime parameter functions for sample size* $(n, T) = (150, 80)$.

C. Construction of Volatility Curves

In what follows we argue along the lines of Müller et al. (2011) under which conditions the central approximation in (22) holds in our context. For this purpose note that given some asset i and a period t for an $0 < s < 1 - \Delta$

$$\begin{aligned} Y_{it,\Delta}(s) &:= \Delta^{-\frac{1}{2}} \log\left(\frac{P_{it}(s+\Delta)}{P_{it}(s)}\right) \\ &= \Delta^{\frac{1}{2}} \mu_{it}(s) + \Delta^{\frac{1}{2}} \alpha_t(s) x_{it}^2(s) + x_{it}(s) W_{it,\Delta}(s) + R_{it,\Delta,1}(s) + R_{it,\Delta,2}(s) + R_{it,\Delta,3}(s) \end{aligned}$$

where

$$\begin{aligned} W_{it,\Delta}(s) &:= \Delta^{-\frac{1}{2}} \left(W_{it}(s) - W_{it}(s-\Delta)\right) \\ R_{it,\Delta,1}(s) &:= \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} \mu_{it}(v) \mathrm{d}v - \Delta^{\frac{1}{2}} \mu_{it}(s) \\ R_{it,\Delta,2}(s) &:= \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} \alpha_t(v) x_{it}^2(v) \mathrm{d}v - \Delta^{\frac{1}{2}} \alpha_t(s) x_{it}^2(s) \\ R_{it,\Delta,3}(s) &:= \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} x_{it}(v) \mathrm{d}W_{it,v} - x_{it}(s) W_{it,\Delta}(s). \end{aligned}$$

The objects $R_{it,\Delta,1} - R_{it,\Delta,3}$ discretization errors that will be shown to become small as the grid size Δ^{-1} increases. This is guaranteed by the following conditions.

> **Assumption 8.** Suppose for all $1 \leq i \leq n$ and $1 \leq t \leq T$ conditions M1-M4 hold for the drift μ_{it} and the volatility x_{it}. Further suppose there exists a $C < \infty$, such that $|\alpha_t(u) - \alpha_t(v)| \leq C|u - v|$ a.s. for all $u, v \in [0, 1]$, i.e. α_t is Lipschitz continuous of order one. Further suppose $\sup_{s\in[0,1]} |\alpha_t(s)| < c$ for some $c < \infty$.

Based on these conditions we conclude the behavior of the discretization errors.

Lemma 3 *Given Assumption 8, it holds that*

$$\sum_{j=1}^{3} E\left[\sup_{s\in[0,1]} R_{it,\Delta,j}(s)\right] = O_p(\Delta^{1/2}) \quad \textit{as} \quad \Delta \to 0. \tag{63}$$

Proof of Lemma 3. As shown in Müller et al. (2011) it holds that $R_{it,\Delta,1} = O_p(\Delta^{3/2})$ and $R_{it,\Delta,3} = O_p(\Delta^{1/2})$. For $R_{it,\Delta,2}$ it holds that

$$
\begin{aligned}
R_{it,\Delta,2} &= \left| \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} \alpha_t(v) x_{it}^2(v) \mathrm{d}v - \Delta^{\frac{1}{2}} \alpha_t(s) x_i t^2(s) \right| \\
&= \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} \left| \alpha_t(v) x_{it}^2(v) - \alpha_t(s) x_{it}^2(s) \right| \mathrm{d}v \\
&= \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} \left| (\alpha_t(v) - \alpha_t(s)) x_{it}^2(v) + \alpha_t(s)(x_{it}^2(v) - x_{it}^2(s)) \right| \mathrm{d}v \\
&\leq \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} \left| (\alpha_t(v) - \alpha_t(s)) x_{it}^2(v) \right| \mathrm{d}v + \Delta^{-\frac{1}{2}} \int_s^{s+\Delta} \left| \alpha_t(s)(x_{it}^2(v) - x_{it}^2(s)) \right| \mathrm{d}v \\
&= O_{a.s.}(\Delta^{\frac{3}{2}})
\end{aligned}
$$

where the $O_{a.s.}$ term is uniform in s. Using a similar argument as in Müller et al. (2011) the result follows. ■

This justifies the small-Δ approximation as indicated in (22) as the terms $\Delta^{\frac{1}{2}} \mu_{it}(s)$ and $\Delta^{\frac{1}{2}} \alpha_t(s) x_{it}^2(s)$ become negligible as $\Delta \to 0$.

Remark In section 7, we argued that it is plausible to proxy x_{it} by $\log(x_{it}) = X_{it}$ in the regression context. Heuristically, the pointwise plausibility of this becomes visible employing a first order Taylor expansion around some $c > 0$. The latter supports $\log(x_{it}) \approx \log(c) + c^{-1}(x_{it} - c)$ for values around unity. This in turn motivates the approximation

$$
\int_0^1 \alpha_t(u) X_{it}(u) \mathrm{d}u \approx - \int_0^1 \alpha_t(u) \mathrm{d}u + c^{-1} \int_0^1 \alpha_t(u) x_{it}(u) \mathrm{d}u.
$$

where the first summand ultimately contributes to the model constant, which is not of inherent interest. For values of c close to one the approximation is particularly reasonable.

D. Empirical Results

	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$\bar{x}$	s
y	-1.74	-0.56	0.00	0.60	1.79	0.02	1.16
$\int_0^1 X(u)\mathrm{d}u$	-16.38	-15.71	-15.18	-14.60	-13.64	-15.12	0.84
$\|\|X\|\|_2^2$	187.20	214.10	231.55	247.80	269.27	230.34	25.13
z	0.02	0.02	0.03	0.05	0.09	0.04	0.03
$\int_0^1 \tilde{E}_{t-1}[X](u)\mathrm{d}u$	-16.16	-15.63	-15.20	-14.69	-13.84	-15.12	0.72
$\|\|\tilde{E}_{t-1}[X]\|\|_2^2$	192.11	216.22	231.45	244.78	261.77	229.70	21.54
$\tilde{E}_{t-1}[z]$	0.02	0.02	0.03	0.05	0.09	0.04	0.02

Table 3: **Summary Statistics.** The quantity q_τ denotes the τ-quantile of the empirical distribution, $\bar{x}$ and s are corresponding mean and standard deviation. The quantities y and z are in %.

Regime k	$\lvert\hat{G}_k\rvert$	$\int_0^1 \tilde{A}_k(u)\mathrm{d}u$	$\lvert\lvert\tilde{A}_k\rvert\rvert_2^2$	$q^{\beta}_{0.05}$	$q^{\beta}_{0.25}$	$q^{\beta}_{0.5}$	$q^{\beta}_{0.75}$	$q^{\beta}_{0.95}$	$\bar{x}^{\beta}$	s^{β}
1	15	0.59	0.48	-103.08	-86.48	-39.24	7.14	67.63	-32.62	61.92
2	57	0.78	0.93	-234.25	-123.75	-50.15	16.42	110.25	-58.18	116.11
3	18	-0.89	1.10	31.92	41.39	47.05	68.80	80.83	54.09	17.35
4	1	-0.23	0.48	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	
5	12	1.01	1.18	-72.49	-61.93	-54.47	-43.54	-38.11	-54.17	13.04
6	20	-1.58	2.80	52.08	74.59	91.32	130.16	161.09	99.11	38.06
7	3	1.03	1.32	-29.61	-24.90	-19.01	-15.51	-12.71	-20.61	9.49
8	3	-0.96	1.19	11.26	15.85	21.58	23.32	24.71	18.92	7.82
9	1	0.59	3.29	-1.09	-1.09	-1.09	-1.09	-1.09	-1.09	
10	1	0.10	0.95	3.26	3.26	3.26	3.26	3.26	3.26	
11	1	-1.33	12.10	-7.13	-7.13	-7.13	-7.13	-7.13	-7.13	
12	1	-1.04	5.56	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	
13	1	0.98	2.06	2.81	2.81	2.81	2.81	2.81	2.81	
14	1	-0.19	1.80	3.10	3.10	3.10	3.10	3.10	3.10	

Table 4: **Estimated Regimes and Parametric Nuissance.** The quantity q^{β}_{τ} denotes the τ-quantile of the empirical distribution of the $\tilde{\beta}_{t,k}$ within regime k, $\bar{x}^{\beta}$ and s^{β} are corresponding mean and standard deviation.

Regime k	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	s
1	-0.60	-0.30	-0.06	0.25	0.78	0.43
2	-0.59	-0.29	-0.05	0.24	0.74	0.42
3	-1.19	-0.39	0.07	0.46	0.97	0.67
4	-0.79	-0.35	-0.09	0.33	0.87	0.52
5	-1.25	-0.41	0.08	0.48	0.97	0.69
6	-0.78	-0.39	-0.06	0.32	1.00	0.55
7	-1.24	-0.39	0.08	0.51	0.94	0.68
8	-0.72	-0.40	-0.08	0.31	0.97	0.55
9	-0.83	-0.42	-0.07	0.34	1.14	0.61
10	-0.77	-0.40	-0.06	0.36	1.00	0.56
11	-0.68	-0.32	-0.03	0.27	0.80	0.46
12	-0.56	-0.30	-0.06	0.25	0.73	0.42
13	-1.19	-0.33	0.07	0.47	0.87	0.67
14	-0.96	-0.53	-0.10	0.42	1.35	0.71

Table 5: **Risk Premiums.** The table reports distributional features of the estimated risk premiums $\int_0^1 \tilde{A}_k(u)X_{it}(u)\mathrm{d}u$ within the regimes. Column names in analogy to before.

References

Anderson, R.M. (2011). “Time-varying risk premia?” *Journal of Mathematical Economics.*, 47: 253–259.

Ang, A. and Hodrick, A.R. and Xing, Y. and Zhang, X. (2006). “The Cross-Section of Volatility and Expected Returns?” *The Journal of Finance.*, 61 (1): 259–299.

Bada O. and Gualtieri J. and Kneip A. and Sickles R.C. (2015). “Panel Data Models with Multiple Jump Discontinuities in the Parameters.” *Working Paper.*.

Baker, S.R. and Bloom, N. and Davis, S.J. (2016). “Measuring economic policy uncertainty.” *The Quarterly Journal of Economics.*, 131 (4): 1593–1636.

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

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

Barndorff-Nielsen, O.E. and Shephard, N. (2002). “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models.” *Journal of the Royal Statistical Society. Series B.*, 64 (2): 253–280.

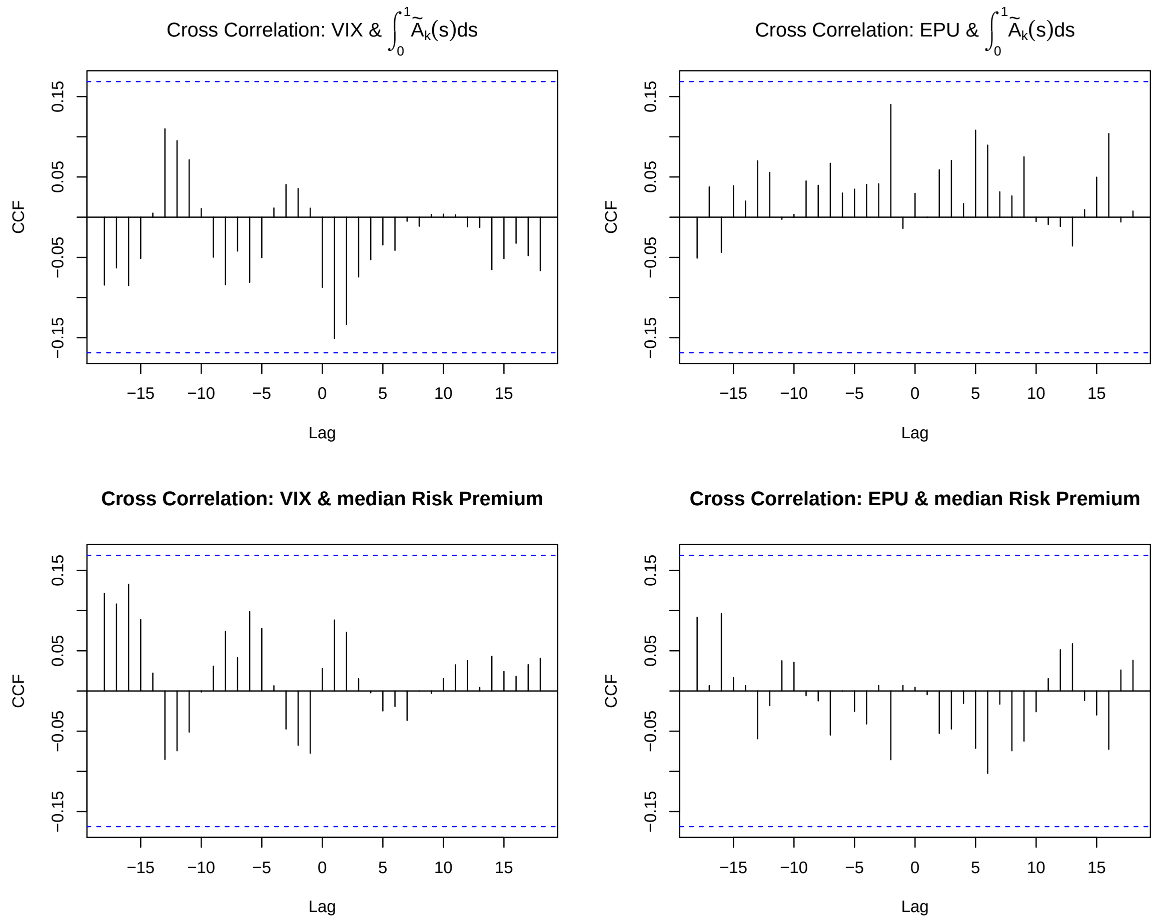

Figure 11: *Cross Correlations.* Median Risk Premiums are calculated as medians of the empirical distribution of $\{\int_0^1 \tilde{A}_k(u)X_{it}(u)du, \quad 1 \leq i \leq n\}$ at a day t.

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

Hastie, T. and Tibshirani, R. and Friedman, J. (2009). "The Elements of Statistical Learning." *Springer, Berlin: Springer series in statistics.*, Second Edition.

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

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

Kneip, A. and Poss, D. and Sarda, P. (2016). "Functional linear Regression with Points of Impact." *The Annals of Statistics*, 44 (1): 1–30.

Hamilton, J. D. (2010). "Regime switching models." In Macroeconometrics and Time Series Analysis, pp. 202-209. Palgrave Macmillan UK.

Hall, P. and Hooker, G. (2016). "Truncated linear models for functional data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78 (3): 637 – 653.

Hall, P. and Horowitz, J.L. (2007). "Methodology and Convergence Rates for Functional Linear Regression." *The Annals of Statistics*, 35 (1): 70 –91.

Horváth, L. and Kokoszka, P. and Rice, G. (2014). "Testing stationarity of functional time series." *Journal of Econometrics*, 179 (1): 66– 82.

Ghysels, E. and Santa-Clara, P. and Valkanov, R. (2004). "The MIDAS touch: Mixed data sampling regression models." *Working Paper*.

Hörmann, S. and Kidziński, L. and Kokoszka, P. (2015). "Estimation in functional lagged regression." *Journal of Time Series Analysis*, 36 (3): 541–561.

Hall, P. and Hosseini-Nasab, M. (2006). "On properties of functional principal components analysis." *Journal of the Royal Statistical Society B*, 68 (1): 109 –126.

Ramsay, J. and Silverman, B. W. (2005) **Functional Data Analysis.** Springer Science & Business Media.

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

Su, L. and Shi, Z. and Phillips, P.C.B. (2017). "Identifying latent structures in panel data." *Econometrica*, 84(6), 2215-2264.

Vogt, M. and Linton, O. (2017). "Classification of Nonparametric Regression Functions in Longitudinal Data Models." *Journal of the Royal Statistical Society: Series B.*, 79 (1).

Yao, F. and Müller, H. G. and Wang, J. L. (2005). "Functional data analysis for sparse longitudinal data." *Journal of the American Statistical Association*, 100 (470), 577-590.

Zeileis, A. and Hornik, K. and Smola, A. and Karatzoglou, A. (2004). "kernlab-an S4 package for kernel methods in R." *Journal of Statistical Software*, 11 (9): 1 –20.