Supplemental Paper for:

Parameter Regimes in Partial Functional Panel Regression

by Dominik Liebl and Fabian Walders

Throughout this appendix we use the symbols C and c to denote generic positive constants.

A Technical Appendix

In this part we use the following notation for norms in addition to the ones introduced in the main paper. Given a mapping $F_1 : L^2([0,1]) \to \mathbb{R}$, we use as norm of F_1 the operator norm $||F_1||_{H'} := \sup_{||f_1||_2=1} |F_1(f_1)|$. Further, for an integral operator $F_2 : L^2([0,1]) \to L^2([0,1])$ with kernel $f_2 \in L^2([0,1] \times [0,1])$, denote its Hilbert-Schmidt norm as $||F_2||_{\mathcal{S}} := ||f_2||_2$, where in this case $||\cdot||_2$ is the L^2 norm in $L^2([0,1] \times [0,1])$.

For the following proofs we make use of the following closed form solutions of our least squares estimators $\hat{a}_{j,t}$ and $\hat{\beta}_t$:

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

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

where

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

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

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

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

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

For the sake of readability we will proof the lemma and theorems for $P = 1$, while the generalization to $P > 1$ is straightforward and does not add any additional insights. In this spirit we ease our notation by dropping boldface notation and the dependence on coordinate labels p.

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

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

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

$$E\left[\left|\left|\hat{K}_{X,t} - K_{X,k}\right|\right|_2^2\right] = O(n^{-1}), \tag{8}$$

where the index k is such that $t \in G_k$, which we use in what follows without further reference. In Equation (8) $K_{X,k}$ denotes the covariance function in the k-th regime, i.e. $K_{X,k}(u, v) := E[(X_{it}(u) - E[X_{it}](u))(X_{it}(v) - E[X_{it}](v))]$ and in analogy $K_{z,k} := E[(z_{it} - E[z_{it}])^2]$. Further, it obviously holds that

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

where

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

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

$$E\left[\left|\hat{\lambda}_{j,t} - \lambda_{j,k}\right|^q\right] = O(n^{-q/2}) \ \ q = 1, 2 \ldots \tag{9}$$

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

As a final observation, note that combining the results in Shin (2009) and Hall and Horowitz (2007) allows to conclude that for any $1 \leq t \leq T$

$$||\hat{\Phi}_t - \Phi_k||^2_{H'} = \left|\left|\sum_{j=1}^{m} \frac{\langle \hat{K}_{zX,t}, \hat{\phi}_{j,t}\rangle}{\hat{\lambda}_{j,t}} \hat{\phi}_{j,t} - \sum_{j=1}^{\infty} \frac{\langle K_{zX,k}, \phi_{j,k}\rangle}{\lambda_{j,k}} \phi_{j,k}\right|\right|^2_2$$
$$= O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right), \tag{10}$$

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

$$\Phi_k(g) := \sum_{j=1}^{\infty} \frac{\langle K_{zX,k}, \phi_{j,k}\rangle}{\lambda_{j,k}} \langle \phi_{j,k}, g\rangle \tag{11}$$

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

A.1 Proof of Theorem 4.1

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

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

with $\hat{B}_t := [\hat{K}_{z,t} - \hat{\Phi}_t(\hat{K}_{zX,t})]$. Regarding the inverse in $\hat{\beta}_t$ note that it follows from (6), (7), and (10) in analogy to Shin (2009) that

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

as $n \to \infty$, which certainly implies $\hat{B}_t^{-1} = B_k^{-1} + o_p(1)$ by the continuous mapping theorem, whereas $B_k = E[\mathbf{s}_{it}^2] > 0$ follows from Assumption 6. To see this also consider the decomposition shown in (25). As in Shin (2009), we assess the difference

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

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

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

where, in analogy to her work,

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

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

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

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

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

as a consequence of (6). The second term, $R_{1,2,t}$, in $R_{0,3,t}$ is defined as

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

with

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

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

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

with

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

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

and

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

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

$$||\hat{\phi}_{t,j} - \phi_{j,k}||_2^q \leq \left[\frac{8^{1/2} \mathcal{D}_t}{\min_{1 \leq l \leq j} \{\lambda_{j,k} - \lambda_{j+1,k}\})} \right]^q \quad \text{almost surely} \tag{12}$$

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

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

of which the second coincides with their work and the first one is a straightforward derivative of their arguments. Denoting the complement of a set A as A^c, note that $\mathbb{P}(\mathcal{F}^c_{1,n,t}) = o(1)$ as well as $\mathbb{P}(\mathcal{F}^c_{2,n,t}) = o(1)$ due Assumptions 4–5 and root-n consistency of the empirical covariance operator and its corresponding eigenvalues as well as assuming the constants in $\mathcal{F}_{1,n,t}$ and $\mathcal{F}_{2,n,t}$ to be appropriate. Since $\mathbb{P}(\mathcal{F}^c_{3,n,t}) \leq \mathbb{P}(\mathcal{F}^c_{1,n,t}) + \mathbb{P}(\mathcal{F}^c_{2,n,t})$, we conclude $\mathbb{P}(\mathcal{F}^c_{3,n,t}) = o(1)$. We also show that this property holds uniformly over $1 \leq t \leq T$ as $(n, T) \to \infty$ in the proof of Lemma 4.1 below. Equation (5.21) in Hall and Horowitz (2007), reads in our notation as

$$||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 \leq 8 \left(1 - 4 C n^{\frac{2(1+\mu)}{\mu+2\nu}} \mathcal{D}_t^2 \right)^{-1} R^{(\phi)}_{j,t}, \tag{13}$$

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

The inequality in (13) is valid on $\mathcal{F}_{2,n,t}$, whereas the constant C on the right hand side is the constant in $\mathcal{F}_{1,n,t}$. On this event $\mathcal{F}_{1,n,t}$ it further holds that

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

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

$$||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 \leq 16R_{j,t}^{(\phi)}. \tag{14}$$

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

$$E\left[R_{j,t}^{(\phi)}\right] = O\left(j^2 n^{-1}\right) \tag{15}$$

uniformly in $1 \leq j \leq m$ (see also the corresponding proof of Equation (5.22) in Section 5.3 in Hall and Horowitz, 2007). Note that (14) obviously implies that on $\mathcal{F}_{3,n,t}$,

$$||\hat{\phi}_{j,t} - \phi_{j,k}||_2 \leq 4\left(R_{j,t}^{(\phi)}\right)^{\frac{1}{2}} \tag{16}$$

of which the right hand side has the property $E\left[\left(R_{j,t}^{(\phi)}\right)^{1/2}\right] \leq E\left[R_{j,t}^{(\phi)}\right]^{1/2} = O\left(jn^{-1/2}\right)$ uniformly over $1 \leq j \leq m$, what follows from Jensen's inequality and (15).

These observations imply that

$$\begin{aligned}\mathbb{P}\left(nm^{-3}\sum_{j=1}^{m}||\hat{\phi}_{j,t} - \phi_{j,k}||_2^2 > c\right) &\leq \mathbb{P}\left(16nm^{-3}\sum_{j=1}^{m}R_{j,t}^{(\phi)} > c\right) + \mathbb{P}\left(\mathcal{F}_{3,n,t}^c\right) \\ &\leq \frac{nm^{-3}\sum_{j=1}^{m}E\left[R_{j,t}^{(\phi)}\right]}{c/16} + o(1)\end{aligned} \tag{17}$$

by the Markov inequality. The numerator on the right hand side of (17) is bounded above as a consequence of (15) and Assumptions 4 & 5, and thus $\sum_{j=1}^{m}||\phi_{j,t} - \phi_{j,k}||_2^2 = O_p\left(n^{-1}m^3\right)$. From this and Assumptions 4 & 5, of which the former is slightly stronger than in Hall and Horowitz (2007) and Shin (2009), we conclude for the first summand in $R_{3,3,t}$,

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

because $\nu > 3 - \mu/2$. Note that from our observations for (16), we can further conclude

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

using similar arguments as before. We have $n^{-1/2}m^2 = n^{\frac{2-\mu/2-\nu}{\mu+2\nu}} = o(1)$ by Assumption 4, which allows to conclude in combination with (6) and Assumption 2, that $R_{3,1,t} = O_p(n^{-1/2})$.

Using similar arguments as for (17), allows us to conclude for the second term in $R_{3,3,t}$:

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

where the numerator on the right hand side of the last inequality is bounded above thanks to Assumptions 4–5 as well as our observation in (16). An analogue argument shows $R_{3,2,t} = O_p(n^{-1/2})$ (see also points 3 and 4 in Assumption 2 to see this).

Combining arguments implies $\hat{\beta}_t-\beta_t = O_p(n^{-1/2})$ for every $1\le t\le T$, which concludes the proof of the first result in Theorem 4.1. Turning to $\hat{\alpha}_t$ note that

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

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

$$\sum_{j=1}^{m}(\hat{a}_{j,t}-a^*_{j,t})^2\le 2\sum_{j=1}^{m}(\hat{\lambda}^{-1}_{j,t}\langle\hat{K}^{\#}_{yX,t}-\hat{\beta}_t\hat{K}^{\#}_{zX,t},\hat{\phi}_{j,t}\rangle-a^*_{j,t})^2+2\sum_{j=1}^{m}(\hat{\lambda}^{-1}_{j,t}\langle r_{y,t}r_{x,t}-\hat{\beta}_t r_{z,t}r_{x,t},\hat{\phi}_{j,t}\rangle)^2 \tag{18}$$

with $\hat{K}^{\#}_{yX,t} := n^{-1}\sum_{i=1}^{n}(y_{it}-E[y_{it}])(X_{it}-E[X_{it}])$, $\hat{K}^{\#}_{zX,t} := n^{-1}\sum_{i=1}^{n}(z_{it}-E[z_{it}])(X_{it}-E[X_{it}])$, $r_{x,t} := E[X_{it}]-\bar{X}_t$, $r_{y,t} := E[y_{it}]-\bar{y}_t$ and $r_{z,t} := E[z_{it}]-\bar{z}_t$. Note that $||r_{x,t}||_2, |r_{y,t}|$ and $|r_{z,t}|$ all correspond to errors from parametric estimation problems and are thus of order $n^{-1/2}$. Bounds on $\hat{\lambda}_{j,t}-\lambda_{j,k}$ as well as $||\hat{\phi}_{j,t}-\phi_{j,k}||_2$ are asymptotically equivalent for data centered around their arithmetic mean and data centered around their population expectation. Together with the above arguments it follows that the first term in (18) is asymptotically equivalent to the corresponding term in Shin (2009), implying $\sum_{j=1}^{m}(\hat{\lambda}^{-1}_{j,t}\langle\hat{K}^{\#}_{yX,t}-\hat{\beta}_t\hat{K}^{\#}_{zX,t},\hat{\phi}_{j,t}\rangle-a^*_{j,t})^2 = O_p\left(n^{\frac{1-2\nu}{\mu+2\nu}}\right)$. Now, define the event

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

for which we conclude $\mathbb{P}(\mathcal{F}^c_{4,n,t}) = o(1)$ for any $1\le t\le T$ as $n\to\infty$ as a consequence of (9). On this event the second term in (18) can be bounded according to

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

Finally combining arguments yields $||\hat{\alpha}_t - \alpha_t||_2^2 = O_p(n^{\frac{1-2\nu}{\mu+2\nu}})$ for any $1 \leq t \leq T$ as $n \to \infty$, which concludes the proof of the second part of Theorem 4.1. ■

A.2 Proof of Lemma 4.1

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

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

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

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

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

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

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq t\leq T} |\hat{K}_{z,t} - K_{z,k}|^2 > c\right) \\
&\leq K \max_{1\leq k\leq K} |G_k| \frac{n^{-1}E\left[|(z_{it} - E[z_{it}])^2 - K_{z,k}|^2\right]}{c} + K \max_{1\leq k\leq K} |G_k| \frac{E\left[(\bar{z}_t - E[z_{it}])^4\right]}{c} \\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1) \qquad (19)
\end{aligned}$$

  and similarly

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

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

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\le t\le T}||\hat{K}_{X\epsilon,t}||_2^2 > c\right) \le \sum_{t=1}^{T}\mathbb{P}\left(||\hat{K}^{\#}_{X\epsilon,t}||_2^2 > c/4\right) + \sum_{t=1}^{T}\mathbb{P}\left(||r_{x,t}||_2^2 r_{\epsilon,t}^2 > c/4\right)\\
&\le K\max_{1\le k\le K}|G_k|\frac{n^{-1}\sigma_{\epsilon,k}^2 E\left[||X_{it}-E[X_{it}]||_2^2\right]}{c} + K\max_{1\le k\le K}|G_k|\frac{E[(\bar{\epsilon}_t)^2]E\left[||\bar{X}_t - E[X_{it}]||_2^2\right]}{c}\\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1) \qquad (20)
\end{aligned}
$$

Similar arguments can be used to show

$$
\mathbb{P}\left(\max_{1\le t\le T}|\hat{K}_{z\epsilon,t}|^2 > c\right) = O(n^{\delta-1}) + O(n^{\delta-2}) = o(1). \qquad (21)
$$

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

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\le t\le T}||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c\right)\\
\le&\sum_{t=1}^{T}\mathbb{P}\left(||\hat{K}^{\#}_{zX,t} - K_{zX,k}||_2^2 > c\right) + \sum_{t=1}^{T}\mathbb{P}\left(||r_{x,t}||_2^2 r_{z,t}^2 > c\right)\\
\le&K\max_{1\le k\le K}|G_k|\frac{n^{-1}E\left[||(z_{it}-E[z_{it}])(X_{it}-E[X_{it}]) - K_{zX,k}||_2^2\right]}{c}\\
&+K\max_{1\le k\le K}|G_k|\frac{E\left[r_{z,t}^2\right]E\left[||r_{x,t}||_2^2\right]}{c}\\
&= O(n^{\delta-1}) + O(n^{\delta-2}) = o(1). \qquad (22)
\end{aligned}
$$

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

  To see this, note that $E[R_{j,t}^{(\phi)}]$ as in (15) does not vary over the index $t \in G_k$ within a regime k, but potentially across regimes $k = 1, \dots, K$. This due to the stationarity of the functional regressor within regimes as postulated in Assumption 1. We thus conclude:

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\le t\le T}\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\phi_{j,k}-\hat{\phi}_{j,t}||_2^2>c\right)\\
\le&\sum_{t=1}^{T}\mathbb{P}\left(\sum_{j=1}^{m}\hat{\lambda}_{j,t}^{-2}||\phi_{j,k}-\hat{\phi}_{j,t}||_2^2>c\right)\\
\le&\sum_{t=1}^{T}\left(\mathbb{P}\left(16\cdot 4C_\lambda^{-2}\sum_{j=1}^{m}j^{2\mu}R_{j,t}^{(\phi)}>c\right)+\mathbb{P}\left(\mathcal{F}_{1,n,t}^c\right)+\mathbb{P}\left(\mathcal{F}_{2,n,t}^c\right)+\mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right)\right)\\
\le&K\max_{1\le k\le K}|G_k|\frac{\sum_{j=1}^{m}E\left[R_{j,t}^{(\phi)}\right]j^{2\mu}}{c\cdot C_\lambda^2/(16\cdot 4)}+\sum_{t=1}^{T}\left(\mathbb{P}\left(\mathcal{F}_{1,n,t}^c\right)+\mathbb{P}\left(\mathcal{F}_{2,n,t}^c\right)+\mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right)\right)\\
=&O\left(Tn^{-1}m^{3+2\mu}\right)+\sum_{t=1}^{T}\left(\mathbb{P}\left(\mathcal{F}_{1,n,t}^c\right)+\mathbb{P}\left(\mathcal{F}_{2,n,t}^c\right)+\mathbb{P}\left(\mathcal{F}_{4,n,t}^c\right)\right), \qquad (23)
\end{aligned}
$$

where C_λ is the constant from point 1 in Assumption 2. To obtain the second inequality, we used once more that $\hat{\lambda}_{j,t}\ge\lambda_{j,k}/2$ for $1\le j\le m$ on $\mathcal{F}_{4,n,t}$. The sequence in (23) is a null sequence because on the one hand

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

thanks to Assumption 4 and on the other hand since

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

as we argue next. First we observe

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

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

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

by the fact that $\mathcal{D}_t \ge \sup_j |\hat{\lambda}_{j,t} - \lambda_{j,k}|$ almost surely as well as Assumptions 2–5. In lines of our arguments from the proof of Theorem 4.1, we conclude $K \max_{1\le k\le K} |G_k|\mathbb{P}(\mathcal{F}^c_{3,n,t}) = o(1)$. Beyond that it holds

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

again thanks to Assumption 4. Note that our result in (23) implies in particular that

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

which will be used without further reference in what follows.

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

Now, turning to our concrete arguments for $\hat{\alpha}_t^{(\Delta)}$, we note that the scaling which distinguishes $\hat{\alpha}_t^{(\Delta)}$ from $\hat{\alpha}_t$ is composed by $\hat{\sigma}_{\epsilon,t}$ and the empirical eigenvalues $\hat{\lambda}_{j,t}$,$1 \le j \le m$. While the latter can be treated in a comparably simple way, $\hat{\sigma}_{\epsilon,t}$ requires closer attention. We thus begin focusing on this object and its constituents. For this purpose, define the event $\mathcal{S}_{n,t}$ for later use according to

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

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

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

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

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

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

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

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

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

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

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

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

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

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

we note this expression can be most easily handled using the almost sure bound in (12) according to

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

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

$$\mathbb{P}\left(\max_{1\le t\le T} ||\hat{\Phi}_t - \Phi_k||_{H'}^2 > c \right) = o(1). \tag{24}$$

To proceed we work again on the differences $(\hat{\beta}_t - \beta_t) = \hat{B}_t^{-1}\left(R_{0,1,t} + R_{0,2,t} + R_{0,3,t}\right)$ with $R_{0,1,t}$, $R_{0,2,t}$ and $R_{0,3,t}$ as in the proof of Theorem 4.1. Addressing the inverse in these differences, define the t-wise event $Q_{n,t} := \{|\hat{B}_t - B_k| \le \frac{1}{2}B_k\}$. For this event, note that $\sum_{t=1}^{T} \mathbb{P}\left(Q_{n,t}^c\right) \le R_{6,1} + R_{6,2}$, where $R_{6,1} := \sum_{t=1}^{T} \mathbb{P}\left(|\hat{K}_{z,t} - K_{z,k}|^2 > c\right) = o(1)$ as shown in (19). For $R_{6,2}$ we use the arguments in Shin (2009) to obtain

$$R_{6,2} := \sum_{t=1}^{T} \mathbb{P}\left(||\hat{\Phi}_t - \Phi_k||_{H'}^2 ||K_{zX,k}||_2^2 + (||\hat{\Phi}_t - \Phi_k||_{H'} + ||\Phi_k||_{H'})^2 ||\hat{K}_{zX,t} - K_{zX,k}||_2^2 > c \right) \tag{25}$$

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

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

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

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

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

For this we argue

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

due to (20) and (21). Further note for $R_{0,2,t}$

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

as a consequence of (20) and (24). For the remaining terms, we argue along the same lines as in the proof of Theorem 4.1:

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

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

because of (22). The remaining term was shown to be bounded according to $R_{1,2} \leq R_{2,1,t} + R_{2,2,t}$, where the two summands are defined above. While $R_{2,1} = O(n^{-1/2})$ deterministically and independently of t, note for the second summand $R_{2,2,t} \leq R_{3,1,t} + R_{3,2,t} + R_{3,3,t}$ as before and further:

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

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

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

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

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

where the first term is a null sequence as implied by (23). The second term is of the order $O(n^{\delta-1}m) = o(1)$ which follows from analogous arguments as used for $R_{3,2,t}$.

Combining our above arguments, we conclude $\mathbb{P}(\max_{1\leq t\leq T}(\hat{\beta}_t - \beta_t)^2 > c) = o(1)$ as claimed in the lemma.

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

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

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

Ad $R_{8,1}$:

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

due to Assumptions 2–5.

Ad $R_{8,2}$:

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

which follows from our above observations.

Ad $R_{8,3}$:

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

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

as a consequence of (23).

Ad $R_{8,4}$:

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

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

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

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

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

Ad $R_{9,1}$:

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

Ad $R_{9,2}$:

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

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

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

by (21) and the above results. Further

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

by (20) and the above results on $\hat{\alpha}_t$.

Ad $R_{9,3}$:

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

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

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

by (19) and the above results. Further it holds that

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

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

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

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

with

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

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

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

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

where

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

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

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

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

which follows from our above arguments. Further we conclude

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

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

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

where

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

and

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

Note for $R_{14,1}$:

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

with

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

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

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

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

Ad $R_{15,1}$:

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

Ad $R_{15,2}$:

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

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

Ad $R_{15,3}$:

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

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

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

which follows from (23) and our above arguments.

Combining arguments implies the last statement in Lemma 4.1. ■

A.3 Proof of Theorem 4.2

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

$$\begin{aligned}\mathbb{P}\left(\{(1),\ldots,(\hat{\kappa})\} \neq \{([1],\ldots,[\kappa]\}\right) \leq& \mathbb{P}\left(\{(1),\ldots,(\kappa)\} \neq \{[1],\ldots,[\kappa]\}\right) + \mathbb{P}(\hat{\kappa}\neq\kappa)\\ =& o(1) + o(1).\end{aligned} \tag{26}$$

In order to prove that the first probability on the right hand side of (26) is a null sequence, suppose that $t \in G_k$, with $1 \leq k \leq K$. As indicated, there are $\kappa \geq 1$ indexes in $S^{(j)}$ being elements of G_k. For the corresponding distances it holds that $\Delta_{t[1]} = \cdots = \Delta_{t[\kappa]} = 0$ by definition. The remaining distances are bounded away from zero by $0 < C_\Delta \leq \Delta_{t[\kappa+1]} \leq \cdots \leq \Delta_{t[|S^{(j)}|]}$ due to Assumption 7.

As stated in Lemma 4.1, $\max_{1\leq t\leq T} ||\hat{\alpha}_t^{(\Delta)} - \alpha_t^{(\Delta)}||_2^2 = o_p(1)$ implying that $\max_{1\leq s\leq T} |\hat{\Delta}_{ts} - \Delta_{ts}| = o_p(1)$, which holds for any reference period t. Combining arguments allows to conclude $\max_{1\leq s\leq \kappa} \hat{\Delta}_{t(s)} = o_p(1)$ and $\min_{\kappa<s\leq|S^{(j)}|} \hat{\Delta}_{t(s)} \geq C_\Delta + o_p(1)$ as well as $\max_{1\leq s\leq \kappa} \hat{\Delta}_{t[s]} = o_p(1)$ and $\min_{\kappa<s\leq|S^{(j)}|} \hat{\Delta}_{t[s]} \geq C_\Delta + o_p(1)$. This implies that the first probability on the right hand side of (26) tends to zero. Further note that the specification of the threshold in Assumption 7 immediately implies $\mathbb{P}\left(\hat{\Delta}_{t[\kappa]} < \tau_{nT}\right) \to 1$ and $\mathbb{P}\left(\hat{\Delta}_{t[\kappa+1]} > \tau_{nT}\right) \to 1$ as $n \to \infty$ in light of the preceding arguments. As a consequence of this $\mathbb{P}\left(\hat{\Delta}_{t[\kappa]} < \tau_{nT} < \hat{\Delta}_{t[\kappa+1]}\right) \to 1$ as $n \to \infty$, implying that the second probability on the right hand side of (26) is a null sequence. ■

Remark 1

For the calculation of the convergence rate of our estimator $\tilde{A}_k$, the classification error is negligible as a consequence of Theorem 4.2. To see this note that an analogous argument as in Vogt and Linton (2017) holds in our context: let $s(n,T)$ be an arbitrary deterministic sequence such that $s(n,T) \to 0$ as $n, T \to \infty$. Now, note that for any constant $C > 0$

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

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

In light of this remark, the proof of Theorem 4.3 starts from the ideal *oracle* estimators $\tilde{A}_k^*$ rather than their contaminated counterparts.

Remark 2

For the proof of Theorem 4.3, we work with classification-error-free oracle variants of the estimators $\tilde{\phi}_{j,k}, \tilde{\lambda}_{j,k}, \tilde{K}_{X,k}, \tilde{K}_{zX,k}, \tilde{K}_{z,k}$ and $\tilde{\Gamma}_k$. Such estimators, calculated from $\{(z_{it}, X_{it}) : 1 \leq i \leq n,\ t \in G_k\}$, are denoted $\tilde{\phi}_{j,k}^*, \tilde{\lambda}_{j,k}^*, \tilde{K}_{X,k}^*, \tilde{K}_{zX,k}^*, \tilde{K}_{z,k}^*$ and $\tilde{\Gamma}_k^*$. In analogy to before, we further denote the Hilbert Schmidt norm of the difference between $\tilde{\Gamma}_k^*$ and the population counterpart Γ_k as $\widetilde{\mathcal{D}}_k^* := ||\tilde{\Gamma}_k^* - \Gamma_k||_{\mathcal{S}}$. Beyond these quantities, the estimator $\tilde{A}_k^*$ makes implicitly use of the operator $\tilde{\Phi}_k^*$ which estimates, in analogy to $\hat{\Phi}_t$, the operator Φ_k as in (11). $\tilde{\Phi}_k^*$ is defined according to

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

for any $g \in L^2([0,1])$, where $\tilde{m} \equiv \tilde{m}_k$ for simplicity of notation.

Assessing the asymptotic properties of the classification-error-free estimators, note that due to Assumption 1 for every regime G_k, the random variables $\{X_{it} : 1 \le i \le n, t \in G_k\}$ are L^4_m-approximable. Thus, for suitably large constants, the following inequalities from Hörmann and Kokoszka (2010) hold:[1]

$$E\left[\left(\widetilde{\mathcal{D}}_k^*\right)^2\right] \le C(n|G_k|)^{-1} \tag{27}$$

$$E\left[\left|\left|\tilde{K}^*_{X,k} - K_{X,k}\right|\right|_2^2\right] \le C(n|G_k|)^{-1} \tag{28}$$

$$E\left[\left|\tilde{\lambda}^*_{j,k} - \lambda_{j,k}\right|^2\right] \le E\left[\left(\widetilde{\mathcal{D}}_k^*\right)^2\right] \le C(n|G_k|)^{-1} \tag{29}$$

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

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

and further

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

which can be shown by straightforward moment calculations. In addition to that, bounds on $\left|\left|\tilde{\phi}^*_{j,k} - \phi_{j,k}\right|\right|_2^2$ can be obtained in analogy to the almost sure bound in (12) and the asymptotic bound as in (13)–(15). We make the latter precise defining the analogues to $\mathcal{F}_{1,n,t}$–$\mathcal{F}_{3,n,t}$ as

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

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

$$||\tilde{\phi}^*_{j,k} - \phi_{j,k}||_2^2 \le 8\left(1 - 4C(n|G_k|)^{\frac{2(1+\mu)}{\mu+2\nu}}\left(\widetilde{\mathcal{D}}_k^*\right)^2\right)^{-1}\tilde{R}^{(\phi)}_{j,k}, \tag{30}$$

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

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

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

$$||\tilde{\phi}^*_{j,k} - \phi_{j,k}||_2^2 \leq 16\tilde{R}^{(\phi)}_{j,k}. \tag{31}$$

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

As a further important observation we note that

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

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

$$z_{it} - E[z_{it}] = \langle \zeta, X_{it} - E[X_{it}]\rangle + s_{it} \tag{32}$$

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

$$||\hat{\zeta} - \zeta||_2^2 = \left|\left|\tilde{\Phi}^*_k - \Phi_k\right|\right|^2_{H'}. \tag{33}$$

The argumentation in Hall and Horowitz (2007) (cf. their Theorem 1 and corresponding proof) transfers mutatis mutandis to a setup with weakly dependent regressors (L^4_m dependence) and weakly dependent errors (m-dependence) as is the case in our auxiliary regression (32). This can be shown using the fundamental results formulated in Hörmann and Kokoszka (2010). As $\hat{\zeta}$ is calculated from a sample of size $n|G_k|$, the results in Hall and Horowitz (2007) together with (33) thus imply

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

as claimed before.

A.4 Proof of Theorem 4.3

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

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

In the proof of Lemma 4.1 it was shown that $\mathbb{P}\left(Q_{n,t}^c\right) = o(|G_k|^{-1})$. Regarding the remaining term, note that due to the exogeneity of the regressors and stationary distributions within the regimes the following holds: for any $c_j > 0$, $j = 1, 2, 3$, there exist constants $C_j = C_j(c_j)$, $j = 1, 2, 3$ such that

$$\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k} R_{0,1,t}^2 > c_1\right) \leq \frac{nE[R_{0,1,t}^2]}{c_1} \leq C_1$$

$$\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k} R_{0,2,t}^2 > c_2\right) \leq \frac{nE[R_{0,2,t}^2]}{c_2} \leq C_2$$

$$\mathbb{P}\left(n|G_k|^{-1}\sum_{t\in G_k} R_{1,1,t}^2 > c_3\right) \leq \frac{nCE[||\hat{K}_{zX,t} - K_{zX}||_2^2]}{c_3} \leq C_3.$$

Further we observe that $R_{2,1} = o(n^{-1/2})$. Using that $\mathbb{P}(\bigcup_{t\in G_k} \mathcal{F}_{3,n,t}^c) \leq \sum_{t=G_k} \mathbb{P}(\mathcal{F}_{3,n,t}^c) = o(1)$ as shown in the proof of Lemma 4.1, we argue that for any constant c_4, there exists a constant C_4, such that

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

due to the stationarity of (X_{it}, z_{it}), $t \in G_k$ and (12). Beyond that we argue for $R_{3,2,t}$ that for any $c_5 > 0$ it follows from similar arguments that there exists a $C_5 > 0$ such that

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

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

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

thanks to Assumptions 2–5, stationarity and once more the bound in (12). Combining arguments allows us to conclude that $|G_k|^{-1} \sum_{t\in G_k} (\hat{\beta}_t - \beta_t)^2 = O_p(n^{-1})$ as $n, T \to \infty$. We use this finding in a moment to obtain the convergence rate for $\tilde{A}_k$. To assess the underlying problem, we use the following notation:

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

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

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

where

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

and

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

The upper bound

$$\begin{aligned}
||\tilde{A}_k^* - A_k||_2^2 &= \left|\left| \sum_{j=1}^{\tilde{m}} \left(\tilde{a}_{j,k}^{(1)} + \tilde{a}_{j,k}^{(2)} \right) \tilde{\phi}_{j,k}^* - A_k \right|\right|_2^2 && (34) \\
&\leq 2 \left|\left| \sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}^{(1)} \tilde{\phi}_{j,k}^* - A_k \right|\right|_2^2 + 2 \sum_{j=1}^{\tilde{m}} \left(\tilde{a}_{j,k}^{(2)} \right)^2 && (35)
\end{aligned}$$

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 second remark and Assumptions 1-5, it holds that $\left|\left|\sum_{j=1}^{\tilde{m}} \tilde{a}_{j,k}^{(1)} \tilde{\phi}_{j,k}^{*} - A_k\right|\right|_2^2 = O_p\left(n^{\frac{(1+\delta)(1-2\nu)}{\mu+2\nu}}\right)$. The remaining term in (35) we split according to

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

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

Ad $R_{16,1}$:

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

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

Ad $R_{16,2}$:

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

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

Ad $R_{16,3}$:

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

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

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

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

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

as well as

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

Further, we use similar arguments as before (see the proof of Theorem 4.1) to obtain

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

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

A.5 Threshold Choice

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

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

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

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

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

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

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

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

where $F^{-1}_{\underline{m}}$ denotes the quantile function of the $\chi^2_{\underline{m}}$-distribution. As we consider a context where $|G_k|$ is large ($|G_k| \propto T$ in Assumption A3), we expect the value of $\varepsilon/|G_k|$ to be very close to zero. This motivates setting $\tau_{nT} = (2/n)F^{-1}_{\underline{m}}(p_\tau)$, for some p_τ very close to one as mentioned in Section 5. Note that according to Theorem A in Inglot (2010) and our assumptions in Section 4

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

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

References

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

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

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

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

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

Kneip, A., D. Poß, and P. Sarda (2016). Functional linear regression with points of impact. *The Annals of Statistics 44*(1), 1–30.

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

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