ASYMPTOTIC LAWS FOR CHANGE POINT ESTIMATION IN INVERSE REGRESSION

Sophie Frick[1], Thorsten Hohage[2], Axel Munk[3]

[1] *Institute for Mathematical Stochastics,*
[2] *Institute for Numerical and Applied Mathematics,*
Georgia Augusta Universität Göttingen

Supplementary Material

In this supplement the proofs of the results of the paper are collected.

S1 Proofs of Theorem 3.1, Proposition 3.3, and Corollaries 3.2 and 3.4

This section is dedicated to the proof of Theorem 3.1 and Corollaries 3.2 and 3.4 in Section 3. The proof is separated in four parts. We start with some technical results, then we calculate the entropy numbers of the considered function spaces, which yield the basic arguments to show consistency of the estimator in (3.1). Finally, we give the proof of the asymptotic results given in Theorem 3.1 and Corollaries 3.2 and 3.4. This section is dedicated to the proof of Theorem 3.1 and Corollaries 3.2 and 3.4 in Section 3. The proof is separated in four parts. We start with some technical results, then we calculate the entropy numbers of the considered function spaces, which yield the basic arguments to show consistency of the estimator in (3.1). Finally, we give the proof of the asymptotic results given in Theorem 3.1 and Corollaries 3.2 and 3.4.

S1.1 Some technical lemmata

Before we give the proofs of the main results in Section 3, we need some technical lemmata.

Lemma S1.1. *For* $p \in [1, \infty)$ *the mapping* $\mathfrak{F} : (\Theta_k, |\cdot|_\infty) \to (\mathbf{F}_k, \|\cdot\|_{L^p})$, $\theta \mapsto f(\cdot, \theta)$ *is Hölder continuous with index* $1/p$, *and there exists a constant* L *independent of* k *such* *that for all $\theta^{(1)}, \theta^{(2)} \in \Theta$ the following estimate holds true:*

$$\left\|\mathfrak{F}(\theta^{(2)}) - \mathfrak{F}(\theta^{(1)})\right\|_{L^p} \leq \left(L(k+1)\left|\theta^{(2)} - \theta^{(1)}\right|_\infty\right)^{1/p}.$$

Proof. Let $\theta^{(1)} = (\vartheta_1^{(1)}, \tau_1^{(1)}, \dots, \vartheta_k^{(1)}), \theta^{(2)} = (\vartheta_1^{(2)}, \tau_1^{(2)}, \dots, \vartheta_k^{(2)}) \in \Theta_k$ and consider the mapping $\lambda : [0,1] \to L^1([a,b])$,

$$\lambda(t) := f(\cdot, \theta^{(1)} + t(\theta^{(2)} - \theta^{(1)})).$$

Setting $\vartheta_i(t) := \vartheta_i^{(1)} + t(\vartheta_i^{(2)} - \vartheta_i^{(1)})$ for $i = 1, \dots, k+1$ and and $\tau_i(t) := \tau_i^{(1)} + t(\tau_1^{(2)} - \tau_i^{(1)})$ for $i = 0, \dots, k+1$ we get from the integral form of the mean value theorem that

$$\begin{aligned}\left\|\mathfrak{F}(\theta^{(2)}) - \mathfrak{F}(\theta^{(1)})\right\|_{L^p}^p = \|\lambda(1) - \lambda(0)\|_{L^p}^p &\leq \sum_{i=1}^{k+1} \int_{\tau_{i-1}(t)}^{\tau_i(t)} \int_0^1 \left|\left(\vartheta_i^{(2)} - \vartheta_i^{(1)}\right)^\top \frac{\partial \mathfrak{f}}{\partial \vartheta}(y, \vartheta_i(t))\right|^p \,\mathrm{d}t\mathrm{d}y \\ &+ \sum_{i=1}^{k} \int_0^1 \left|[f(\tau_i(t), \theta(t))]\right|^p dt \left|\tau_i^{(2)} - \tau_i^{(1)}\right| \\ &\leq (1+k)\, C \left|\theta^{(2)} - \theta^{(1)}\right|_\infty\end{aligned}$$

with the constant $C := \sup_{\vartheta \in \Psi, y \in [a,b]} \max((2|\mathfrak{f}(y,\vartheta)|)^p, (b-a)\mathrm{diam}_\infty(\Psi)^{p-1}|\frac{\partial \mathfrak{f}}{\partial \vartheta}(y,\vartheta)|_1^p)$, which is finite since Ψ is compact. □

Lemma S1.2. *Suppose that Assumption **C** holds true. Then $\Lambda : \Theta_k \to L^2(I)$ is continuously differentiable and the derivative is given by*

$$(\Lambda'[\theta]e_i)(x) = \begin{cases} \int_a^b \varphi(x,y) \frac{\partial}{\partial \theta_i} f(y,\theta) \mathrm{d}y & i \neq 0 \bmod (r+1), \\ \varphi(x, \tau_{\frac{i}{r+1}}) \, [f(\cdot,\theta)] \, (\tau_{\frac{i}{r+1}}) & i = 0 \bmod (r+1). \end{cases} \tag{S1.1}$$

Proof. We show that the mapping $\Lambda_0 : \Psi \times [a,b]^2 \to L^2(I)$

$$\Lambda_0(\vartheta, \tau_1, \tau_2) := \int_{\tau_1}^{\tau_2} \varphi(\cdot), y) \mathfrak{f}(y, \vartheta) \, \mathrm{d}y$$

is continuously differentiable with derivative

$$\begin{aligned}\Lambda_0'[\vartheta, \tau_1, \tau_2](\delta\vartheta, \delta\tau_1, \delta\tau_2) = &\int_{\tau_1}^{\tau_2} \varphi(\cdot), y) \frac{\partial \mathfrak{f}}{\partial \vartheta}(y, \vartheta) \delta\vartheta \, \mathrm{d}y \\ &- \varphi(\cdot, \tau_1) \mathfrak{f}(\vartheta)(\tau_1, \vartheta) \, \delta\tau_1 + \varphi(\cdot, \tau_2) \mathfrak{f}(\vartheta)(\tau_2, \vartheta) \, \delta\tau_2\end{aligned} \tag{S1.2}$$

from which the assertion follows immediately. We write $\Lambda_0 = \Phi_0 \circ \mathfrak{F}$ as the composition of the mappinng $\mathfrak{F} : \Psi \times [a,b]^2 \to C([a,b]) \times [a,b]^2$, $\mathfrak{F}(\vartheta, \tau_1, \tau_2) := (\mathfrak{f}(\cdot, \vartheta), \tau_1, \tau_2)^\top$, which is continuously differentiable with derivate $\mathfrak{F}'[\vartheta, \tau_1, \tau_2](\delta\vartheta, \delta\tau_1, \delta\tau_2) = (\frac{\partial \mathfrak{f}}{\partial \vartheta}(\cdot, \vartheta)\delta\vartheta, \delta\tau_1, \delta\tau_2)^\top$ by the first property in Definition 2.1, and the integral operator $\Phi_0 : C([a,b]) \times [a,b]^2 \to$

$L^2(I)$, $(\Phi_0(g,\tau_1,\tau_2))(x) := \int_{\tau_1}^{\tau_2} \varphi(x,y)g(y)\,\mathrm{d}y$ which is continuously differentiable with derivative $\Phi_0'[g,\tau_1,\tau_2](\delta g,\delta\tau_1,\delta\tau_2) = \Phi_0(\delta g,\tau_1,\tau_2) - \varphi(\cdot,\tau_1)g(\tau_1)\delta\tau_1 + \varphi(\cdot,\tau_2)g(\tau_2)\delta\tau_2$ by the fundamental theorem of calculus and Assumption **C**. Now (S1.2) follows from the chain rule $\Lambda_0'[\bar\theta](\delta\bar\theta) = \Phi_0'[\mathfrak{F}(\bar\theta)]\mathfrak{F}'[\bar\theta](\delta\bar\theta)$ with $\bar\theta := (\vartheta,\tau_1,\tau_2)$. □

Corollary S1.3. *Suppose that Assumptions **B** and **C** are met. Then, uniformly for all $f \in F_k([a,b])$, it holds*

$$o_P(1) + s_l\|\Phi f\|_n^2 \le \|\Phi f\|^2_{L^2([a,b])} \le s_u\|\Phi f\|_n^2 + o_P(1)$$

*with constants s_l, s_u depending on the design density (cf. Assumption **B**), only.*

Proof. The claim follows from Boysen, Bruns, and Munk (2009, Lemma 4.3) together with Assumption **C**i). □

S1.2 Entropy results

In order to show consistency of the least squares estimator $\hat f_n$ in (3.1), we apply uniform deviation inequalities from empirical process theory. To this end, it is necessary to calculate the *entropy* of the space of interest, which is defined in the following way.

Definition S1.4. *Given a subset $\mathcal{G}$ of a linear space G, a semi-norm $\|\cdot\| : G \to [0,\infty)$, and a real number $\delta > 0$, the δ-**covering number** $\boldsymbol{N}(\delta,\mathcal{G},\|\cdot\|)$ is defined as the smallest value of N such that there are functions $g_1,\dots,g_N$ with*

$$\min_{1\le j\le N}\|g - g_j\| \le \delta \quad \text{for all } g \in \mathcal{G}.$$

*Moreover, the δ-**entropy** $\boldsymbol{H}$ and the **entropy integral** $\boldsymbol{J}$ of $\mathcal{G}$ are defined as*

$$\boldsymbol{H}(\delta,\mathcal{G},\|\cdot\|) = \log \boldsymbol{N}(\delta,\mathcal{G},\|\cdot\|) \text{ and}$$

$$\boldsymbol{J}(\delta,\mathcal{G},\|\cdot\|) := \max\left(\delta, \int_0^\delta \boldsymbol{H}^{1/2}(u,\mathcal{G},\|\cdot\|)\mathrm{d}u\right),$$

respectively.

We are interested in the entropy of the set

$$\mathbf{G}_k := \{\Phi f \in L^2(I) \mid f \in \boldsymbol{F}_k[a,b]\}, \tag{S1.3}$$

where Φ is a known integral operator with kernel φ as defined in (1.2). In order to deduce consistency of $\hat f_n$, additionally we have to know the entropy of the set $\boldsymbol{F}_k$. By definition, all functions $f \in \boldsymbol{F}_k$ are determined by a parameter vector θ. Thus the core of the problem reduces to determination of the entropy of the parameter set Θ_k.

Lemma S1.5. *Let $\boldsymbol{F}_k$ and $d = (k+1)r + k$ be as in Definition 2.4. Then there exists a constant $T_{\mathcal{F}}, \tilde{T}_{\mathcal{F}} > 0$ depending only on the considered function class $\mathcal{F}$ in Definition 2.1, such that*

$$\boldsymbol{H}(\delta, \boldsymbol{F}_k, \|\cdot\|_{L^1}) \le d \log\left(\frac{(k+1)T_{\mathcal{F}} + \delta}{\delta}\right), \tag{S1.4}$$

$$\boldsymbol{H}(\delta, \boldsymbol{G}_k, \|\cdot\|_n) \le d \log\left(\frac{(k+1)\tilde{T}_{\mathcal{F}} + \delta}{\delta}\right). \tag{S1.5}$$

Proof. Note that the diameter of Θ_k with respect to the maximum norm is bounded by a constant M independent of k and recall that $\mathfrak{F} : (\Theta_k, |\cdot|_\infty) \to (\mathbf{F}_k, \|\cdot\|_{L^1})$, $\theta \mapsto f(\cdot, \theta)$ is Lipschitz continuous with constant $L(k+1)$ (cf. Lemma S1.1). Hence, (S1.4) with $T_{\mathcal{F}} := 2ML$ follows from the fact that the number of balls with radius $\delta/(L(k+1))$ which are needed to cover a subset of $\mathbb{R}^d$ with diameter bounded by M can be estimated by $(2ML(k+1)+\delta)^d/\delta^d$ (cf. del Barrio, Deheuvels, and van de Geer (2007, Lem. 2.5)). Analogously, we obtain (S1.5) with $\tilde{T}_{\mathcal{F}} := 2ML\|\Phi\|_{L^1 \to L^\infty}$. □

S1.3 Consistency

Theorem S1.6. *Let Φ be an operator satisfying Assumption **C** and $f_0 = f(\cdot, \theta_0) \in F_k$. Furthermore, assume that Assumption **A1** and **B** are met. Then, for $\hat{f}_n = f(\cdot, \hat{\theta}_n)$, the least squares estimator in (3.1), it holds that*

$$\|\Phi \hat{f}_n - \Phi f_0\|_n = o_P(1).$$

Proof. Due to Inequality (3.1) we have

$$\|\Phi \hat{f}_n - Y\|_n^2 \le \|\Phi f_0 - Y\|_n^2 + o_p(n^{-1}).$$

Inserting $Y = \Phi f_0 + \varepsilon$ leads to

$$\|\Phi \hat{f}_n - \Phi f_0\|_n^2 - 2\langle \Phi \hat{f}_n - \Phi f_0, \varepsilon\rangle_n + \|\varepsilon\|_n^2 \le \|\varepsilon\|_n^2 + o_p(n^{-1})$$

which implies

$$\begin{aligned}
\|\Phi \hat{f}_n - \Phi f_0\|_n^2 &\le 2\langle \Phi \hat{f}_n - \Phi f_0, \varepsilon\rangle_n + o_p(n^{-1}) \\
&= 2(\langle \Phi \hat{f}_n, \varepsilon\rangle_n - \langle \Phi f_0, \varepsilon\rangle_n) + o_p(n^{-1}) \\
&\le 4 \sup_{g \in \mathbf{G}_k} |\langle g, \varepsilon\rangle_n| + o_p(n^{-1}).
\end{aligned}$$

Lemma S1.5 gives boundedness of the entropy $\boldsymbol{H}(\delta, \mathbf{G}_k, P_n)$ uniformly in n, for all $\delta > 0$ and so $n^{-1}\boldsymbol{H}(\delta, \mathbf{G}_k, P_n) \to 0$ as $n \to \infty$. With this result it follows directly from van de Geer (2000, Theorem 4.8) that $\sup_{g \in \mathbf{G}_k} |\langle g, \varepsilon\rangle_n| = o_P(1)$. □

Corollary S1.7. *Under the assumptions of Theorem S1.6 one has*

$$\|\Phi \hat{f}_n - \Phi f_0\|_{L^2(I)} = o_P(1).$$

Proof. Since the design in Definition (1.1) is assumed to satisfy Assumption **B** the claim follows directly from Theorem S1.6 and Corollary S1.3. □

Lemma S1.8. *Under the assumptions of Theorem S1.6 it holds that*

$$\|\Phi \hat{f}_n - \Phi f_0\|_{L^2(I)} = o_p(1) \quad \textit{implies} \quad \|f(\cdot,\theta_0) - f(\cdot,\hat{\theta}_n)\|_{L^2([a,b])} = o_p(1).$$

Proof. The operator $\Phi : (\mathbf{F}_k, \|\cdot\|_{L^2([a,b])}) \longrightarrow (L^2(I), \|\cdot\|_{L^2(I)})$ is linear and bounded and hence continuous. According to Assumption **C**ii) it is injective and it follows from Lemma S1.5 that the set $(\mathbf{F}_k, \|\cdot\|_{L^2([a,b])})$ is totally bounded. Since it also contains functions with less than k change points, it is additionally closed and therefore compact. Hence $\Phi : \mathbf{F}_k \longrightarrow \{\Phi f \in L^2(I) : f \in \mathbf{F}_k\}$ is a bijective continuous mapping from a compact set to a Hausdorff space, hence a homeomorphism (see tom Dieck (2008, Prop 1.5.3)). □

Lemma S1.9. *Assume that $f_0 = f(\cdot,\theta_0) \in \boldsymbol{F}_k$ with $\sharp\mathcal{J}(f_0) = k$ and let $\{f(\cdot,\theta_n)\}_{n\in\mathbb{N}}$ be a sequence in $\boldsymbol{F}_k$. Then*

$$\|f(\cdot,\theta_0) - f(\cdot,\theta_n)\|_{L^2([a,b])} = o(1) \quad \textit{implies} \quad |\theta_0 - \theta_n|_\infty = o(1).$$

Proof. Due to the definition of $\mathcal{J}(\cdot)$ in Subsection 2.2, the assumption $\sharp\mathcal{J}(f_0) = k$ implies that $f(\cdot,\theta_0)$ has precisely k change points. That means, $f(\cdot,\theta_0) \equiv f(\cdot,\theta)$ implies $\theta = \theta_0$, i.e. for all $\theta_0 \neq \theta \in \Theta_k$ we have $\|f(\cdot,\theta_0) - f(\cdot,\theta)\|_{L^2([a,b])} > 0$. Now assume that $\|f(\cdot,\theta_0) - f(\cdot,\theta_n)\|_{L^2([a,b])} = o(1)$ but that there exist a subsequence $\{\theta_{k_n}\}_{n\in\mathbb{N}}$ and a constant $c_1 > 0$, such that $|\theta_0 - \theta_{k_n}|_\infty > c_1$ for all $n \in \mathbb{N}$. Since Θ_k is compact, we can choose a further subsequence of this subsequence, which converges to some $\hat{\theta} \in \Theta_k$. W.l.o.g we assume $\lim_{n\to\infty} |\hat{\theta} - \theta_{k_n}|_\infty = 0$. By construction $|\theta_0 - \hat{\theta}|_\infty > c_1$ and so uniqueness of θ_0 implies $\|f(\cdot,\theta_0) - f(\cdot,\hat{\theta})\|_{L^2([a,b])} > c_2 > 0$ for some constant c_2. Since the mapping $\theta \mapsto \|f(\cdot,\theta) - f(\cdot,\theta_0)\|_{L^2([a,b])}$ is continuous by Lemma S1.1, there exists some $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$ we have

$$\|f(\cdot,\theta_0) - f(\cdot,\theta_{k_n})\|_{L^2([a,b])} > \frac{1}{2}c_2 > 0.$$

This is a contradiction to $\|f(\cdot,\theta_0) - f(\cdot,\theta_n)\|_{L^2([a,b])} = o(1)$ and the claim follows. □

Corollary S1.10. *Under the assumptions of Theorem S1.6 it holds that*

$$\|f(\cdot,\theta_0) - f(\cdot,\hat{\theta}_n)\|_{L^2([a,b])} = o_P(1).$$

Moreover, if the true function f_0 has exactly k change points it also holds that

$$|\theta_0 - \hat{\theta}_n|_\infty = o_P(1).$$

Proof. This follows from Theorem S1.6 by application of Lemma S1.8 and S1.9. □

S1.4 Asymptotic normality

In this subsection we show asymptotic normality of the least squares estimator $\hat{\theta}_n$ in (3.1). Therefore, we focus on the stochastic process $\|Y - \Phi\hat{f}_n\|_n^2 = n^{-1}\sum_{i=1}^n (y_i - \Phi\hat{f}_n(x_i))^2$ for the random observations (Y, X) as in (1.1), which henceforth we write as the empirical expectation

$$\mathbb{E}_n m(\cdot,\cdot,\theta) := n^{-1}\sum_{i=1}^n m(x_i, y_i, \theta),$$

with m defined as

$$m(x, y, \theta) := (y - (\Lambda(\theta))(x))^2. \tag{S1.6}$$

Hence, $\hat{\theta}_n$ the least squares estimator is the minimizer of $\theta \longmapsto \mathbb{E}_n m(\cdot,\cdot,\theta)$. Let $\mathbf{E}\varepsilon_1 = 0$ and $\mathbf{E}\varepsilon_1^2 = \sigma^2$ then expectation of $m(\cdot,\cdot,\theta)$ can be calculated as

$$\begin{aligned}\mathbf{E}m(\cdot,\cdot,\theta) &= \mathbf{E}(\Phi f(\cdot,\theta_0) - \Phi f(\cdot,\theta))^2 + \sigma^2 \\ &= \mathbf{E}(\Phi f(\cdot,\theta_0) - \Phi f(\cdot,\theta))^2 + \mathbf{E}m(\cdot,\cdot,\theta_0).\end{aligned} \tag{S1.7}$$

By Lemma S1.2, the function $\theta \mapsto m(\cdot, y, \theta)$ is differentiable with derivative $\partial/\partial\theta\, m(\cdot, y, \theta) = 2(\Lambda(\theta) - y)\Lambda'[\theta]$ such that for all $h_1, h_2 \in \mathbb{R}^d$

$$\mathbf{E}\left(\frac{\partial m}{\partial\theta}(\cdot,\cdot,\theta_0)h_1\right)\left(\frac{\partial m}{\partial\theta}(\cdot,\cdot,\theta_0)h_2\right) = 4\sigma^2\mathbf{E}(\Lambda'[\theta_0]h_1)(\Lambda'[\theta_0]h_2) = 4\sigma^2 h_1^\top V_{\theta_0} h_2. \tag{S1.8}$$

Classical conditions for asymptotic normality of $\hat{\theta}_n$ require that the function $\theta \mapsto m(x, y, \theta)$ is twice differentiable, which is not the case on our situation. Therefore, we follow a different route according to Theorem 5.23 (Chapter 5.3) in van der Vaart (1998) where a second order expansion of the expectation $\theta \mapsto \mathbf{E}m(\cdot,\cdot,\theta)$ instead of the function m itself is sufficient to obtain the desired normality.

Theorem S1.11. *For each θ in an open subset of Euclidean space let $(x, y) \mapsto m(x, y, \theta)$ be a measurable function such that $\theta \mapsto m(x, y, \theta)$ is differentiable at θ_0 for $\boldsymbol{P}$-almost every (x, y) and such that, for every θ_1 and θ_2 in a neighborhood of θ_0 and a measurable function $\dot{m}$ with $\boldsymbol{E}\dot{m}^2 < \infty$*

$$|m(x, y, \theta_1) - m(x, y, \theta_2)| \le \dot{m}(x, y)|\theta_1 - \theta_2|_\infty. \tag{S1.9}$$

Furthermore, assume that the map $\theta \mapsto \boldsymbol{E}m(\cdot,\cdot,\theta)$ has the asymptotic behavior

$$\boldsymbol{E}m(\cdot,\cdot,\theta) = \boldsymbol{E}m(\cdot,\cdot,\theta_0) + \frac{1}{2}(\theta - \theta_0)^\top V(\theta - \theta_0) + o(|\theta_0 - \theta|_\infty^2), \qquad \text{as } |\theta_0 - \theta|_\infty \to 0 \tag{S1.10}$$

at a point of minimum θ_0 with a nonsingular symmetric matrix V. If $\mathbb{E}_n m(\cdot,\cdot,\hat{\theta}_n) \le \inf_\theta \mathbb{E}_n m(\cdot,\cdot,\theta) + o_P(n^{-1})$ and $\hat{\theta}_n \xrightarrow{P} \theta_0$, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V^{-1}\frac{1}{\sqrt{n}}\sum_{1=1}^n \frac{\partial m}{\partial\theta}(x_i, y_i, \theta_0) + o_P(1).$$

In particular, the sequence $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $V^{-1}\boldsymbol{E}\frac{\partial m}{\partial\theta}(\cdot,\cdot,\theta_0)\frac{\partial m}{\partial\theta}(\cdot,\cdot,\theta_0)^\top V^{-1}$.

Proof. Along the lines of the proof of van der Vaart (1998, Thm 2.23). □

Proof. (of Theorem 3.1) We show that the assumptions of Theorem S1.11 are satisfied: It follows from Lemma S1.1 and the assumed boundedness of $\Phi : L^1([a,b]) \to L^\infty(I)$ that $\Lambda = \Phi \cdot \mathfrak{F} : (\Theta, |\cdot|_\infty) \to (L^\infty, \|\cdot\|_{L^\infty})$ is Lipschitz continuous, which implies condition (S1.9) is satisfied with constant $\dot{m}$. Moreover, (S1.10) with $V = V_{\theta_0}$ follows from Lemma S1.2. According to this theorem, together with (S1.8), the sequence $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $\sigma^2 V_{\theta_0}^{-1}$, which proves (i). Part (ii) follows from van der Vaart (1998, Cor. 5.53). Part (iii) is now a consequence of part (ii) and Lemma S1.1. Finally, part (iv) follows from part (iii) with $p = 1$ and the boundedness of $\Phi : L^1([a,b]) \to L^\infty(I)$. □

Proof. (of Corollary 3.2) Due to the differentiability of h in Definition 2.6 Lemma S1.1 and S1.2 hold analogously for the reduced parameter domain by the chain rule. Moreover, the mapping $\delta\tilde{\theta} \mapsto \Lambda'[h(\tilde{\theta}]h'[\tilde{\theta}]\delta\tilde{\theta}$ is injective by Assumption **C** and the injectivity assumption in Definition 2.6, and hence $V_{\tilde{\theta}}$ is nonsingular. Therefore, the proof of the corollary is completely analogous to the proof of Theorem 3.1. □

Proof. (of Corollary 3.4) Statements (i) - (iv) from Theorem 3.1 are valid for the reduced parameter vectors $\tilde{\theta}_0$ and $\tilde{\theta}_n$ by Corollary 3.2. In order to show (3.6), we skip the dependencies of the parameter components, for the sake of simplicity and consider the pieces $\mathfrak{f}(y, \vartheta^i)$ instead of $\mathfrak{f}(y, \vartheta^i(\tilde{\theta}))$ for all $i = 1, ..., k+1$, keeping in mind that for all occurring derivatives we actually need to apply the chain rule.

Now f has a kink in τ_i for all $i = 1, ..., k$. W.l.o.g. we assume that $\tau_i > \hat{\tau}_i$, then we have

$$\int_{\tau_i}^{\hat{\tau}_i} \left(\mathfrak{f}(y, \vartheta^{i+1}) - \mathfrak{f}(y, \hat{\vartheta}^i)\right)^p \mathrm{d}y \leq \int_{\tau_i}^{\hat{\tau}_i} \Big(|\mathfrak{f}(y, \vartheta^{i+1}) - \mathfrak{f}(\tau_i, \vartheta^{i+1})| + |\mathfrak{f}(\tau_i, \vartheta^{i+1}) - \mathfrak{f}(\tau_i, \vartheta^i)| + |\mathfrak{f}(\tau_i, \vartheta^i) - \mathfrak{f}(\tau_i, \hat{\vartheta}^i)| + |\mathfrak{f}(\tau_i, \hat{\vartheta}^i) - \mathfrak{f}(y, \hat{\vartheta}^i)|\Big)^p \mathrm{d}y.$$

By the mean value theorem we have $|\mathfrak{f}(\tau_i, \vartheta^i) - \mathfrak{f}(\tau_i, \hat{\vartheta}^i)| = O(|\vartheta^i - \hat{\vartheta}^i|)$. The term $|\mathfrak{f}(\tau_i, \vartheta^{i+1}) - \mathfrak{f}(\tau_i, \vartheta^i)|$ vanishes because there is a kink at τ_i. Finally, remembering the definition of the modulus of continuity ν in (3.5), we get

$$\sup_{y \in [\tau_i, \hat{\tau}_i]} (|\mathfrak{f}(y, \vartheta^{i+1}) - \mathfrak{f}(\tau_i, \vartheta^{i+1})|, |\mathfrak{f}(\tau_i, \hat{\vartheta}^i) - \mathfrak{f}(y, \hat{\vartheta}^i)|) = \nu(\mathcal{F}, |\tau_i - \hat{\tau}_i|)\,.$$

Hence, it follows from (ii) that

$$\begin{aligned}\int_{\tau_i}^{\hat{\tau}_i} \left(\mathfrak{f}(y, \vartheta^{i+1}) - \mathfrak{f}(y, \hat{\vartheta}^i)\right)^p \mathrm{d}y &= O(|\tau_i - \hat{\tau}_i|)(\nu(\mathcal{F}, |\tau_i - \hat{\tau}_i|) + |\vartheta^i - \hat{\vartheta}^i|)^2 \\ &= O_P(n^{-\frac{1}{2}}(\nu(\mathcal{F}, n^{-\frac{1}{2}})^p + n^{-p/2})).\end{aligned}$$

Since this holds for all $i = 1, ..., k$, this proves (3.6). □

S1.5 Proof of Proposition 3.3

Proof. (Proposition 3.3) It is obvious from the definition (3.4) that the matrix V_θ is symmetric and positive semi-definite, so we have to study under which conditions it is positive definite. Since $h^\top V_\theta h = \int_a^b |\Lambda'[\theta_0]h|^2 s\, dy$ for $h \in \mathbb{R}^d$, it follows from Assumption **B** on s that that $h^\top V_\theta h = 0$ is equivalent to $\Lambda'[\theta]h = 0$. Hence, V_{θ_0} is non-singular if and only if $\Lambda'[\theta]$ is injective. It follows from Lemma S1.2 that $\Lambda'[\theta] = \Phi \circ \mathfrak{F}'[\theta]$ is the composition of the integral operator $\Phi : \mathcal{M} \to L^2(I)$, which is injective by Assumption **C**ii) and the formal derivative $\mathfrak{F}'[\theta_0] : \Theta_k \to \mathcal{M}$, of the mapping $\mathfrak{F}\theta := f(\cdot, \theta)$ given by

$$\mathfrak{F}'[\vartheta_1, \tau_1, \vartheta_2, \ldots, \tau_k, \vartheta_{k+1}](\delta\vartheta_1, \delta\tau_1, \delta\vartheta_2 \ldots, \delta\tau_k, \delta\vartheta_{k+1})$$
$$:= \sum_{j=1}^{k+1} (\delta\vartheta_j)^\top \frac{\partial \mathfrak{f}}{\partial \vartheta_j}(\cdot, \vartheta_j) \mathbf{1}_{[\tau_{j-1}, \tau_j)} + \sum_{j=1}^{k} [f(\cdot, \theta)]\,(\tau_j)\delta\tau_j\,.$$

Since the mappings $\mathfrak{F}'_{[\tau_{j-1},\tau_j)}[\vartheta_j]\delta\vartheta_j = (\delta\vartheta_j)^\top \frac{\partial \mathfrak{f}}{\partial \vartheta_j}(\cdot, \vartheta_j)$ are assumed to be injective (see Definition 2.1) $\mathfrak{F}$ is injective if and only if $[f(\cdot,\theta)](\tau_j) \neq 0$ for $j = 1, \ldots, k$, i.e. if and only if $f(\cdot, \theta)$ has jumps at all change points. □

S2 Proof of Theorem 3.5

From Inequality (3.2) we obtain the basic inequality

$$\|\Phi \hat{f}_{\lambda_n} - \Phi f_0\|_n^2 \leq 2\langle \Phi f_0 - \Phi \hat{f}_{\lambda_n}, \varepsilon\rangle_n + \lambda_n(\sharp J(f_0) - \sharp J(\hat{f}_{\lambda_n})) + o(n^{-1}). \tag{S2.11}$$

Again we have to consider the behavior of the empirical process $\langle \Phi f_0 - \Phi \hat{f}_{\lambda_n}, \varepsilon\rangle_n$, and therefore the entropy of the respective function space to gain a bound for this process. We use the results from Boysen, Bruns, and Munk (2009).

Lemma S2.1. *Suppose that Assumptions A and A1 are satisfied. Then, for all $\Phi f \in \boldsymbol{G}_\infty = \{\Phi f \in L^2(I) \mid f \in \boldsymbol{F}_\infty\}$, we have*

$$|\langle \Phi f, \varepsilon\rangle_n| = O_P(n^{-\frac{1}{2}}) \|\Phi f\|_n^{1-\epsilon} (\sharp J(f))^{\frac{1}{2}(1+2\epsilon)},$$

for any $\epsilon > 0$.

Proof. For fixed number of jumps k, we find from Lemma S1.5 that

$$H(\delta, G_k, P_n) \leq d \log\left(\frac{\tilde{T}_{\mathcal{F}}\sqrt{k+1} + \delta}{\delta}\right),$$

with $d = (k+1)r + k$ and a constant $\tilde{T}_{\mathcal{F}}$, which is independent of k. Using this entropy bound, it follows along the lines of the proof of Lemma 4.18 in Boysen, Bruns, and Munk (2009) that

$$\sup_{g \in G_k,\, \|g\|_n \leq \delta} \frac{|\langle g, \varepsilon\rangle_n|}{\sqrt{k}\delta\left(1 + \log\left(\frac{\tilde{T}_{\mathcal{F}}\sqrt{k}+\delta}{\delta}\right)\right)} = O_P(n^{-\frac{1}{2}}),$$

holds uniformly for all k. For all $\Phi f \in \mathbf{G}_\infty$ this implies

$$\frac{|\langle \Phi f, \varepsilon\rangle_n|}{\sqrt{\sharp J(f)}\|\Phi f\|_n\left(1+\log\left(\frac{\tilde{T}_{\mathcal{F}}\sqrt{\sharp J(f)}+\|\Phi f\|_n}{\|\Phi f\|_n}\right)\right)}$$
$$\leq \sup_{\substack{g\in G_{\sharp J(f)},\\ \|g\|_n\leq\|\Phi f\|_n}} \frac{|\langle g, \varepsilon\rangle_n|}{\sqrt{\sharp J(f)}\|\Phi f\|_n\left(1+\log\left(\frac{\tilde{T}_{\mathcal{F}}\sqrt{\sharp J(f)}+\|\Phi f\|_n}{\|\Phi f\|_n}\right)\right)} = O_P(n^{-\frac{1}{2}}).$$

Analogously to the proof of Corollary 4.19 in Boysen, Bruns, and Munk (2009), this directly yields the claim. □

Lemma S2.2. *Let $f_0 \in \boldsymbol{F}_\infty$ and $\{f_n\}_{n\in\mathbb{N}}$ a sequence in $\boldsymbol{F}_{\sharp J(f_0),D}$, with*

$$\|f_0 - f_n\|_{L^2([a,b])} = o(1).$$

Then, there exists an $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$

$$\sharp J(f_0) = \sharp J(f_n).$$

Proof. W.l.o.g let $\sharp J(f_0) = 1$. Now we assume that there exists a subsequence f_{k_n} with no jumps, i.e. $f_{k_n} \in \boldsymbol{F}_{0,D}$ for all n. Furthermore f_{k_n} is a subsequence of a converging sequence, and thus converges to the same limit function f_0. As shown in the proof of Lemma S1.8, the set $\boldsymbol{F}_{0,D}$ is compact thus the limit function of f_{k_n} has to be contained in $\boldsymbol{F}_{0,D}$, which leads the contradiction

$$f_0 \in \boldsymbol{F}_{0,D}.$$

□

Now we are prepared for the proof of Theorem 3.5.

Proof. (of Theorem 3.5) Throughout the proof w.l.o.g we assume that $\epsilon \leq 1$. From Lemma S2.1 and (S2.11), it follows that

$$\begin{aligned}\|\Phi\hat{f}_{\lambda_n} - \Phi f_0\|_n^2 &\leq O_P(n^{-\frac{1}{2}})\|\Phi\hat{f}_{\lambda_n} - \Phi f_0\|_n^{1-\frac{1}{2}\epsilon}(\sharp\mathcal{J}(\hat{f}_{\lambda_n} - f_0))^{\frac{1}{2}(1+\epsilon)} \\ &\quad +\lambda_n(\sharp\mathcal{J}(f_0) - \sharp\mathcal{J}(\hat{f}_{\lambda_n})) + o(n^{-1}) \\ &\leq O_P(n^{-\frac{1}{2}})\|\Phi\hat{f}_{\lambda_n} - \Phi f_0\|_n^{1-\frac{1}{2}\epsilon}\sharp\mathcal{J}(\hat{f}_{\lambda_n})^{\frac{1}{2}(1+\epsilon)} - \lambda_n\sharp\mathcal{J}(\hat{f}_{\lambda_n}) + \lambda_n\sharp\mathcal{J}(f_0),\end{aligned}$$

where we took into account that λ_n is assumed to converge slower than n^{-1} and that we have $\sharp\mathcal{J}(f_0) < \infty$, which implies that $\sharp\mathcal{J}(\hat{f}_{\lambda_n} - f_0) = O_P(\sharp\mathcal{J}(\hat{f}_{\lambda n}))$.

Choosing $f \equiv 0$ on the right hand side of Equation (3.2) implies $\lambda_n\sharp\mathcal{J}(\hat{f}_{\lambda_n}) \leq \|Y\|_n^2 = O_P(1)$ and hence, we have

$$\sharp\mathcal{J}(\hat{f}_{\lambda_n}) = O_P(\lambda_n^{-1}). \tag{S2.12}$$

We assumed that $\lambda_n^{-1} n^{-1/(1+\epsilon)} \to 0$, for $n \to \infty$, which gives

$$n^{-1} = o(\lambda_n^{1+\epsilon}). \tag{S2.13}$$

By compactness of Ψ we have that $\sup_{f \in \mathbf{F}_\infty} \|f\|_\infty \le R$ and thus

$$\sup_{f \in \mathbf{F}_\infty} \|\Phi f\|_n \le \|\varphi\|_\infty R < \infty \tag{S2.14}$$

Inserting (S2.14), (S2.12) and (S2.13) into (S2.12), we obtain

$$\begin{aligned} \|\Phi \hat{f}_{\lambda_n} - \Phi f_0\|_n^2 &\le o_P(\lambda_n^{\frac{1+\epsilon}{2}}) O_P(\lambda_n^{\frac{1-\epsilon}{2}}) \|\Phi \hat{f}_{\lambda_n} - \Phi f_0\|_n^{1-\frac{1}{2}\epsilon} \sharp\mathcal{J}(\hat{f}_n) - \lambda_n \sharp\mathcal{J}(\hat{f}_{\lambda_n}) + \lambda_n \sharp\mathcal{J}(f_0) \\ &= (o_p(\lambda_n) - \lambda_n) \sharp\mathcal{J}(\hat{f}_{\lambda_n}) + \lambda_n \sharp\mathcal{J}(f_0). \end{aligned} \tag{S2.15}$$

Since $o_p(\lambda_n) - \lambda_n$ becomes negative for increasing n, this implies

$$\|\Phi \hat{f}_{\lambda_n} - \Phi f_0\|_n^2 = O_P(\lambda_n)$$

and with Corollary S1.3,

$$\|\Phi \hat{f}_{\lambda_n} - \Phi f_0\|_{L^2(I)}^2 = O_P(\lambda_n) + o_P(1) = o_P(1). \tag{S2.16}$$

Again considering Equation (S2.15) we find that this is equivalent to

$$0 \le (o_p(\lambda_n) - \lambda_n) \sharp\mathcal{J}(\hat{f}_{\lambda_n}) + \lambda_n \sharp\mathcal{J}(f_0),$$

which means

$$(1 - o_P(1)) \sharp\mathcal{J}(\hat{f}_{\lambda_n}) \le \sharp\mathcal{J}(f_0).$$

Because $\sharp\mathcal{J}(f_0)$ and $\sharp\mathcal{J}(\hat{f}_{\lambda_n})$ are integers, this implies $P(\sharp\mathcal{J}(\hat{f}_{\lambda_n}) \le \sharp\mathcal{J}(f_0)) \to 1$. For $\sharp\mathcal{J}(\hat{f}_{\lambda_n}) \le \sharp\mathcal{J}(f_0)$ in turn, it holds that $f_0, \hat{f}_{\lambda_n} \in \boldsymbol{F}_{\sharp\mathcal{J}(f_0)}$ and Lemma S1.8 together with (S2.16), yields

$$\|f_0 - \hat{f}_{\lambda_n}\|_{L^2([a,b])} = o_P(1).$$

Using Lemma S2.2 this implies that $\lim_{n\to\infty} P\left(\sharp\mathcal{J}(f_0) = \sharp\mathcal{J}(\hat{f}_{\lambda_n})\right) = 1$, which is the claim. □

S3 Proofs of the Theorems 4.2, 4.3 and 4.4

Proof. (Theorem 4.2) ad **C**i): It is straightforward to show that $\|\Phi f\|_{L^\infty} \le \|\phi\|_{L^\infty} \|f\|_{L^1}$, so $\phi \in \mathcal{L}\left(L^1([a,b]), L^\infty([c,d])\right)$. Since $\phi \in BV([ac, bd])$ there exist monotonically increasing and bounded functions ϕ_1, ϕ_2 such that $\phi = \phi_1 - \phi_2$. Setting $\varphi_i(x,y) := \phi_i(xy)$ for $i = 1, 2$ we obtain for $x, x+\delta \in [a,b]$ with $\delta > 0$

$$\begin{aligned} |(\Phi f)(x) - (\Phi f)(x+\delta)| &= \left| \int_a^b (\varphi_1(x,y) - \varphi_1(x+\delta,y) - \varphi_2(x,y) + \varphi_2(x+\delta,y)) f(y) \mathrm{d}y \right| \\ &\le \|f\|_\infty \left[\int_a^b |\varphi_1(x+\delta,y) - \varphi_1(x,y)| \, \mathrm{d}y + \int_a^b |\varphi_2(x+\delta,y) - \varphi_2(x+\delta,y)| \, \mathrm{d}y \right] \quad \text{(S3.17)} \\ &= \|f\|_\infty \left[\int_a^b (\varphi_1(x+\delta,y) - \varphi_1(x,y)) \, \mathrm{d}y + \int_a^b (\varphi_2(x+\delta,y) - \varphi_2(\delta,y)) \, \mathrm{d}y \right] \end{aligned}$$

using the monotonicity of ϕ_i in the last line. The integrals on the left hand side can be estimated by

$$\begin{aligned}\int_a^b [\phi_i((x+\delta)y) - \phi_i(xy)]\,\mathrm{d}y &= \frac{1}{x+\delta}\int_{a(x+\delta)}^{b(x+\delta)} \phi_i(u)\mathrm{d}u - \frac{1}{x}\int_{ax}^{bx}\phi_i(u)\mathrm{d}u \\ &= \left(\frac{1}{x+\delta} - \frac{1}{x}\right)\int_{ax}^{bx}\phi_i(u)\mathrm{d}u + \frac{1}{x+\delta}\left(\int_{bx}^{b(x+\delta)}\phi_i(u)\mathrm{d}u - \int_{ax}^{a(x+\delta)}\phi_i(u)\mathrm{d}u\right) \\ &\le \left(\frac{b-a}{x+\delta}\delta + \frac{\delta b - \delta a}{x+\delta}\right)\|\phi_i\|_\infty \le \frac{b-a}{a}2\delta\|\phi_i\|_\infty\,,\end{aligned}$$

so $|(\Phi f)(x) - (\Phi f)(x+\delta)| \le \frac{b-a}{a}2\delta(\|\phi_1\|_\infty + \|\phi_2\|_\infty)\,\|f\|_\infty$.

ad **C**ii): Assume that the Müntz condition (4.1) holds true and that

$$(\Phi\mu)|_{\left[\frac{\rho_1}{a},\frac{\rho_2}{b}\right]} \equiv 0$$

for some Borel measure $\mu \in \mathcal{B}([a,b])$. Since $xy \in [\rho_1,\rho_2]$ if $x \in [\frac{\rho_1}{a},\frac{\rho_2}{b}]$ and $y \in [a,b]$ and since the series expansion of ϕ converges absolutely and hence uniformly on the compact interval $[\rho_1,\rho_2]$, integration and summation may be interchanged, and we obtain

$$(\Phi\mu)(x) = \int_a^b \phi(xy)\mathrm{d}\mu(y) = \sum_{j=0}^\infty x^j \int_a^b \alpha_j y^j \mathrm{d}\mu(y) = \sum_{j=0}^\infty c_j x^j, \qquad x \in \left[\frac{\rho_1}{a},\frac{\rho_2}{b}\right]$$

with $c_j := \alpha_j \int_a^b y^j \mathrm{d}\mu(y)$. In order to see that the power series $\sum_{j=0}^\infty c_j x^j$ converges absolutely and uniformly for $x \in [\frac{\rho_1}{a},\frac{\rho_2}{b}]$, note that $|c_j| \le |\mu|([a,b])|\alpha_j|b^j$, so

$$\sum_{j=0}^\infty \left|c_j x^j\right| \le |\mu|([a,b])\sum_{j=0}^\infty |\alpha_j|\rho_2^j < \infty, \qquad x \in \left[\frac{\rho_1}{a},\frac{\rho_2}{b}\right].$$

Since a power series with positive radius of convergence vanishes identically if and only if all its coefficients vanish, we obtain

$$\int_a^b y^j \mathrm{d}\mu(y) = 0 \qquad \text{for all } j \in J.$$

By the *Müntz-Theorem* 4.1 this implies that $\int_a^b g(y)\mathrm{d}\mu(y) = 0$ for all $g \in C([a,b])$, so $\mu \equiv 0$, i.e. $\Phi : \mathcal{B}([a,b]) \to L^2([a,b])$ is injective.

If the Müntz condition (4.1) is violated, then the converse implication of the full Müntz Theorem 4.1 entails that the closure of $\mathrm{span}(\{y^j : j \in J\})$ does not coincide with $C([a,b])$, and as a consequence of the Hahn-Banach theorem (cf. Yosida (1995, §IV.6)) there exists a functional $\overline{\mu}_0 \neq 0$ in the dual space $C([a,b])'$ which vanishes on $\mathrm{span}(\{y^j : j \in J\})$. By the Riesz representation theorem (cf. Rudin (1987, Thm 6.19)) $\overline{\mu}_0$ can be expressed by a (signed) Borel measure $\mu_0 \in \mathcal{B}([a,b])$ via $\overline{\mu}_0(g) = \int_a^b g d\mu_0$, and our previous computations show that $\Phi\mu_0 = 0$. □

Proof. (Theorem 4.3) ad **C**i): Obviously, $\|\Phi f\|_{L^\infty} \leq \|\phi\|_{L^\infty}\|f\|_{L^1}$, so $\phi \in \mathcal{L}\left(L^1([a,b]), L^\infty([a,b])\right)$. As in the proof of Theorem 4.2, we can write $\phi = \phi_1 - \phi_2$ with bounded, monotonically increasing functions ϕ_1, ϕ_2, and define $\varphi_i(x,y) := \phi_i(x-y)$ such that eq. (S3.17) holds true. Here the integrals on the left hand side of (S3.17) can be estimated by

$$\int_a^b (\phi_i(x+\delta-y) - \phi_i(x-y))\,\mathrm{d}y \quad = \quad -\int_{x-b}^{x-b+\delta} \phi_i(u)\mathrm{d}u + \int_{x-a}^{x-a+\delta} \phi_i(u)\mathrm{d}u \leq 2|\delta|\|\phi_i\|_\infty,$$

and we obtain $|(\Phi f)(x) - (\Phi f)(x+\delta)| \leq 2\delta(\|\phi_1\|_\infty + \|\phi_2\|_\infty)\|f\|_\infty$.

ad **C**ii): Take $\mu = f + \sum_{j=1} \gamma_j \delta_{x_j} \in \mathcal{M}([a,b])$ and assume that

$$(\Phi\mu)(x) = \int_a^b \phi(x-y)f(y)\,\mathrm{d}y + \sum_{j=1}^n \gamma_j \phi(x-x_j) = 0, \qquad \text{for all } x \in [a,b].$$

Extending f by 0 on $\mathbb{R} \setminus [a,b]$, it follows from the Plancherel theorem and the Fourier convolution theorem that

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} f(x)(\Phi\mu)(x)\,\mathrm{d}x + \sum_{k=1}^n \gamma_k (\Phi\mu)(x_k) \\ &= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} \phi(x-y)f(y)\,\mathrm{d}y\,\mathrm{d}x + \sum_{j=1}^n \gamma_j \int_{-\infty}^{\infty} f(x)\phi(x-x_j)\,\mathrm{d}x \\ &\quad + \sum_{k=1}^n \gamma_k \int_{-\infty}^{\infty} \phi(x_k - y)f(y)\,\mathrm{d}y + \sum_{k=1}^n \sum_{j=1}^n \gamma_k\gamma_j\phi(x_j - x_k) \\ &= \int_{-\infty}^{\infty} \left| \widehat{f}(\xi) + \sum_{j=1}^n \gamma_j e^{-2\pi i \xi x_j} \right|^2 \widehat{\phi}(\xi)\,\mathrm{d}\xi\,. \end{aligned}$$

Using the assumption $\widehat{\phi} > 0$ a.e., we find that $\widehat{f}(\xi) + \sum_{j=1}^n \gamma_j e^{-2\pi i \xi x_j} = 0$ for a.e. $\xi \in \mathbb{R}$. Since $\lim_{|\xi|\to 0} \widehat{f}(\xi) = 0$ by the Riemann-Lebesgue lemma, this implies $\widehat{f} = 0$ and $\gamma_1 = \cdots = \gamma_n = 0$, so $\mu = 0$. □

Proof. (Theorem 4.4) ad **C**ii): This follows from the first part of Theorem 4.3 since analytic functions are of bounded variation.

ad **C**ii): Assume that $\Phi\mu = 0$ for $\mu = f + \sum_{j=1}\gamma_j\delta_{x_j} \in \mathcal{M}([a,b])$. Since ϕ is analytic, it has a holomorphic extension to a neighborhood $\mathcal{U}$ of $\mathbb{R}$ in $\mathbb{C}$. By a compactness argument $\mathcal{U}_0 := \bigcap_{y\in[a,b]} \mathcal{U} - y$ is also a neighborhood of $\mathbb{R}$ in $\mathbb{C}$. Define

$$g(z) := \int_a^b \phi(z-y)f(y)\,dy + \sum_{j=1}^n \gamma_j\phi(z-x_j), \qquad z \in \mathcal{U}_0.$$

Interchanging differentiation and integration, it follows that g is holomorphic, and $g(x) =$

$(\Phi\mu)(x) = 0$ for $x \in [a, b]$. Hence, g vanishes identically. Therefore,

$$0 = \int_{-\infty}^{\infty} e^{-2\pi i\xi x} g(x)\, dx = \widehat{\phi}(\xi) \left(\widehat{f}(\xi) + \sum_{j=1}^{n} \gamma_j e^{2\pi i \xi x_j} \right), \qquad \xi \in \mathbb{R}.$$

Since we have assumed that $\widehat{\phi} \neq 0$ a.e., it follows that the term in parenthesis vanishes a.e., and hence $\mu = 0$. □

References

Boysen, L., Bruns, S., and Munk, A. (2009), "Jump estimation in inverse regression," *Electron. J. Statist.*, 3, 1322–1359.

del Barrio, E., Deheuvels, P., and van de Geer, S. (2007), *Lectures on empirical processes*, EMS Series of Lectures in Mathematics, European Mathematical Society (EMS), Zürich, theory and statistical applications, With a preface by Juan A. Cuesta Albertos and Carlos Matrán.

Rudin, W. (1987), *Real and complex analysis*, New York: McGraw-Hill Book Co., 3rd ed.

tom Dieck, T. (2008), *Algebraic Topology*, EMS Textbooks in Mathematics, European Mathematical Society.

van de Geer, S. A. (2000), *Applications of empirical process theory*, vol. 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge: Cambridge University Press.

van der Vaart, A. W. (1998), *Asymptotic statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge: Cambridge University Press.

Yosida, K. (1995), *Functional analysis*, Classics in Mathematics, Berlin: Springer-Verlag, reprint of the sixth (1980) edition.