Estimation of a regression function corresponding to latent variables [1]

Michael Kohler[a], Florian Müller[a,2] and Harro Walk[b]
a Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@mathematik.tu-darmstadt.de, fmueller@mathematik.tu-darmstadt.de
b Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany, email: walk@mathematik.uni-stuttgart.de.

May 13, 2014

Abstract
The problem of estimation of a univariate regression function from latent variables given an independent and identically distributed sample of the observable variables in the corresponding common factor analysis model is considered. Nonparametric least squares estimates of the regression function are defined. The strong consistency of the estimates is shown for subgaussian random variables whose characteristic function vanishes nowhere. This consistency result does not require any assumptions on the structure or the smoothness of the regression function.

Key words and phrases: Common factor analysis, latent variables, least squares, nonparametric regression, consistency.

AMS classification: Primary 62G08; secondary 62G20.

1 Introduction

One of the fundamental problems of modern statistics is the problem of nonparametric regression. Here one considers an $\mathbb{R}^d \times \mathbb{R}$-valued random vector (Z_1, Z_2) with $\mathbf{E}Z_2^2 < \infty$ and the dependency of Z_2 on the value of Z_1 is of interest. More precisely, the goal is to find a function $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(Z_1)$ is a "good approximation" of Z_2. If the main aim of the analysis is minimization of the mean squared prediction error or L_2 risk

$$\mathbf{E}\{|f(Z_1) - Z_2|^2\}, \tag{1}$$

then the optimal function is the so-called regression function $m : \mathbb{R}^d \to \mathbb{R}$, $m(z_1) = \mathbf{E}\{Z_2 | Z_1 = z_1\}$. Indeed, for an arbitrary (measurable) function $f : \mathbb{R}^d \to \mathbb{R}$ we have

$$\mathbf{E}\{|f(Z_1) - Z_2|^2\} = \mathbf{E}\{|m(Z_1) - Z_2|^2\} + \int |f(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1) \tag{2}$$

[1]Running title: *Latent regression estimation*
[2]corresponding author, phone: +49 (0) 6151 - 16-2494

(cf., e.g., Section 1.1 in Györfi et al. (2002)). Since the integral on the right-hand side of (2) is always nonnegative, (2) implies that the regression function is the optimal predictor in view of minimization of the L_2 risk:

$$\mathbf{E}\{|m(Z_1) - Z_2|^2\} = \min_{f:\mathbb{R}^d \to \mathbb{R}} \mathbf{E}\{|f(Z_1) - Z_2|^2\}. \tag{3}$$

In addition, any function f is a good predictor in the sense that its L_2 risk is close to the optimal value, if and only if the so-called L_2 error

$$\int |f(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1) \tag{4}$$

is small. This motivates to measure the error caused by using a function f instead of the regression function by the L_2 error (4).

In applications, usually the distribution of (Z_1, Z_2) (and hence also the regression function) is unknown. But often it is possible to observe a sample of the underlying distribution. This leads to the regression estimation problem. Here (Z_1, Z_2), $(Z_{1,1}, Z_{2,1})$, $(Z_{1,2}, Z_{2,2}, \dots)$ are independent and identically distributed random vectors. The set of data

$$\mathcal{D}_n = \{(Z_{1,1}, Z_{2,1}), \dots, (Z_{1,n}, Z_{2,n})\}$$

is given, and the goal is to construct an estimate $m_n(\cdot) = m_n(\cdot, \mathcal{D}_n) : \mathbb{R}^d \to \mathbb{R}$ of the regression function such that the L_2 error $\int |m_n(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1)$ is small.

The problem of nonparametric regression in the above described setting is nowadays well understood. In particular it was shown in Stone (1977) that there exists universally consistent regression estimates, i.e., that suitably defined nearest neighbor estimates m_n satisfy

$$\mathbf{E} \int |m_n(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1) \to 0 \quad (n \to \infty)$$

for all distributions of (Z_1, Z_2) satisfying $\mathbf{E} Z_2^2 < \infty$. Related results have been shown for many other estimates, which include, e.g., results on the kernel regression estimate (cf., e.g., Nadaraya (1964, 1970), Watson (1964), Devroye and Wagner (1980) or Devroye and Krzyżak (1989)), the partitioning regression estimate (cf., e.g., Györfi (1981), Nobel (1996) or Beirlant and Györfi (1998)), the nearest neighbor regression estimate (cf., e.g., Devroye (1982), Devroye, Györfi, Krzyżak and Lugosi (1994), Mack (1981) or Zhao (1987)), least squares estimates (cf., e.g., Lugosi and Zeger (1995)) or smoothing spline estimates (cf., e.g., Kohler and Krzyżak (2001)). A summary of the main results can be found in the monograph by Györfi et al. (2002).

The main aim of the current paper is to try to achieve similar results in a setting, where the variables Z_1 and Z_2 cannot be observed explicitely, instead only observable indicators corresponding to these random variables are given. Such models are studied in various areas including psychology, social sciences, education or economics, where theoretical concepts such as intelligence, desirability or welfare are of great importance. These concepts cannot be measured directly but observable indicators (or manifest variables), such as scores on IQ-tests, are given. To model such concepts theoretically usually latent variable models are used.

The starting point of latent variable models are structural assumptions on the underlying latent variable model. Let X and Y be $\mathbb{R}^{d_X}$- and $\mathbb{R}^{d_Y}$-valued observable random variables (manifest variables). In order to analyze the relation between X and Y we assume that they depend linearly on some hidden and unobservable variables Z_1 and Z_2, where Z_1 and Z_2 are d_{Z_1}- and d_{Z_2}-dimensional random vectors, resp. We assume a simple structure in terms of a single cause of variation (i.e., a single latent variable) for each manifest variables. I.e., each of the components of the manifest variables is influenced by at most one of the components of the latent variables. In order to simplify the notation, we will assume furthermore that the two latent variables Z_1 and Z_2 are real-valued, in which case our model is given by

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(d)} \\ Y^{(1)} \\ Y^{(2)} \\ \vdots \\ Y^{(l)} \end{pmatrix} = \begin{pmatrix} 1 \cdot Z_1 \\ a_2 \cdot Z_1 \\ \vdots \\ a_d \cdot Z_1 \\ 1 \cdot Z_2 \\ b_2 \cdot Z_2 \\ \vdots \\ b_l \cdot Z_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_d \\ \delta_1 \\ \delta_2 \\ \vdots \\ \delta_l \end{pmatrix}. \tag{5}$$

Here $a_2, \ldots, b_l$ are real numbers and Z_1, Z_2, $\epsilon_1, \ldots \epsilon_d$, $\delta_1, \ldots, \delta_l$ are real random variables such that $\mathbf{E}\{\epsilon_j\} = \mathbf{E}\{\delta_k\} = 0$ holds.

One possibility to fit such a (or even a much more general) latent variable model to data is to assume that the underlying distribution is Gaussian, and therefore it is uniquely determined by its covariance structure. Then the maximum likelihood principle can be used to fit the latent variable model to observed data.

In this paper we try to avoid as far as possible assumptions on the distributions of the random variables occuring in the above latent variable model and we use techniques of nonparametric regression in order to analyze the dependency between the latent variables Z_1 and Z_2.

Of course, in order to be able to do this we need assumptions which ensure that the joint distribution of the latent variables is uniquely determined by the joint distribution of the observable manifest variables. The main trick here is that we make an independence assumption: We assume that random errors occurring in the manifest variables are independent from the latent variables, furthermore we need to assume that the characteristic function of the manifest variables does not vanish at any point.

Our main result is a consistency result of a nonparametric regression estimate of the regression function $m(z_1) = \mathbf{E}\{Z_2|Z_1 = z_1\}$ which holds whenever suitable exponential moments of the random variables in (5) exists. Here we do not need any assumption on structure or the smoothness of the regression function.

Known applications of regression estimation to latent variables in the literature are mainly restricted to parametric models, often formulated with so-called structural equations models, for surveys see, e.g., Marsh, Wen and Hau (2004) or Schumacker and Marcoulides (1998). In Paul et al. (2008) a high-dimensional linear regression problem is considered, where a low dimensional latent variable model determines the response variable. Principal component analysis is used to estimate the underlying latent variables, and it is assumed that all variables have a Gaussian distribution. A generalization of Gaussian latent variable models to the case that the manifest variables are indirect observations of normal underlying variables can be done via generalized linear latent variable models, cf., e.g., Conne, Ronchetti and Victoria-Feser (2010). Surveys on latent variables and its applications can be found, e.g., in Bollen (2002) and Skrondal and Rabe-Hesketh (2007).

The only nonparametric regression estimate we know in this context was proposed in Kelava et al. (2012), however the consistency result there is valid only for Lipschitz continuous regression functions where a bound on the Lipschitz constant must be known. Our method extends this results to non smooth regression functions by modifying the estimate there. Instead of trying to estimate the latent variable model (5) such that the independence assumption between latent variables and the errors in the manifest variables is satisfied, we estimate it such that certain relations between the characteristic function of the random variables in model (5) hold.

1.1 Notation

Throughout this paper we use the following notation: the sets of integers, rational numbers and real numbers are denoted by $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{R}$ respectively. For a real-valued function $f : \mathcal{D} \to \mathbb{R}$ with domain $\mathcal{D}$, let

$$\|f\|_\infty := \sup_{x \in \mathcal{D}} |f(x)|$$

denote its supremum norm. In the same spirit, we will use

$$\|f\|_{\infty,I} := \sup_{x \in I} |f(x)|$$

to denote the supremum of f over some subset $I \subset \mathcal{D}$.

We will call a random variable Z subgausian if there exists $c > 0$ such that

$$\mathbf{E}\{\exp(c|Z_1|^2)\} < \infty$$

holds.

If X is a $\mathbb{R}^d$-valued random variable, we denote the j-th component of X by $X^{(j)}$. The characteristic function of X is given by

$$\varphi_X(u) = \mathbf{E}\{e^{i \cdot u^T X}\} \quad (u \in \mathbb{R}^d).$$

Given an independent and identically distributed sample $X_1, X_2, X_3, \ldots$ of X, we will denote the empirical characteristic function by

$$\varphi_{X_1^n}(u) = \frac{1}{n} \sum_{j=1}^{n} e^{i \cdot u^T X_j}.$$

The distribution of X will be denoted by $\mathbf{P}_X$. For $f : \mathcal{D} \to \mathbb{R}$ we write

$$x = \arg\min_{z \in \mathcal{D}} f(z)$$

in case that

$$x \in \mathcal{D} \quad \text{and} \quad f(x) = \min_{z \in \mathcal{D}} f(z).$$

To simplify notation concerning rates of convergence, we will use a generic constant $C > 0$, i.e., the value of C might be different in each equation containing C.

1.2 Outline

The estimate is defined in Section 2, the main result is formulated in Section 3, and the proofs are given in Section 4.

2 Definition of the estimate

In the sequel we consider the common factor analysis (5) with the following additional assumptions:

A1) $d \geq 3$ and $l \geq 3$

A2) the coefficients $a_2, \ldots, a_d \in \mathbb{R}$ and $b_2, \ldots, b_l \in \mathbb{R}$ are all different from zero

A3) Z_1 and Z_2 are square integrable real-valued random variables with $\mathbf{E}\{Z_1^2\} > 0$ and $\mathbf{E}\{Z_2^2\} > 0$

A4) $\epsilon_1, \ldots, \epsilon_d$ and $\delta_1, \ldots, \delta_l$ are real-valued random variables with $\mathbf{E}\{\epsilon_j\} = \mathbf{E}\{\delta_k\} = 0$ for all $j \leq d$ and $k \leq l$

A5) (Z_1, Z_2), $\epsilon_1, \ldots, \epsilon_d$, $\delta_1, \ldots, \delta_l$ are independent

A6) the characteristic function of $(X, Y) := (X^{(1)}, \ldots, X^{(d)}, Y^{(1)}, \ldots, Y^{(l)})$ vanishes nowhere

Let

$$(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$$

be independent and identically distributed, and denote the latent variables and random errors corresponding to $X_i = (X_i^{(1)}, \dots, X_i^{(d)})^T$ and $Y_i = (Y_i^{(1)}, \dots, Y_i^{(l)})^T$ by $Z_{1,i}$, $Z_{2,i}$, $\epsilon_{1,i}$, $\dots$, $\delta_{l,i}$. Given

$$\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$$

we consider the problem of constructing an estimate $m_n(\cdot) = m_n(\cdot, \mathcal{D}_n) : \mathbb{R} \to \mathbb{R}$ of $m : \mathbb{R} \to \mathbb{R}$ defined by

$$m(z_1) = \mathbf{E}\{Z_2 | Z_1 = z_1\}$$

such that

$$\int |m_n(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1)$$

is small.

We start with the definition of the estimate by estimating the coefficients a_j and b_k. We set $a_1 := 1$ and $b_1 := 1$. By our independence assumption we observe that for $i \neq j$ we have

$$\begin{aligned}
\mathbf{E}\{X^{(j)} \cdot X^{(i)}\} &= \mathbf{E}\{(a_j Z_1 + \epsilon_j) \cdot (a_i Z_1 + \epsilon_i)\} \\
&= \mathbf{E}\{a_j a_i Z_1^2\} + \mathbf{E}\{\epsilon_j \cdot a_i Z_1\} + \mathbf{E}\{\epsilon_i \cdot a_j Z_1\} + \mathbf{E}\{\epsilon_i \cdot \epsilon_j\} \\
&= a_j a_i \mathbf{E}\{Z_1^2\}.
\end{aligned}$$

Thus, since Z_1, Z_2 are square integrable we have

$$a_2 = \frac{\mathbf{E}\{X^{(2)} \cdot X^{(3)}\}}{\mathbf{E}\{X^{(1)} \cdot X^{(3)}\}} \quad \text{and} \quad a_j = \frac{\mathbf{E}\{X^{(2)} \cdot X^{(j)}\}}{\mathbf{E}\{X^{(1)} \cdot X^{(2)}\}} \quad \text{for } j \in \{3, \dots, d\},$$

and similarly

$$b_2 = \frac{\mathbf{E}\{Y^{(2)} \cdot Y^{(3)}\}}{\mathbf{E}\{Y^{(1)} \cdot Y^{(3)}\}} \quad \text{and} \quad b_k = \frac{\mathbf{E}\{Y^{(2)} \cdot Y^{(k)}\}}{\mathbf{E}\{Y^{(1)} \cdot Y^{(2)}\}} \quad \text{for } k \in \{3, \dots, l\}.$$

We set $\hat{a}_1 = \hat{b}_1 = 1$ and construct our estimates as the empirical analogues

$$\hat{a}_2 = \frac{\frac{1}{n}\sum_{i=1}^n X_i^{(2)} \cdot X_i^{(3)}}{\frac{1}{n}\sum_{i=1}^n X_i^{(1)} \cdot X_i^{(3)}} \quad \text{and} \quad \hat{a}_j = \frac{\frac{1}{n}\sum_{i=1}^n X_i^{(2)} \cdot X_i^{(j)}}{\frac{1}{n}\sum_{i=1}^n X_i^{(1)} \cdot X_i^{(2)}}$$

and

$$\hat{b}_2 = \frac{\frac{1}{n}\sum_{j=1}^n Y_j^{(2)} \cdot Y_j^{(3)}}{\frac{1}{n}\sum_{j=1}^n Y_j^{(1)} \cdot Y_j^{(3)}} \quad \text{and} \quad \hat{b}_k = \frac{\frac{1}{n}\sum_{j=1}^n Y_j^{(2)} \cdot Y_j^{(k)}}{\frac{1}{n}\sum_{j=1}^n Y_j^{(1)} \cdot Y_j^{(2)}}$$

for $j, k > 2$.

Let $\hat{z}_{1,i}$ be an estimate of $Z_{1,i}$. By $X^{(j)} - a_j Z_1 = \epsilon_j$ for $j \in \{1, \dots, d\}$ we obtain estimates $\hat{\epsilon}_{j,i}$ of $\epsilon_{j,i} = X_i^{(j)} - a_j Z_{1,i}$ via

$$\hat{\epsilon}_{j,i} = X_i^{(j)} - \hat{a}_j \hat{z}_{1,i}.$$

The estimates of $\delta_{k,i}$ are similarly defined as

$$\hat{\delta}_{k,i} = Y_i^{(k)} - \hat{b}_k \hat{z}_{2,i}.$$

Next we define the estimates $(\hat{z}_{1,i}, \hat{z}_{2,i})$ of $(Z_{1,i}, Z_{2,i})$ for $i = 1, \dots, n$. It follows from Kelava et al. (2012)(cf., Lemma 1 below), that the following three relations between the characteristic functions of (X, Y), (Z_1, Z_2), ϵ_1 and δ_1 and their partial derivatives hold:

$$\varphi_{(X,Y)}(u_1, 0, \dots, 0, v_1, 0, \dots, 0) = \varphi_{(Z_1,Z_2)}(u_1, v_1) \cdot \varphi_{\epsilon_1}(u_1) \cdot \varphi_{\delta_1}(v_1),$$

$$\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(u_1, 0, \dots, 0, v_1, 0, \dots, 0) = a_2 \cdot \frac{\partial}{\partial u_1}\varphi_{(Z_1,Z_2)}(u_1, v_1) \cdot \varphi_{\epsilon_1}(u_1) \cdot \varphi_{\delta_1}(v_1)$$

and

$$\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u_1, 0, \dots, 0, v_1, 0, \dots, 0) = b_2 \cdot \frac{\partial}{\partial v_1}\varphi_{(Z_1,Z_2)}(u_1, v_1) \cdot \varphi_{\epsilon_1}(u_1) \cdot \varphi_{\delta_1}(v_1).$$

In the definition of our estimate we choose $(\hat{z}_{1,i}, \hat{z}_{2,i})$ such that these relations approximately hold between the corresponding empirical characteristic functions. More precisely, we choose $\hat{K}_n \in \mathbb{N}_0$ depending on the data $X_1, \dots, X_n, Y_1, \dots, Y_n$ by

$$\hat{K}_n := \min\left\{ \lfloor n^{\frac{1}{11}} \rfloor, \max\{K \in \mathbb{N} : \inf_{u,v \in [-\sqrt{K}, \sqrt{K}]} \left|\varphi_{(X^{(1)},Y^{(1)})_1^n}(u\pi, v\pi)\right| \geq \frac{1}{\log(n)}\} \right\}. \quad (6)$$

We will show in the proofs that $\hat{K}_n$ tends to infinity almost surely. Let us define a truncation height by $\hat{L}_n := \sqrt{\hat{K}_n}$ and set $\hat{L}_n^* := 2 \cdot \hat{L}_n = 2 \cdot \sqrt{\hat{K}_n}$. We construct an estimated sample $(\hat{z}_{1,1}, \hat{z}_{2,1}), \dots, (\hat{z}_{1,n}, \hat{z}_{2,n})$ of (Z_1, Z_2) by minimizing

$$\begin{aligned}
T_n(\hat{z}_1, \hat{z}_2) := & \\
\sup_{u_1,v_1 \in [-2\hat{K}_n, 2\hat{K}_n]} & \left\{ \left| \varphi_{(X^{(1)},Y^{(1)})_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}) - \varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}(\frac{\pi u_1}{\hat{L}_n^*}) \cdot \varphi_{\hat{\delta}_{1,1}^n}(\frac{\pi v_1}{\hat{L}_n^*}) \right| \right. \\
& + \left| \frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, 0, \dots, 0, \frac{\pi v_1}{\hat{L}_n^*}, 0, \dots, 0) \right. \\
& \qquad \left. - \hat{a}_2 \cdot \frac{\partial}{\partial u_1}\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}(\frac{\pi u_1}{\hat{L}_n^*}) \cdot \varphi_{\hat{\delta}_{1,1}^n}(\frac{\pi v_1}{\hat{L}_n^*}) \right| \\
& + \left| \frac{\partial}{\partial v_2}\varphi_{(X,Y)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, 0, \dots, 0, \frac{\pi v_1}{\hat{L}_n^*}, 0, \dots, 0) \right.
\end{aligned}$$

$$- \hat{b}_2 \cdot \frac{\partial}{\partial v_1} \varphi_{(\hat{z}_1, \hat{z}_2)_1^n}\left(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\right) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}\left(\frac{\pi u_1}{\hat{L}_n^*}\right) \cdot \varphi_{\hat{\delta}_{1,1}^n}\left(\frac{\pi v_1}{\hat{L}_n^*}\right) \Bigg| \Bigg\} \tag{7}$$

under the constraint $\frac{1}{n} \cdot \sum_{j=1}^n \hat{z}_{2,j}^8 \leq \sqrt{\hat{K}_n}$. We estimate the regression function m by a least squares estimate

$$m_n(\cdot) = \arg \min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n |\hat{z}_{i,2} - f(\hat{z}_{i,1})|^2, \tag{8}$$

where $\mathcal{F}_n$ is the set of trigonometric polynomials

$$\mathcal{F}_n = \Bigg\{ f : \mathbb{R} \to \mathbb{R} \Bigg| \\ f(z) = \sum_{k=0}^{\hat{K}_n} a_k \cos\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z\right) + \sum_{k=0}^{\hat{K}_n} b_k \sin\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z\right), \quad a_k, b_k \in [-\beta_n, \beta_n] \Bigg\}. \tag{9}$$

Here the degree $\hat{K}_n$ is defined in (6) and $\beta_n > 0$ is a parameter of the estimate.

Our main result is the following theorem.

Theorem 1. *Assume that in the model (5) the assumptions* $A1), \ldots A6)$ *hold. Furthermore assume that* Z_1 *and* Z_2 *as well as* $\epsilon_1, \ldots, \delta_l$ *are subgaussian.*

Define $\hat{K}_n$ *by (6) and* $\mathcal{F}_n$ *as in* (9) *with* $\hat{L}_n^* = 2 \cdot \hat{L}_n = 2 \cdot \sqrt{\hat{K}_n}$ *and* $\beta_n = n^{\frac{1}{11}}$. *Define the estimated sample* $(\hat{z}_{1,i}, \hat{z}_{2,i})$ *of* (Z_1, Z_2) *by minimizing (7) under the constraint* $\frac{1}{n} \cdot \sum_{j=1}^n \hat{z}_{2,j}^8 \leq \sqrt{\hat{K}_n}$. *Define* m_n *as in* (8).

Then the L_2*-error of* m_n *converges to* 0 *almost surely, i.e.,*

$$\int |m_n(x) - m(x)|^2 \mathbf{P}_{Z_1}(dx) \to 0 \quad (n \to \infty)$$

with probability one.

Remark 1. The above consistency result holds whenever several regularity assumption in the common factor analysis model (5) are satisfied. First of all, we need that in our common factor analysis model the distribution of the observable variables uniquely determines the distribution of the latent variables, therefore we assume that the errors are independent of the latent variables and have mean zero, and that the characteristic function of the observable random variables does not vanish at any point. Secondly we need that all random variables satisfy some exponential moment condition. Otherwise we do not need any regularity conditions, in particular we do not need that the regression function satisfies some structural assumption or that it is smooth, and we do not make any parametric assumption on the underlying distribution.

Remark 2. It should be possible to extend the above result to a latent variable model where the dimension of Z_1 is multivariate and to derive in this model a consistency result of the corresponding multivariate regression function. In order to simplify the notation we did not try to do this in the current paper.

Remark 3. It is an open problem how to extend the above result to a more general least squares estimates (based on a general function space) and whether the above consistency result holds under weaker moment conditions.

Remark 4. As soon as one imposes smoothness assumption on the regression function, it should be possible to derive in the above common factor analysis model even results concerning the rate of convergence of the L_2 error. It is an open problem what the optimal minimax rates of convergence are in the above model and how one can define estimates which achieve them.

3 Proofs

In the proofs we will need bounded variants of the random variables occurring in the latent variable model. We define the truncated Z_1 by

$$Z_{1,\hat{L}_n} := \begin{cases} Z_1 & \text{for } |Z_1| < \hat{L}_n \\ sign(Z_1) \cdot \hat{L}_n & \text{for } |Z_1| \geq \hat{L}_n. \end{cases} \tag{10}$$

Truncated variants of $Z_2, X^{(1)}, \ldots, Y^{(1)}, \epsilon_1, \ldots, \delta_1$ are defined similarly.

Furthermore we need a set of trigonometric polynomials in two variables corresponding to $\mathcal{F}_n$. Let $\mathcal{G}_n$ be the set consisting of all functions of the form

$$\begin{aligned} g(z_1, z_2) \quad = \quad & \sum_{j,k=0}^{2\cdot\hat{K}_n} a_{j,k} \cdot \cos\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right) \cdot \cos\left(j \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_2\right) \\ & + \sum_{j,k=0}^{2\cdot\hat{K}_n} b_{j,k} \cdot \cos\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right) \cdot \sin\left(j \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_2\right) \\ & + \sum_{j,k=0}^{2\cdot\hat{K}_n} c_{j,k} \cdot \sin\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right) \cdot \cos\left(j \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_2\right) \\ & + \sum_{j,k=0}^{2\cdot\hat{K}_n} d_{j,k} \cdot \sin\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right) \cdot \sin\left(j \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_2\right) \end{aligned}$$

where $a_{j,k}, b_{j,k}, c_{j,k}, d_{j,k} \in [-10 \cdot \hat{K}_n \cdot \beta_n^2, 10 \cdot \hat{K}_n \cdot \beta_n^2]$.

Lemma 1. *Let $X^{(1)}, \ldots X^{(d)}$, $Y^{(1)}, \ldots, Y^{(l)}$, Z_1, Z_2, $\epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_l$ be real-valued random variables such that the assumptions A1) to A6) are fulfilled. If* (5) *holds for some constants $a_2, \ldots, a_d, b_2, \ldots, b_l$ we have*

$$\varphi_{(X,Y)}(u_1, 0, \ldots, 0, v_1, 0, \ldots, 0) = \varphi_{(Z_1,Z_2)}(u_1, v_1) \cdot \varphi_{\epsilon_1}(u_1) \cdot \varphi_{\delta_1}(v_1),$$

$$\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(u_1,0,\dots,0,v_1,0,\dots,0) = a_2\cdot\frac{\partial}{\partial u_1}\varphi_{(Z_1,Z_2)}(u_1,v_1)\cdot\varphi_{\epsilon_1}(u_1)\cdot\varphi_{\delta_1}(v_1)$$

and

$$\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u_1,0,\dots,0,v_1,0,\dots,0) = b_2\cdot\frac{\partial}{\partial v_1}\varphi_{(Z_1,Z_2)}(u_1,v_1)\cdot\varphi_{\epsilon_1}(u_1)\cdot\varphi_{\delta_1}(v_1).$$

Furthermore, the coefficients $a_2,\dots,a_d$,$b_2,\dots,b_l$ *and the distribution of* (Z_1,Z_2) *as well as the distributions of the errors* $\epsilon_1,\dots,\epsilon_d,\delta_1,\dots,\delta_l$ *are uniquely determined by the distribution of* $(X^{(1)},\dots,X^{(d)},Y^{(1)},\dots,Y^{(l)})$.

Proof. This is Lemma 1 in Kelava et al. (2012). □

Lemma 2. *Let* $(\hat{K}_n)_n$ *be a sequence of* $\mathbb{R}$*-valued random variables with* $\hat{K}_n\to\infty$ *almost surely and let* $(\beta_n)_n$ *be a sequence of* $\mathbb{R}$*-valued random variables. If*

$$\hat{K}_n^3\cdot\beta_n^2\cdot\sup_{k,j\in\mathbb{Z},|k|\le 2\cdot\hat{K}_n,|j|\le 2\cdot\hat{K}_n}\left|\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}\left(j\cdot\frac{\pi}{\hat{L}_n^*},k\cdot\frac{\pi}{\hat{L}_n^*}\right)-\varphi_{(Z_1,Z_2)}\left(j\cdot\frac{\pi}{\hat{L}_n^*},k\cdot\frac{\pi}{\hat{L}_n^*}\right)\right|\to 0$$

almost surely, then

$$\sup_{g\in\mathcal{G}_n}\left|\frac{1}{n}\sum_{l=1}^n g(\hat{z}_{1,l},\hat{z}_{2,l})-\int g(Z_1,Z_2)\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\right|\to 0$$

almost surely.

Proof. Set

$$A_{n,1}:=\sup_{k,j\in\mathbb{Z},|k|\le 2\cdot\hat{K}_n,|j|\le 2\cdot\hat{K}_n}\left|\frac{1}{n}\sum_{l=1}^n\cos\left(k\cdot\frac{\pi}{\hat{L}_n^*}\cdot\hat{z}_{1,l}\right)\cdot\cos\left(j\cdot\frac{\pi}{\hat{L}_n^*}\cdot\hat{z}_{2,l}\right)\right.$$
$$\left.-\int\cos\left(k\cdot\frac{\pi}{\hat{L}_n^*}\cdot z_1\right)\cdot\cos\left(j\cdot\frac{\pi}{\hat{L}_n^*}\cdot z_2\right)\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\right|,$$

$$A_{n,2}:=\sup_{k,j\in\mathbb{Z},|k|\le 2\cdot\hat{K}_n,|j|\le 2\cdot\hat{K}_n}\left|\frac{1}{n}\sum_{l=1}^n\cos\left(k\cdot\frac{\pi}{\hat{L}_n^*}\cdot\hat{z}_{1,l}\right)\cdot\sin\left(j\cdot\frac{\pi}{\hat{L}_n^*}\cdot\hat{z}_{2,l}\right)\right.$$
$$\left.-\int\cos\left(k\cdot\frac{\pi}{\hat{L}_n^*}\cdot z_1\right)\cdot\sin\left(j\cdot\frac{\pi}{\hat{L}_n^*}\cdot z_2\right)\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\right|,$$

$$A_{n,3}:=\sup_{k,j\in\mathbb{Z},|k|\le 2\cdot\hat{K}_n,|j|\le 2\cdot\hat{K}_n}\left|\frac{1}{n}\sum_{l=1}^n\sin\left(k\cdot\frac{\pi}{\hat{L}_n^*}\cdot\hat{z}_{1,l}\right)\cdot\cos\left(j\cdot\frac{\pi}{\hat{L}_n^*}\cdot\hat{z}_{2,l}\right)\right.$$

$$- \int \sin\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right) \cdot \cos\left(j \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_2\right) \mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\Bigg|$$

and

$$A_{n,4} := \sup_{k,j \in \mathbb{Z}, |k| \le 2\cdot\hat{K}_n, |j| \le 2\cdot\hat{K}_n} \Bigg| \frac{1}{n}\sum_{l=1}^{n} \sin\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot \hat{z}_{1,l}\right) \cdot \sin\left(j \cdot \frac{\pi}{\hat{L}_n^*} \cdot \hat{z}_{2,l}\right) - \int \sin\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right) \cdot \sin\left(j \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_2\right) \mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\Bigg|.$$

By assumption the product of $\hat{K}_n^3 \beta_n^2$ and the absolute value of

$$\Theta_{j,k} := \frac{1}{n}\sum_{l=1}^{n} \exp\big(i\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l} + \frac{\pi \cdot k}{\hat{L}_n^*}\hat{z}_{2,l}\big)\big) - \mathbf{E}\left\{\exp\big(i \cdot \big(\frac{\pi \cdot j}{\hat{L}_n^*} \cdot Z_1 + \frac{\pi \cdot k}{\hat{L}_n^*} \cdot Z_2\big)\big)\right\}$$

converges uniformly to 0 a.s. for every $(j,k) \in \{-2\cdot\hat{K}_n, -2\cdot\hat{K}_n+1, \dots, 2\cdot\hat{K}_n\}^2$. Thus $\hat{K}_n^3\beta_n^2 \cdot |(\Theta_{j,k} + \Theta_{j,(-k)})|$ also converges to 0 uniformly almost surely. By

$$Re \exp(ix) = \cos(x) \quad \text{and} \quad \cos(x+y) + \cos(x-y) = 2\cos(x)\cos(y) \quad (x, y \in \mathbb{R})$$

we obtain

$$\begin{aligned}
& Re\left(\exp\big(i\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l} + \frac{\pi \cdot k}{\hat{L}_n^*}\hat{z}_{2,l}\big)\big)\right) + Re\left(\exp\big(i\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l} + \frac{\pi \cdot (-k)}{\hat{L}_n^*}\hat{z}_{2,l}\big)\big)\right) \\
&= \cos\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l} + \frac{\pi \cdot k}{\hat{L}_n^*}\hat{z}_{2,l}\big) + \cos\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l} + \frac{\pi \cdot (-k)}{\hat{L}_n^*}\hat{z}_{2,l}\big) \\
&= 2\cos\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l}\big) \cdot \cos\big(\frac{\pi \cdot k}{\hat{L}_n^*}\hat{z}_{2,l}\big).
\end{aligned}$$

Together with

$$\begin{aligned}
& Re\left(\Theta_{j,k} + \Theta_{j,(-k)}\right) = \\
& \frac{1}{n}\sum_{l=1}^{n} Re\left(\exp\big(i\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l} + \frac{\pi \cdot k}{\hat{L}_n^*}\hat{z}_{2,l}\big)\big)\right) + Re\left(\exp\big(i\big(\frac{\pi \cdot j}{\hat{L}_n^*}\hat{z}_{1,l} + \frac{\pi \cdot (-k)}{\hat{L}_n^*}\hat{z}_{2,l}\big)\big)\right) \\
& - \mathbf{E}\Bigg\{ Re\left(\exp\big(i\big(\frac{\pi \cdot j}{\hat{L}_n^*} \cdot Z_1 + \frac{\pi \cdot (-k)}{\hat{L}_n^*} \cdot Z_2\big)\big)\right) \\
& \qquad + Re\left(\exp\big(i\big(\frac{\pi \cdot j}{\hat{L}_n^*} \cdot Z_1 + \frac{\pi \cdot k}{\hat{L}_n^*} \cdot Z_2\big)\big)\right)\Bigg\},
\end{aligned}$$

we get

$$\hat{K}_n^3\beta_n^2 \sup_{k,j \in \mathbb{Z}, |k| \le 2\cdot\hat{K}_n, |j| \le 2\cdot\hat{K}_n} |Re(\Theta_{j,k} + \Theta_{j,(-k)})| = 2 \cdot \hat{K}_n^3\beta_n^2 A_{n,1}.$$

Since

$$\begin{aligned}
&\hat{K}_n^3\beta_n^2 \sup_{k,j\in\mathbb{Z},|k|\leq 2\cdot\hat{K}_n,|j|\leq 2\cdot\hat{K}_n} |Re(\Theta_{j,k}+\Theta_{j,(-k)})| \\
\leq\quad & \hat{K}_n^3\beta_n^2 \sup_{k,j\in\mathbb{Z},|k|\leq 2\cdot\hat{K}_n,|j|\leq 2\cdot\hat{K}_n} |(\Theta_{j,k}+\Theta_{j,(-k)})| \to 0 \quad (a.s.)
\end{aligned}$$

we obtain $\hat{K}_n^3\beta_n^2 A_{n,1}\to 0$ almost surely. With $Im(\exp(ix)) = sin(x)$,

$$\sin(x-y)+\sin(x+y) = 2\sin(x)\cos(y)$$

and

$$\cos(x-y)-\cos(x+y) = 2\sin(x)\sin(y)$$

analogous results follow for $A_{n,2}, A_{n,3}$ and $A_{n,4}$. Because of

$$\begin{aligned}
&\sup_{g\in\mathcal{G}_n}\left|\frac{1}{n}\sum_{l=1}^n g(\hat{z}_{1,l},\hat{z}_{2,l}) - \int g(z_1,z_2)\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\right| \\
&= \sup_{\substack{a_{j,k},\,b_{j,k},\,c_{j,k},\,d_{j,k}\\ \in[-10\cdot\hat{K}_n\beta_n^2,10\cdot\hat{K}_n\beta_n^2]}} \Bigg|\frac{1}{n}\sum_{l=1}^n\sum_{j,k=0}^{2\cdot\hat{K}_n}\Big[a_{j,k}\cos\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big) \\
&\qquad + b_{j,k}\cos\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big) + c_{j,k}\sin\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big) \\
&\qquad + d_{j,k}\sin\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big)\Big] \\
&\quad - \int\Big[a_{j,k}\cos\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}z_2\big) + b_{j,k}\cos\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}z_2\big) \\
&\quad + c_{j,k}\sin\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}z_2\big) + d_{j,k}\sin\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}z_2\big)\Big]d\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\Bigg| \\
&\leq 10\cdot\hat{K}_n\cdot\beta_n^2\sum_{j,k=0}^{2\cdot\hat{K}_n}\Big|\frac{1}{n}\sum_{l=1}^n\cos\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big) \\
&\qquad\qquad - \int\cos\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}z_2\big)\ d\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\Big| \\
&\quad + \Big|\frac{1}{n}\sum_{l=1}^n\cos\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big) - \int\cos\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}z_2\big)\ d\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\Big| \\
&\quad + \Big|\frac{1}{n}\sum_{l=1}^n\sin\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big) - \int\sin\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\cos\big(j\frac{\pi}{\hat{L}_n^*}z_2\big)\ d\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\Big| \\
&\quad + \Big|\frac{1}{n}\sum_{l=1}^n\sin\big(k\frac{\pi}{\hat{L}_n^*}\hat{z}_{1,l}\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}\hat{z}_{2,l}\big) - \int\sin\big(k\frac{\pi}{\hat{L}_n^*}z_1\big)\sin\big(j\frac{\pi}{\hat{L}_n^*}z_2\big)\ d\mathbf{P}_{(Z_1,Z_2)}(d(z_1,z_2))\Big|
\end{aligned}$$

$$\leq 10 \cdot \hat{K}_n \beta_n^2 \cdot (2 \cdot \hat{K}_n + 1)^2 \cdot \sum_{r=1}^{4} A_{n,r} \leq 90 \cdot \hat{K}_n^3 \beta_n^2 \cdot \sum_{r=1}^{4} A_{n,r}$$

for $\hat{K}_n \geq 1$, this implies the assertion. □

Lemma 3. *Let $\mathcal{F}_n$ be the set consisting of all functions of the form*

$$f(z_1) \quad = \quad \sum_{k=0}^{\hat{K}_n} a_k \cdot \cos\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right) + \sum_{k=0}^{\hat{K}_n} b_k \cdot \sin\left(k \cdot \frac{\pi}{\hat{L}_n^*} \cdot z_1\right)$$

where $a_k, b_k \in [-\beta_n, \beta_n]$. Choose $\hat{K}_n$ depending on the data $X_1^{(1)}, \ldots, X_n^{(1)}, Y_1^{(1)}, \ldots Y_n^{(1)}$ as in (6). Assume $\hat{K}_n \to \infty$ almost surely, set $\hat{L}_n^* = 2 \cdot \hat{L}_n = 2 \cdot \sqrt{\hat{K}_n}$ and $\beta_n = n^{\frac{1}{11}}$. Furthermore, assume that Z_1 and Z_2 are subgaussian. Define

$$m_n(\cdot) = \arg\min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} |\hat{z}_{i,2} - f(\hat{z}_{i,1})|^2.$$

Then

$$\hat{K}_n^3 \cdot \beta_n^2 \cdot \sup_{k,j \in \mathbb{Z}, |k| \leq 2\hat{K}_n, |j| \leq 2\hat{K}_n} \left| \varphi_{(\hat{z}_1, \hat{z}_2)_1^n}\left(j \cdot \frac{\pi}{\hat{L}_n^*}, k \cdot \frac{\pi}{\hat{L}_n^*}\right) - \varphi_{(Z_1, Z_2)}\left(j \cdot \frac{\pi}{\hat{L}_n^*}, k \cdot \frac{\pi}{\hat{L}_n^*}\right) \right| \to 0$$

almost surely implies

$$\int |m_n(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1) \to 0$$

almost surely.

Proof. We use a variant of the decomposition for the L_2 error of least square estimates (cf., e.g., Lemma 10.1 in Györfi et al. (2002)) and obtain, for $\delta > 0$,

$$\begin{aligned}
&\int |m_n(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1) = \mathbf{E}\big\{|m_n(Z_1) - Z_2|^2 \big| \mathcal{D}_n\big\} - \mathbf{E}\big\{|m(Z_1) - Z_2|^2\big\} \\
&= \mathbf{E}\big\{|m_n(Z_1) - Z_2|^2 \big| \mathcal{D}_n\big\} - (1+\delta)^8 \inf_{f \in \mathcal{F}_n} \mathbf{E}\{|f(Z_1) - Z_2|^2\} \\
&\quad + (1+\delta)^8 \left[\inf_{f \in \mathcal{F}_n} \mathbf{E}\{|f(Z_1) - Z_2|^2\} - \mathbf{E}\big\{|m(Z_1) - Z_2|^2\big\}\right] \\
&\quad + (1+\delta)^8 \mathbf{E}\big\{|m(Z_1) - Z_2|^2\big\} - \mathbf{E}\big\{|m(Z_1) - Z_2|^2\big\} \\
&= T_{1,n} + T_{2,n} + T_{3,n}.
\end{aligned}$$

In the next steps of the proof we will establish upper bounds for $T_{1,n}, T_{2,n}$ and $T_{3,n}$. Let $\varepsilon > 0$ be arbitrary. Since $\mathbf{E}\big\{|m(Z_1) - Z_2|^2\big\}$ is independent of n, we may choose δ such that

$$T_{3,n} = ((1+\delta)^8 - 1)\mathbf{E}\big\{|m(Z_1) - Z_2|^2\big\} \leq \varepsilon$$

holds. By the standard decomposition of the L_2-risk (cf., e.g., Section 1.1 in Györfi et al. (2002)) we see that

$$T_{2,n} = (1+\delta)^8 \inf_{f \in \mathcal{F}_n} \int |f(z_1) - m(z_1)|^2 \mathbf{P}_{Z_1}(dz_1).$$

The function m can be approximated by smooth functions with compact support in $L_2(\mathbf{P}_{Z_1})$ (cf., e.g., Theorem A.1 in Györfi et al. (2002)). Thus it suffices to show that continuously differentiable functions with compact support can be approximated arbitrarily well in $L_2(\mathbf{P}_{Z_1})$ by functions $f \in \bigcup_{n\in\mathbb{N}} \mathcal{F}_n$. Let m^* be a smooth function with compact support. By our assumption on $\hat{K}_n$, we have $\hat{L}_n^* \to \infty$ *a.s.* and hence the support of m^* is *a.s.* contained in $[-\hat{L}_n^*, \hat{L}_n^*]$ for n large enough. Thus, for n large enough, there exists a uniquely determined Lipschitz continuous and $2\hat{L}_n^*$ periodic continuation of m^* from $[-\hat{L}_n^*, \hat{L}_n^*]$ to $\mathbb{R}$. Denote this continuation of m^* by m_n^*. Clearly, m_n^* is bounded by $\|m^*\|_\infty$. Additionally, the coefficients of the Fourier series of m_n^* are bounded by $\|m_n^*\|_\infty$. Thus the coefficients are bounded by $\|m^*\|_\infty$. By $\beta_n = n^{\frac{1}{11}}$, we see that the $\hat{K}_n$-th partial sum of the Fourier series of m_n^*, which we denote by $S_{\hat{K}_n, m_n^*}$, is contained in $\mathcal{F}_n$ for n large enough. Since m^* is Lipschitz continuous, every m_n^* is also Lipschitz continuous with the same Lipschitz constant K.

Now, if f is a Lipschitz continuous function with Lipschitz constant λ and period $2L$, then $x \mapsto f\big(\frac{L}{\pi} \cdot x)$ ist 2π periodic and is Lipschitz continuous with Lipschitz constant $\frac{L}{\pi} \cdot \lambda$. Using this we can apply Corollary 1 from Jackson (1930) and obtain

$$\inf_{f\in\mathcal{F}_n} \|f - m^*\|_{\infty,[-\hat{L}_n^*, \hat{L}_n^*]} \leq \|S_{\hat{K}_n, m^*} - m^*\|_{\infty,[-\hat{L}_n^*, \hat{L}_n^*]} \leq C \frac{\hat{L}_n \ln(\hat{K}_n)}{\hat{K}_n}. \tag{11}$$

for some $C > 0$. Here we have used that the trigonometric polynomials contained in $\mathcal{F}_n$ are of degree $\hat{K}_n$. Outside of this interval, m^* vanishes for $\hat{L}_n^*$ large enough. Z_1 and $\hat{L}_n^*$ are independent since $\hat{L}_n^*$ depends only on the data $(X_1, Y_1), \ldots$. Using the same upper bound for the coeficents of $S_{\hat{K}_n, m_n^*}$ as above as well as the inequality of Markov we get that almost surely, for n large enough,

$$\begin{aligned}
\inf_{f\in\mathcal{F}_n} \int_{\{|x|\geq \hat{L}_n^*\}} |f(x) - m^*(x)|^2 \mathbf{P}_{Z_1}(dx) &\leq \int_{\{|x|\geq \hat{L}_n^*\}} S_{\hat{K}_n, m_n^*}(x)^2 \mathbf{P}_{Z_1}(dx) \\
&\leq ((2\cdot \hat{K}_n + 1)\cdot \|m^*\|_\infty)^2 \cdot \mathbf{P}[|Z_1| \geq \hat{L}_n^*] \\
&\leq 9 \cdot \|m^*\|_\infty^2 \cdot \hat{K}_n^2 \cdot \mathbf{E}\Big\{\exp(c\cdot |Z_1|^2) \frac{1}{\exp(c\cdot (\hat{L}_n^*)^2)}\Big\} \\
&\leq C\hat{K}_n^2 \cdot \mathbf{E}\big\{\exp(-c\cdot \hat{K}_n)\big\} \qquad (12)
\end{aligned}$$

holds for some $C \geq 0$ depending only on m^*. Here we have used that Z_1 is subgaussian. By definition $\hat{L}_n^* \to \infty$ and $\hat{K}_n \to \infty$ almost surely, which implies that (11) tends to 0 almost surely. Using dominated convergence, we see that (12) tends to 0. This yields

$T_{2,n} \to 0$ almost surely.

To establish an upper bound for $T_{1,n}$ we use an approximation through trigonometric polynomials in two variables. Consider $h_{1,n} : [-\hat{L}_n, \hat{L}_n] \to \mathbb{R}$, $h_{1,n}(z_2) = z_2$. Define $\tilde{h}_{1,n} : \mathbb{R} \to \mathbb{R}$ as the uniquely determined $4\hat{L}_n^*$-periodic continuation of

$$x \mapsto \begin{cases} 2\hat{L}_n - x & \text{for } x \in [\hat{L}_n, 3\hat{L}_n] \\ x & \text{for } x \in (-\hat{L}_n, \hat{L}_n). \end{cases}$$

For every $n \in \mathbb{N}$, $\tilde{h}_{1,n}$ is a periodic continuation of h_1 and is Lipschitz continuous with Lipschitz constant one. The Fourier coefficients of $\tilde{h}_{1,n}$ are bounded by $\|\tilde{h}_{1,n}\|_\infty = \hat{L}_n$. Together with

$$\hat{L}_n \leq \hat{K_n} \leq n^{\frac{1}{11}} = \beta_n,$$

this implies that the $\hat{K}_n$-th partial sum of the Fourier series $g_{1,n}$ of $\tilde{h}_{1,n}$ is contained in $\mathcal{F}_n$. For a $f_n \in \mathcal{F}_n$ we define $g_{f_n,n}(z_1, z_2) := f_n(z_1) - g_{1,n}(z_2)$. Now $g_{f_n,n}$ is a trigonometric polynomial in two variables of degree $\hat{K}_n$. Using a similar argument as above and applying Corollary 1 from Jackson (1930) again, we obtain

$$\sup_{z_1,z_2 \in [-\hat{L}_n^*, \hat{L}_n^*]} \left|\left(f_n(z_1) - \tilde{h}_{1,n}(z_2)\right) - g_{f_n,n}(z_1, z_2)\right| = \sup_{z_2 \in [-\hat{L}_n^*, \hat{L}_n^*]} \left|\tilde{h}_{1,n}(z_2) - g_{1,n}(z_2)\right| \to 0$$

for $n \to \infty$. Since $\tilde{h}_{1,n}$ is a continuation of $h_{1,n}$, we also observe that $g_{f_n,n}$ uniformly approximates $f(z_1) - z_2$ for $z_1, z_2 \in [-\hat{L}_n, \hat{L}_n]$. Additionally, although $g_{f_n,n}$ depends on $f_n \in \mathcal{F}_n$ the above error is independent of f_n by our construction of $g_{f_n,n}$.

Now we use the decomposition

$$\begin{aligned} T_{1,n} = \sup_{f \in \mathcal{F}_n} \Big[& \mathbf{E}\big\{|m_n(Z_1) - Z_2|^2 \big| \mathcal{D}_n\big\} - (1+\delta)^4 \frac{1}{n} \sum_{j=1}^n |m_n(\hat{z}_{1,j}) - \hat{z}_{2,j}|^2 \\ & + (1+\delta)^4 \Big(\frac{1}{n} \sum_{j=1}^n |m_n(\hat{z}_{1,j}) - \hat{z}_{2,j}|^2 - \frac{1}{n} \sum_{j=1}^n |f(\hat{z}_{1,j}) - \hat{z}_{2,j}|^2 \Big) \\ & + (1+\delta)^4 \frac{1}{n} \sum_{j=1}^n |f(\hat{z}_{1,j}) - \hat{z}_{2,j}|^2 - (1+\delta)^8 \mathbf{E}\{|f(Z_1) - Z_2|^2\} \Big] \\ = \sup_{f \in \mathcal{F}_n} & \big[T_{1,n}^{(1)} + T_{1,n}^{(2)} + T_{1,n}^{(3)}\big]. \end{aligned}$$

By our definition of m_n we have $T_{1,n}^{(2)} \leq 0$. This yields

$$T_{1,n} \leq \sup_{f \in \mathcal{F}_n} \big[T_{1,n}^{(1)} + T_{1,n}^{(3)}\big] = T_{1,n}^{(1)} + \sup_{f \in \mathcal{F}_n} T_{1,n}^{(3)}.$$

Next, we will use the elementary inequality

$$(a+b)^2 - (1+\delta)a^2 \leq (1+\frac{1}{\delta})b^2 \quad (\delta > 0\,; a, b \in \mathbb{R}). \tag{13}$$

We have

$$\begin{aligned}
T_{1,n}^{(1)} &= \mathbf{E}\big\{|m_n(Z_1) - Z_2|^2 \big| \mathcal{D}_n\big\} - (1+\delta)^4 \cdot \frac{1}{n}\sum_{j=1}^{n} |m_n(\hat{z}_{1,j}) - \hat{z}_{2,j}|^2 \\
&= \mathbf{E}\big\{|m_n(Z_1) - Z_2|^2 \big| \mathcal{D}_n\big\} - (1+\delta) \cdot \mathbf{E}\big\{|m_n(Z_1) - Z_{2,\hat{L}_n}|^2 \big| \mathcal{D}_n\big\} \\
&\quad + (1+\delta) \cdot \mathbf{E}\big\{|m_n(Z_1) - Z_{2,\hat{L}_n}|^2 \big| \mathcal{D}_n\big\} - (1+\delta)^2 \cdot \mathbf{E}\big\{g_{m_n,n}(Z_1, Z_{2,\hat{L}_n})^2 \big| \mathcal{D}_n\big\} \\
&\quad + (1+\delta)^2 \cdot \mathbf{E}\big\{g_{m_n,n}(Z_1, Z_{2,\hat{L}_n})^2 \big| \mathcal{D}_n\big\} - (1+\delta)^3 \cdot \mathbf{E}\big\{g_{m_n,n}(Z_1, Z_2)^2 \big| \mathcal{D}_n\big\} \\
&\quad + (1+\delta)^3 \cdot \Big(\mathbf{E}\big\{g_{m_n,n}(Z_1, Z_2)^2 \big| \mathcal{D}_n\big\} - \frac{1}{n}\sum_{j=1}^{n} g_{m_n,n}(\hat{z}_{1,j}, \hat{z}_{2,j})^2\Big) \\
&\quad + (1+\delta)^3 \cdot \frac{1}{n}\sum_{j=1}^{n} g_{m_n,n}(\hat{z}_{1,j}, \hat{z}_{2,j})^2 - (1+\delta)^4 \cdot \frac{1}{n}\sum_{j=1}^{n} |m_n(\hat{z}_{1,j}) - \hat{z}_{2,j}|^2.
\end{aligned}$$

Applying (13) yields

$$\begin{aligned}
T_{1,n}^{(1)} &\leq \left(1 + \frac{1}{\delta}\right) \mathbf{E}\Big\{\Big(m_n(Z_1) - Z_2 - \Big(m_n(Z_1) - Z_{2,\hat{L}_n}\Big)\Big)^2 \Big| \mathcal{D}_n\Big\} \\
&\quad + \big(1 + \frac{1}{\delta}\big)\big(1+\delta\big)\mathbf{E}\Big\{\Big(m_n(Z_1) - Z_{2,\hat{L}_n} - g_{m_n,n}(Z_1, Z_{2,\hat{L}_n})\Big)^2 \Big| \mathcal{D}_n\Big\} \\
&\quad + \big(1 + \frac{1}{\delta}\big)\big(1+\delta\big)^2\mathbf{E}\Big\{\Big(g_{m_n,n}(Z_1, Z_{2,\hat{L}_n}) - g_{m_n,n}(Z_1, Z_2)\Big)^2 \Big| \mathcal{D}_n\Big\} \\
&\quad + \big(1+\delta\big)^3\Big[\mathbf{E}\Big\{g_{m_n,n}(Z_1, Z_2)^2 \Big| \mathcal{D}_n\Big\} - \frac{1}{n}\sum_{j=1}^{n} g_{m_n,n}(\hat{z}_{1,j}, \hat{z}_{2,j})^2\Big] \\
&\quad + \big(1+\delta\big)^3\big(1 + \frac{1}{\delta}\big)\frac{1}{n}\sum_{j=1}^{n}\Big(g_{m_n,n}(\hat{z}_{1,j}, \hat{z}_{2,j}) - (m_n(\hat{z}_{1,j}) - \hat{z}_{2,j})\Big)^2 \\
&= \sum_{k=1}^{5} \tilde{T}_{1,n}^{(k)}.
\end{aligned}$$

By a similar argument, we derive an upper bound for $T_{1,n}^{(3)}$.

$$\begin{aligned}
T_{1,n}^{(3)} &\leq \left(1 + \frac{1}{\delta}\right)(1+\delta)^4\, \frac{1}{n}\sum_{j=1}^{n}\Big(f(\hat{z}_{1,j}) - \hat{z}_{2,j} - g_{f,n}(\hat{z}_{1,j}, \hat{z}_{2,j})\Big)^2 \\
&\quad + \big(1+\delta\big)^5\Big[\frac{1}{n}\sum_{j=1}^{n}\Big(g_{f,n}(\hat{z}_{1,j}, \hat{z}_{2,j})\Big)^2 - \mathbf{E}\Big\{g_{f,n}(Z_1, Z_2)^2\Big\}
\end{aligned}$$

$$
\begin{aligned}
&+ \big(1+\delta\big)^5 \big(1+\frac{1}{\delta}\big)\mathbf{E}\Big\{ \Big(g_{f,n}(Z_1,Z_2) - g_{f,n}(Z_1, Z_{2,\hat{L}_n})\Big)^2 \Big\} \\
&+ \big(1+\delta\big)^6 \big(1+\frac{1}{\delta}\big)\mathbf{E}\Big\{ \Big(g_{f,n}(Z_1,Z_{2,\hat{L}_n}) - \big(f(Z_1) - Z_{2,\hat{L}_n}\big)\Big)^2 \Big\} \\
&+ \big(1+\delta\big)^7 \big(1+\frac{1}{\delta}\big)\mathbf{E}\Big\{ \Big(f(Z_1) - Z_{2,\hat{L}_n} - \big(f(Z_1) - Z_2\big)\Big)^2 \Big\} \\
&= \sum_{k=6}^{10} \tilde{T}_{1,n}^{(k)}
\end{aligned}
$$

Since $m_n \in \mathcal{F}_n$ by our definition of the estimate, it now suffices to show that if we replace above m_n by some $f \in \mathcal{F}_n$ then the above expressions converge to 0 uniformly in $\mathcal{F}_n$. Beforehand, let us loosely classify the above expressions into three types. Let us denote $\tilde{T}_{1,n}^{(1)}$, $\tilde{T}_{1,n}^{(3)}$, $\tilde{T}_{1,n}^{(8)}$ and $\tilde{T}_{1,n}^{(10)}$ as expressions of type *i*. For these, we will use that $Z_{2,\hat{L}_n}$ converges to Z_2. $\tilde{T}_{1,n}^{(2)}$ and $\tilde{T}_{1,n}^{(9)}$ are of type *ii*. For these the approximation properties of $g_{f,n}$ are used. $\tilde{T}_{1,n}^{(5)}$ and $\tilde{T}_{1,n}^{(6)}$ are of type *iii*, where we will use a combination of the arguments used on type *i* and type *ii*. The remaining expressions $\tilde{T}_{1,n}^{(4)}$ and $\tilde{T}_{1,n}^{(7)}$ constitute type *iv*. Expression of type *iv* converge to 0 if Lemma 2 is applicable.

Let us begin with the expressions of type *i*. For $f \in \mathcal{F}_n$ we have

$$
g_{f,n}(x,y) = f(x) - g_{1,n}(y)
$$

where $g_{1,n}$ is our trigonometric approximation of the identity from above. Thus the differences in our expressions of type *i* are independent of $f \in \mathcal{F}_n$. Furthermore by $g_{1,n} \in \mathcal{F}_n$, $g_{1,n}$ is Lipschitz continuous. Since cosine and sine are Lipschitz continous with Lipschitz constant one and the coefficients of $g_{1,n}$, which are the Fourier-coefficients of $\tilde{h}_{1,n}$, are bounded by $\hat{L}_n$, there exist a $C \in \mathbb{R}$ such that $C\hat{K}_n^{\frac{3}{2}}$ is a Lipschitz constant for $g_{1,n}$. Since δ is constant, we only need to show

$$
\hat{K}_n^{\frac{3}{2}} \mathbf{E}\big\{(Z_{2,\hat{L}_n} - Z_2)^2\big\} \to 0 \quad a.s.
$$

By assumption, Z_2 is subgaussian. As in (12) we use that Z_2 is independent from L_n as well as dominated convergence. This yields

$$
\begin{aligned}
\hat{K}_n^{\frac{3}{2}} \mathbf{E}\big\{(Z_2 - Z_{2,\hat{L}_n})^2\big\} &\le \hat{K}_n^{\frac{3}{2}} \mathbf{E}\big\{\mathbb{1}_{\{|Z_2| > \hat{L}_n\}} Z_2^2\big\} \\
&\le \hat{K}_n^{\frac{3}{2}} \mathbf{E}\Big\{ Z_2^2 \frac{\exp\big(\frac{c_2}{2} Z_2^2\big)}{\exp\big(\frac{c_2}{2}\hat{L}_n^2\big)} \Big\} \\
&\le \hat{K}_n^{\frac{3}{2}} \mathbf{E}\Big\{ \frac{2}{c_2} \exp\big(\frac{c_2}{2} Z_2^2\big) \cdot \exp\big(\frac{c_2}{2} Z_2^2\big) \Big\} \cdot \mathbf{E}\Big\{ \exp\big(-\frac{c_2}{2}\hat{L}_n^2\big) \Big\} \\
&= \hat{K}_n^{\frac{3}{2}} \mathbf{E}\Big\{ \frac{2}{c_2} \exp\big(c_2 Z_2^2\big) \Big\} \mathbf{E}\Big\{ \exp\big(-\frac{c_2}{2}\hat{K}_n\big) \Big\} \to 0
\end{aligned} \tag{14}
$$

since $\hat{K}_n$ tends to infinity almost surely by assumption.

Let us consider type *ii* next. By the definition of $g_{f,n}$ the expression is independent of f. By $Z_{2,\hat{L}_n} \in [-\hat{L}_n, \hat{L}_n]$ it suffices to show the uniform convergence of $g_{1,n}$ to $\tilde{h}_{1,n}$ on $[-\hat{L}_n, \hat{L}_n]$. As above, we use that the Lipschitz constant of $\tilde{h}_{1,n}$ is independent of n. Since $g_{1,n}$ is defined as the n-th Fourier polynomial of $\tilde{h}_{1,n}$ and both functions are $2\hat{L}_n^*$ periodic, we use the rescaled functions $x \mapsto g_{1,n}\big(\frac{\hat{L}_n^*}{\pi} \cdot x\big)$ and $x \mapsto \tilde{h}_{1,n}\big(\frac{\hat{L}_n^*}{\pi} \cdot x\big)$ and apply Corollary 1 from Jackson (1930) again to obtain the same upper bound for the speed of convergence as above. This yields

$$\|g_{1,n} - \tilde{h}_{1,n}\|_{\infty,[-\hat{L}_n^*,\hat{L}_n^*]} \leq C\frac{\hat{L}_n \ln(\hat{K}_n)}{\hat{K}_n}.$$

and the right-hand side tends to zero almost surely by our assumptions on $\hat{L}_n$ and $\hat{K}_n$. Thus the integrand in $\tilde{T}_{1,n}^{(2)}$ and $\tilde{T}_{1,n}^{(9)}$ tends uniformly to 0. Thus both $\tilde{T}_{1,n}^{(2)}$ and $\tilde{T}_{1,n}^{(9)}$ tend to 0 almost surely.

Now we deal with type *iii*. Since δ is constant we can neglect the first to factors of both $\tilde{T}_{1,n}^{(5)}$ and $\tilde{T}_{1,n}^{(6)}$. Furthermore, by our definition of $g_{m_n,n}$, we have

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\Big(f(\hat{z}_{1,j}) - \hat{z}_{2,j} - g_{f,n}(\hat{z}_{1,j}, \hat{z}_{2,j})\Big)^2 &= \frac{1}{n}\sum_{j=1}^{n}\Big(m_n(\hat{z}_{1,j}) - \hat{z}_{2,j} - g_{m_n,n}(\hat{z}_{1,j}, \hat{z}_{2,j})\Big)^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}(\hat{z}_{2,j} - g_{1.n}(\hat{z}_{2,j}))^2 = J_n.\end{aligned}$$

We construct an upper bound for J_n by introducing the truncated version $\hat{z}_{2,j,\hat{L}_n}$ of $\hat{z}_{2,j}$. This yields

$$\begin{aligned}J_n &\leq 3\frac{1}{n}\sum_{j=1}^{n}|\hat{z}_{2,j} - \hat{z}_{2,j,\hat{L}_n}|^2 + |\hat{z}_{2,j,\hat{L}_n} - g_{1,n}(\hat{z}_{2,j,\hat{L}_n})|^2 + |g_{1,n}(\hat{z}_{2,j,\hat{L}_n}) - g_{1,n}(\hat{z}_{2,j})|^2 \\ &= 3 \cdot (J_{1,n} + J_{2,n} + J_{3,n}).\end{aligned}$$

Since $\hat{z}_{2,j,\hat{L}_n}$ is bounded by $\hat{L}_n$ we can employ the same argument as we used for type *ii*. Thus $J_{2,n} \to 0$ holds with probability one. As above we use that there exists a constant $C > 0$ such that $C\hat{K}_n^{\frac{3}{2}}$ is a Lipschitz constant for $g_{1,n}$. This yields $J_{3,n} \leq C^2\hat{K}_n^{\frac{9}{4}} \sum_{j=1}^{n}|\hat{z}_{2,j,\hat{L}_n} - \hat{z}_{2,j}|^2$. Using a similar argument as in (14) and the fact that $\frac{1}{n} \cdot \sum_{i=1}^{n}(\hat{z}_{2,j})^8 \leq \sqrt{\hat{K}_n}$ holds by our definition of $\hat{z}_{2,j}$, we see

$$\sum_{j=1}^{n}|\hat{z}_{2,j} - \hat{z}_{2,j,\hat{L}_n}|^2 \leq \sum_{j=1}^{n}(\hat{z}_{2,j})^2\frac{(\hat{z}_{2,j})^6}{(\hat{L}_n)^6}$$

$$
\begin{aligned}
&= \frac{1}{(\hat{L}_n)^6} \frac{1}{n} \sum_{j=1}^{n} (\hat{z}_{2,j})^8 \\
&\leq (\hat{K}_n)^{-2.5}.
\end{aligned}
$$

Using $\hat{K}_n \to \infty$ almost surely, this yields $J_{1,n} \to 0$ and $J_{3,n} \to 0$ almost surely.

It remains to show that expressions of type iv converge to 0 almost surely uniformly over $\mathcal{F}_n$. This follows directly from Lemma 2 if $g^2_{f,n}$ is contained in $\mathcal{G}_n$. Since the needed computation is a straightforward application of the relations between trigonometric functions such as

$$
\begin{aligned}
\sin(x)\sin(y) &= \frac{1}{2}\big(\cos(x-y) - \cos(x+y)\big), \\
\cos(x)\cos(y) &= \frac{1}{2}\big(\cos(x-y) + \cos(x+y)\big)
\end{aligned}
$$

and

$$
\sin(x)\cos(y) = \frac{1}{2}\big(\sin(x-y) + \sin(x+y)\big),
$$

we will not show every detail but just the important steps. For $f_n \in \mathcal{F}_n$, we have

$$
g^2_{f_n,n}(z_1, z_2) = f_n(z_1)^2 + g_{1,n}(z_2)^2 + 2f_n(z_1)g_{1,n}(z_2).
$$

With

$$
\begin{aligned}
f_n(z_1) \cdot g_{1,n}(z_2) \quad = \quad \sum_{k,j=0}^{\hat{K}_n} \Bigg(& a_k \tilde{a}_j \cos\left(k \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_1\right) \cos\left(j \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_2\right) \\
&+ b_k \tilde{a}_j \sin\left(k \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_1\right) \cos\left(j \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_2\right) \\
&+ a_k \tilde{b}_j \cos\left(k \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_1\right) \sin\left(j \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_2\right) \\
&+ b_k \tilde{b}_j \sin\left(k \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_1\right) \sin\left(j \cdot \frac{\pi}{\hat{L}^*_n} \cdot z_2\right) \Bigg).
\end{aligned}
$$

we see that $2f_n(z_1) \cdot g_{1,n}(z_2)$ is a trigonometric polynomial of degree $\hat{K}_n$ in two variables and all coefficients are bounded by $2\beta_n^2$. Furthermore, we have

$$
\begin{aligned}
f_n(z_1)^2 = \sum_{k,j=0}^{\hat{K}_n} & \frac{a_k a_j}{2} \left[\cos\left(\frac{(k-j)\pi}{\hat{L}^*_n} \cdot z_1\right) + \cos\left(\frac{(k+j)\pi}{\hat{L}^*_n} \cdot z_1\right)\right] \\
&+ \frac{b_k a_j}{2} \left[\sin\left(\frac{(k-j)\pi}{\hat{L}^*_n} \cdot z_1\right) + \sin\left(\frac{(k+j)\pi}{\hat{L}^*_n} \cdot z_1\right)\right]
\end{aligned}
$$

$$+\frac{a_k b_j}{2}\Big[\sin\left(\frac{(j-k)\pi}{\hat{L}_n^*}\cdot z_1\right)+\sin\left(\frac{(j+k)\pi}{\hat{L}_n^*}\cdot z_1\right)\Big]$$
$$+\frac{b_k b_j}{2}\Big[\cos\left(\frac{(k-j)\pi}{\hat{L}_n^*}\cdot z_1\right)-\cos\left(\frac{(k+j)\pi}{\hat{L}_n^*}\cdot z_1\right)\Big].$$

By reordering the above sum and conflating the appropriate summands we see that $f_n(z_1)^2$ (and analogously also $g_{1,n}(z_2)^2$) is a trigonometric polynomial of degree $2\cdot\hat{K}_n$ and all coefficients can be bounded by $(2\cdot\hat{K}_n+2)\cdot\beta_n^2$. Thus $g_{f_n,n}^2$ is a trigonometric polynomial of degree $2\cdot\hat{K}_n$ with coefficients bounded by

$$2(2\cdot\hat{K}_n+2)\beta_n^2+2\beta_n^2\leq 10\cdot\hat{K}_n\cdot\beta_n^2.$$

This shows $g_{f_n,n}^2\in\mathcal{G}_n$. Now we may apply Lemma 2 and obtain

$$T_{1,n}^{(k)}\to 0\quad a.s.\quad\text{for } k\in\{4,7\}.$$

This completes the proof. □

Lemma 4. *Let $(\hat{z}_1,\hat{z}_2)_1^n$ be defined as in Section 2. Define $\hat{K}_n$ depending on the data $(X_1,Y_1),\ldots,(X_n,Y_n)$ by (6). Set $\beta_n=n^{\frac{1}{11}}$, and $\hat{L}_n^*=2\hat{L}_n=2\sqrt{\hat{K}_n}$. Furthermore let the assumptions of Theorem 1 hold. Then*

$$\hat{K}_n^3\beta_n^2\sup_{|u|\leq 2\cdot\hat{K}_n,|v|\leq 2\cdot\hat{K}_n}\left|\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}\big(\frac{u\pi}{\hat{L}_n^*},\frac{v\pi}{\hat{L}_n^*}\big)-\varphi_{(Z_1,Z_2)}\big(\frac{u\pi}{\hat{L}_n^*},\frac{v\pi}{\hat{L}_n^*}\big)\right|\to 0$$

almost surely.

Before we prove Lemma 4 we will formulate and prove several auxiliary results.

Lemma 5. *Assume that $W^{(1)}$ and $V^{(1)}$ are real-valued random variables that are subgaussian and that $(W^{(1)},V^{(1)})$,$(W_1^{(1)},V_1^{(1)}),\ldots$ are independent and identically distributed. Then*

$$\frac{\sqrt{n}}{\log(n)}\sup_{u,v\in[-n,n]}\left|\varphi_{(W^{(1)},V^{(1)})}\big(u\pi,v\pi\big)-\varphi_{(W^{(1)},V^{(1)})_1^n}\big(u\pi,v\pi\big)\right|\to 0$$

almost surely.

Remark 5. Under the assumptions of Theorem 1, $Z_1,Z_2,X^{(1)},Y^{(1)},\epsilon_1$ and δ_1 are subgaussian. This yields

$$\frac{\sqrt{n}}{\log(n)}\sup_{u,v\in[-2\hat{K}_n,2\hat{K}_n]}\left|\varphi_{(Z_1,Z_2)}\big(\frac{u\pi}{\hat{L}_n^*},\frac{v\pi}{\hat{L}_n^*}\big)-\varphi_{(Z_1,Z_2)_1^n}\big(\frac{u\pi}{\hat{L}_n^*},\frac{v\pi}{\hat{L}_n^*}\big)\right|$$
$$\leq\frac{\sqrt{n}}{\log(n)}\sup_{u,v\in[-\hat{L}_n,\hat{L}_n]}\left|\varphi_{(Z_1,Z_2)}(u\pi,v\pi)-\varphi_{(Z_1,Z_2)_1^n}(u\pi,v\pi)\right|$$

$$\leq \frac{\sqrt{n}}{\log(n)} \sup_{u,v \in [-n,n]} \left| \varphi_{(Z_1,Z_2)}(u\pi, v\pi) - \varphi_{(Z_1,Z_2)_1^n}(u\pi, v\pi) \right|.$$

Here we have used $\hat{L}_n^* = 2\hat{L}_n = 2\sqrt{\hat{K}_n}$, the upper bound $\hat{K}_n \leq n^{\frac{1}{11}}$, and the fact that $u \in [-2\hat{K}_n, 2\hat{K}_n]$ implies $\frac{u}{\hat{L}_n^*} \in \left[\frac{-2\hat{K}_n}{\hat{L}_n^*}, \frac{2\hat{K}_n}{\hat{L}_n^*}\right]$. Consequently, Lemma 5 implies

$$\frac{\sqrt{n}}{\log(n)} \sup_{u,v \in [-2\hat{K}_n, 2\hat{K}_n]} \left| \varphi_{(Z_1,Z_2)}\left(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*}\right) - \varphi_{(Z_1,Z_2)_1^n}\left(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*}\right) \right| \to 0 \quad a.s.$$

Analogous results for $(X^{(1)}, Y^{(1)})$, ϵ_1 and δ_1 are obtained in the same way.

Proof of Lemma 5. As in the proof of Lemma 3 we use truncated versions of $W^{(1)}$ and $V^{(1)}$. However, we choose truncation heights that do not depend on the data. Set $M_n = \log(n)$. We obtain

$$\begin{aligned}
&\sup_{u,v \in [-n,n]} \left| \varphi_{(W^{(1)},V^{(1)})}(u\pi, v\pi) - \varphi_{(W^{(1)},V^{(1)})_1^n}(u\pi, v\pi) \right| \\
&\leq \sup_{u,v \in [-n,n]} \left| \varphi_{(W^{(1)},V^{(1)})}(u\pi, v\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u\pi, v\pi) \right| \\
&\quad + \sup_{u,v \in [-n,n]} \left| \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u\pi, v\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})_1^n}(u\pi, v\pi) \right| \\
&\quad + \sup_{u,v \in [-n,n]} \left| \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})_1^n}(u\pi, v\pi) - \varphi_{(W^{(1)},V^{(1)})_1^n}(u\pi, v\pi) \right| \\
&= T_{1,n} + T_{2,n} + T_{3,n}
\end{aligned}$$

Next, we show that $\frac{\sqrt{n}}{\log(n)} \cdot T_{1,n}$, $\frac{\sqrt{n}}{\log(n)} \cdot T_{2,n}$ and $\frac{\sqrt{n}}{\log(n)} \cdot T_{3,n}$ converge to zero almost surely. We use $\exp(ix) = \cos(x) + i\sin(x)$ together with the mean value theorem and obtain

$$\begin{aligned}
T_{1,n} &\leq \sup_{u,v \in [-n,n]} \mathbf{E}\Big\{ \left| \cos(W^{(1)}u\pi + V^{(1)}v\pi) - \cos(W^{(1)}_{M_n}u\pi + V^{(1)}_{M_n}v\pi) \right| \\
&\quad + \left| \sin(W^{(1)}u\pi + V^{(1)}v\pi) - \sin(W^{(1)}_{M_n}u\pi + V^{(1)}_{M_n}v\pi) \right| \Big\} \\
&\leq \sup_{u,v \in [-n,n]} \mathbf{E}\Big\{ \left| (W^{(1)}u + V^{(1)}v) - (W^{(1)}_{M_n}u + V^{(1)}_{M_n}v) \right| \cdot \pi \\
&\quad + \left| (W^{(1)}u + V^{(1)}v) - (W^{(1)}_{M_n}u + V^{(1)}_{M_n}v) \right| \cdot \pi \Big\} \\
&\leq 2\pi \sup_{u,v \in [-n,n]} \mathbf{E}\Big\{ |u(W^{(1)} - W^{(1)}_{M_n})| + |v(V^{(1)} - V^{(1)}_{M_n})| \Big\} \\
&\leq 2\pi n \cdot \Big[\mathbf{E}\big\{ |W^{(1)} - W^{(1)}_{M_n}| \big\} + \mathbf{E}\big\{ |(V^{(1)} - V^{(1)}_{M_n}| \big\} \Big].
\end{aligned}$$

Since $W^{(1)}$ and $V^{(1)}$ are subgaussian we apply the same argument as in (14). By the Cauchy-Schwarz inequality we have

$$\mathbf{E}\big\{ |W^{(1)} - W^{(1)}_{M_n}| \big\} \leq \sqrt{\mathbf{E}\big\{ |W^{(1)} - W^{(1)}_{M_n}|^2 \big\}}$$

and a similiar upper bound for $\mathbf{E}\{|(V^{(1)} - V^{(1)}_{M_n}|\}$. This yields

$$T_{1,n} \leq 2\pi n\Big[\sqrt{\frac{2}{c}\mathbf{E}\{\exp(c|W^{(1)}|^2)\} \cdot \exp(-\frac{c}{2}M_n^2)}$$
$$+ \sqrt{\frac{2}{c}\mathbf{E}\{\exp(c|V^{(1)}|^2)\} \cdot \exp(-\frac{c}{2}M_n^2)}\Big]$$
$$\leq C \cdot 2\pi n \cdot \exp(-\tilde{c}M_n^2).$$

By $M_n = \log(n)$ we have

$$\frac{\sqrt{n}}{\log(n)}T_{1,n} \leq \frac{2C\pi \cdot \sqrt{n} \cdot n}{\log(n)} \cdot \exp(-\tilde{c}\log(n)^2) \to 0$$

since for any $\tilde{c} > 0$, $\exp(-\tilde{c}\log(n)^2)$ tends to 0 faster than any polynomial in n.

To establish an upper bound for $T_{3,n}$, we use the same arguments as for $T_{1,n}$. With the Cauchy-Schwarz inequality we get

$$T_{3,n} \leq 2\pi n\Big[\frac{1}{n}\sum_{j=1}^{n}|W_j^{(1)} - W^{(1)}_{M_n,j}| + \frac{1}{n}\sum_{j=1}^{n}|V_j^{(1)} - V^{(1)}_{M_n,j}|\Big] \tag{15}$$
$$\leq 2\pi n \cdot \left[\sqrt{\frac{1}{n}\sum_{j=1}^{n}\frac{2}{c}\exp(c|W_j^{(1)}|^2)} + \sqrt{\frac{1}{n}\sum_{j=1}^{n}\frac{2}{c}\exp(c|V_j^{(1)}|^2)}\right] \cdot \exp(-\tilde{c}M_n^2)$$

for some $\tilde{c} > 0$. Now $W^{(1)}$ and $V^{(1)}$ are subgaussian and our sample consists of independent and identically distributed random copies of $W^{(1)}$ and $V^{(1)}$. The strong law of large numbers implies that the second factor converges with probability one and is thus bounded with probability one. This yields

$$\frac{\sqrt{n}}{\log(n)}T_3 \to 0$$

with probability one by the same argument as above.

To show the convergence of $T_{2,n}$ we will use a discretization argument. Thus we are interested in the Lipschitz continuity of $\varphi_{(W^{(1)}_{M_n}, V^{(1)}_{M_n})}(u,v)$ and $\varphi_{(W^{(1)}_{M_n}, V^{(1)}_{M_n})_1^n}(u,v)$ as functions of u and v. If f is a Lipschitz continuous function with Lipschitz constant l, then for any $c \in \mathbb{R}$ the function

$$x \mapsto f(cx) \qquad (c \in \mathbb{R})$$

is also Lipschitz continuous with Lipschitz constant $|c| \cdot l$. Furthermore the sum of two Lipschitz continuous functions with Lipschitz constants l_1 and l_2 is also Lipschitz continuous with Lipschitz constant $l_1 + l_2$. Since $\sin(x)$ and $\cos(x)$ are Lipschitz continuous with Lipschitz constant one and $W^{(1)}_{M_n}$, $V^{(1)}_{M_n}$ are bounded by M_n we obtain

$$\big|\varphi_{(W^{(1)}_{M_n}, V^{(1)}_{M_n})}(u,v) - \varphi_{(W^{(1)}_{M_n}, V^{(1)}_{M_n})}(u_0,v_0)\big|$$

$$\leq |\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u,v) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u,v_0)\big| + |\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u,v_0) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u_0,v_0)\big|$$
$$\leq 2M_n\big(|u-u_0| + |v-v_0|\big)$$

and similarly

$$\big|\varphi_{(X^{(1)}_{M_n},Y^{(1)}_{M_n})^n_1}(u,v) - \varphi_{(X^{(1)}_{M_n},Y^{(1)}_{M_n})^n_1}(u_0,v_0)\big| \leq 2M_n\big(|u-u_0| + |v-v_0|\big).$$

We use

$$I_j := [-n,n] \cap [j/n - 1/2n, j/n + 1/2n] \quad (j \in \mathbb{Z},\ |j| \leq n^2)$$

to partition $[-n,n]$ into intervals of length $\frac{1}{n}$ or less. By our previous observation on the Lipschitz constants, this yields

$$\sup_{u \in I_j, v \in I_k} \big|\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u\pi, v\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(\frac{j}{n}\pi, \frac{k}{n}\pi)\big| \leq \frac{2\pi M_n}{n}$$

as well as analogous upper bound for the discretization error of $\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})^n_1}$. Thus for any $C > 0$ there exists a $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$\frac{\sqrt{n}}{\log(n)} \sup_{u,v \in [-n,n]} \left|\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u\pi, v\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})^n_1}(u\pi, v\pi)\right| > C$$

implies the existence of a $(\frac{j}{n}, \frac{k}{n})$ with $|j|, |k| \leq n^2$ with

$$\frac{\sqrt{n}}{\log(n)} \left|\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(\frac{j}{n}\cdot\pi, \frac{k}{n}\cdot\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})^n_1}(\frac{j}{n}\cdot\pi, \frac{k}{n}\cdot\pi)\right| > \frac{C}{2}. \tag{16}$$

Splitting φ into real and imaginary parts and applying the Hoeffding inequality (cf., e.g., Lemma A.3 in Györfi et al. (2002)) yields

$$\begin{aligned}
&\mathbf{P}\left\{\left|\varphi_{(W^{(1)},V^{(1)})}(u,v) - \varphi_{(W^{(1)},V^{(1)})^n_1}(u,v)\right| > \epsilon\right\} \\
&\leq \mathbf{P}\left\{\big|Re\big(\varphi_{(W^{(1)},V^{(1)})}(u,v) - \varphi_{(W^{(1)},V^{(1)})^n_1}(u,v)\big)\big| > \frac{\epsilon}{\sqrt{2}}\right\} \\
&\quad + \mathbf{P}\left\{\big|Im\big(\varphi_{(W^{(1)},V^{(1)})}(u,v) - \varphi_{(W^{(1)},V^{(1)})^n_1}(u,v)\big)\big| > \frac{\epsilon}{\sqrt{2}}\right\} \\
&\leq 2 \cdot \left(2e^{\frac{-n\cdot\epsilon^2}{4}}\right) = 4e^{\frac{-n\cdot\epsilon^2}{4}}.
\end{aligned}$$

Combining the above inequality with (16), we obtain, for $n \geq n_0$ and $C > 0$

$$\mathbf{P}\left\{\frac{\sqrt{n}}{\log(n)} \sup_{u,v \in [-n,n]} \left|\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u\pi, v\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})^n_1}(u\pi, v\pi)\right| > C\right\}$$

$$\leq (2n^2+1)^2 \cdot \max_{\substack{j,k\in\mathbb{Z}\\ |j|,|k|\leq n^2}} \mathbf{P}\Big\{ \sup_{u,v\in[-n,n]} \Big| \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(\frac{j}{n}\cdot\pi, \frac{k}{n}\cdot\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})_1^n}(\frac{j}{n}\cdot\pi, \frac{k}{n}\cdot\pi)\Big| > \frac{C}{2}\cdot\frac{\log(n)}{\sqrt{n}}\Big\}$$

$$\leq 4\cdot(2n^2+1)^2\cdot e^{-n\cdot\left(\frac{C\log(n)}{2\sqrt{n}}\right)^2\cdot\frac{1}{4}} = 4\cdot(2n^2+1)^2\cdot e^{-C^2\log(n)^2/16}. \tag{17}$$

This yields

$$\sum_{n=1}^{\infty}\mathbf{P}\left\{\frac{\sqrt{n}}{\log(n)}\sup_{u,v\in[-n,n]}\left|\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u\pi,v\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})_1^n}(u\pi,v\pi)\right| > C\right\} < \infty$$

for any $C>0$. By the first Lemma of Borel-Cantelli

$$\frac{\sqrt{n}}{\log(n)}\sup_{u,v\in[-n,n]}\left|\varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})}(u\pi,v\pi) - \varphi_{(W^{(1)}_{M_n},V^{(1)}_{M_n})_1^n}(u\pi,v\pi)\right|$$

converges to 0 almost surely. Together with the previous results on $T_{1,n}$ and $T_{3,n}$ this implies the assertion. □

Lemma 6. *Assume that $V=(V^{(1)},V^{(2)},\ldots,V^{(d)})$ and $W=(W^{(1)},W^{(2)},\ldots,W^{(l)})$ are $\mathbb{R}^d$-valued and $\mathbb{R}^l$-valued random vectors, resp., such that $V^{(1)}$, $W^{(1)}$, $V^{(2)}$ and $W^{(2)}$ are subgaussian. Assume furthermore that (V,W),(V_1,W_1),… are independent and identically distributed. Then*

$$\frac{\sqrt{n}}{\log(n)^2}\sup_{u_1,v_1\in[-n,n]}\Big|\frac{\partial}{\partial u_2}\varphi_{(V,W)}(u_1\pi,0,\ldots,0,v_1\pi,0,\ldots,0) - \frac{\partial}{\partial u_2}\varphi_{(V,W)_1^n}(u_1\pi,0,\ldots,0,v_1\pi,0,\ldots,0)\Big| \to 0$$

almost surely. The same assertion holds if $\frac{\partial}{\partial u_2}$ is replaced by $\frac{\partial}{\partial v_2}$.

Remark 6. As for the previous lemma, a direct application of Lemma 6 yields

$$\frac{\sqrt{n}}{\log(n)^2}\sup_{u_1,v_1\in[-2\hat{K}_n,2\hat{K}_n]}\Big|\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(\frac{u_1\pi}{\hat{L}_n^*},0,\ldots,0,\frac{v_1\pi}{\hat{L}_n^*},0,\ldots,0) - \frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(\frac{u_1\pi}{\hat{L}_n^*},0,\ldots,0,\frac{v_1\pi}{\hat{L}_n^*},0,\ldots,0)\Big| \to 0 \quad a.s.$$

Similiar results are obtained for $\frac{\partial}{\partial v_2}\varphi_{(X,Y)}$, $\frac{\partial}{\partial z_1}\varphi_{(Z_1,Z_2)}$ and $\frac{\partial}{\partial z_2}\varphi_{(Z_1,Z_2)}$.

Proof of Lemma 6. It is well known that the characteristic function of a real-valued random variable is k-times differentiable if the absolute moments of order k are finite.

Since $V^{(2)}$ and $W^{(2)}$ are subgaussian, they are also square integrable. We obtain the partial derivative of $\varphi_{(V,W)}$ as

$$\begin{aligned}\frac{\partial}{\partial u_2}\varphi_{(V,W)}(u_1,0,\ldots,0,v_1,0,\ldots,0) &= \mathbf{E}\Big\{iV^{(2)}e^{i(u_1V^{(1)}+0\cdot V^{(2)}+\cdots+v_1W^{(1)}+\cdots+0)}\Big\}\\ &= \mathbf{E}\Big\{iV^{(2)}e^{i(u_1V^{(1)}+v_1W^{(1)})}\Big\}.\end{aligned}$$

We also compute the derivative of the empirical variant of the characteristic function

$$\frac{\partial}{\partial u_2}\varphi_{(V,W)_1^n}(u_1,0,\ldots,0,v_1,0,\ldots,0) = \frac{1}{n}\sum_{j=1}^n iV_j^{(2)}e^{i(u_1V_j^{(1)}+v_1W_j^{(1)})}.$$

Now we proceed as in Lemma 5. With $M_n := \log(n)$ we have

$$\begin{aligned}&\sup_{u,v\in[-n,n]}\Big|\mathbf{E}\Big\{iV^{(2)}e^{i(u\pi V^{(1)}+v\pi W^{(1)})}\Big\} - \frac{1}{n}\sum_{j=1}^n iV_j^{(2)}e^{i(u\pi V_j^{(1)}+v\pi W_j^{(1)})}\Big|\\ &\leq \sup_{u,v\in[-n,n]}\Big|\mathbf{E}\Big\{iV^{(2)}e^{i\pi(uV^{(1)}+vW^{(1)})}\Big\} - \mathbf{E}\Big\{iV_{M_n}^{(2)}e^{i\pi(uV_{M_n}^{(1)}+vW_{M_n}^{(1)})}\Big\}\Big|\\ &\quad + \sup_{u,v\in[-n,n]}\Big|\mathbf{E}\Big\{iV_{M_n}^{(2)}e^{i\pi(uV_{M_n}^{(1)}+vW_{M_n}^{(1)})}\Big\} - \frac{1}{n}\sum_{j=1}^n iV_{M_n,j}^{(2)}e^{i\pi(uV_{M_n,j}^{(1)}+vW_{M_n,j}^{(1)})}\Big|\\ &\quad + \sup_{u,v\in[-n,n]}\Big|\frac{1}{n}\sum_{j=1}^n iV_{M_n,j}^{(2)}e^{i\pi(uV_{M_n,j}^{(1)}+vW_{M_n,j}^{(1)})} - \frac{1}{n}\sum_{j=1}^n iV_j^{(2)}e^{i\pi(uV_j^{(1)}+vW_j^{(1)})}\Big|.\\ &= T_{1,n}+T_{2,n}+T_{3,n}.\end{aligned}$$

Using the elementary inequality

$$|ab-cd| \leq |a-c|\cdot|b| + |c|\cdot|b-d| \quad (a,b,c,d\in\mathbb{R})$$

we obtain the desired rate of convergence for $T_{1,n}$ and $T_{3,n}$ as in Lemma 5. The upper bound for $T_{2,n}$ is also obtained analogously to Lemma 5. We have the Lipschitz estimate

$$\begin{aligned}&|iX_{M_n}^{(2)}e^{i\pi(uX_{M_n}^{(1)}+vY_{M_n}^{(1)})} - iX_{M_n}^{(2)}e^{i\pi(u_0X_{M_n}^{(1)}+v_0Y_{M_n}^{(1)})}|\\ &\leq 2\pi(M_n)^2(|u-u_0|+|v-v_0|).\end{aligned}$$

Thus the discretization error that arises over a partion with intervals of length $\frac{1}{n}$ is bounded by some constant times $\frac{(M_n)^2}{n} = \frac{\log(n)^2}{n}$. Now, for n large enough,

$$\frac{\sqrt{n}}{\log(n)^2}\sup_{u,v\in[-n,n]}\Big|\mathbf{E}\Big\{iV_{M_n}^{(2)}e^{i\pi(uV_{M_n}^{(1)}+vW_{M_n}^{(1)})}\Big\} - \frac{1}{n}\sum_{j=1}^n iV_{M_n,j}^{(2)}e^{i\pi(uV_{M_n,j}^{(1)}+vW_{M_n,j}^{(1)})}\Big| > C$$

implies the existence of a $(\frac{j}{n},\frac{k}{n})$ with $|j|,|k|\leq n^2$ with

$$\frac{\sqrt{n}}{\log(n)^2}\Big|\mathbf{E}\Big\{iV_{M_n}^{(2)}e^{i\pi(\frac{j}{n}V_{M_n}^{(1)}+\frac{k}{n}W_{M_n}^{(1)})}\Big\} - \frac{1}{n}\sum_{j=1}^n iV_{M_n,j}^{(2)}e^{i\pi(\frac{j}{n}V_{M_n,j}^{(1)}+\frac{k}{n}W_{M_n,j}^{(1)})}\Big| > \frac{C}{2}.$$

Following the argument in the proof of the previous lemma, we apply the Hoeffding inequality to obtain

$$\mathbf{P}\Big\{\frac{\sqrt{n}}{\log(n)^2}\sup_{u,v\in[-n,n]}\Big|\mathbf{E}\Big\{iX^{(2)}_{M_n}e^{i(u_nX^{(1)}_{M_n}+v_nY^{(1)}_{M_n})}\Big\}-\frac{1}{n}\sum_{j=1}^{n}iX^{(2)}_{M_n,j}e^{i(u_nX^{(1)}_{M_n,j}+v_nY^{(1)}_{M_n,j})}\Big|\geq C\Big\}$$
$$\leq 4\cdot(2n^2+1)^2\cdot\exp\Big(-n\big(\frac{C\cdot\log(n)^2}{\sqrt{n}}\big)^2\cdot\frac{1}{16(M_n)^2}\Big)$$
$$\leq Cn^3\cdot\exp\big(\frac{-\log(n)^2}{16}\big).$$

Again, the right hand side of the above inequality is summable. Thus the first assertion follows as in the proof of Lemma 5. The second assertion follows in the same way. □

Lemma 7. *Define $\hat{K}_n$ as in (6). If $X^{(1)}$ and $Y^{(1)}$ are subgaussian and $\varphi_{(X^{(1)},Y^{(1)})}$ vanishes nowhere, we have*

$$\hat{K}_n\to\infty$$

almost surely.

Proof. It suffices to show that

$$\max\{K\in\mathbb{N}:\inf_{u,v\in[-2\cdot\sqrt{K},2\cdot\sqrt{K}]}\big|\varphi_{(X^{(1)},Y^{(1)})_1^n}(u\pi,v\pi)\big|\geq\frac{1}{\log(n)}\}$$

tends to infinity almost surely. We use Lemma 5. Since $X^{(1)}$ and $Y^{(1)}$ are subgaussian we have

$$\frac{\sqrt{n}}{\log(n)}\sup_{u,v\in[-n,n]}\Big|\varphi_{(X^{(1)},Y^{(1)})}\big(u\pi,v\pi\big)-\varphi_{(X^{(1)},Y^{(1)})_1^n}\big(u\pi,v\pi\big)\Big|\to 0$$

almost surely. By assumption $\varphi_{(X_1,Y_1)}$ is continuous and vanishes nowhere. Thus for every interval I there exist a constant C_I such that

$$\varphi_{(X_1,Y_1)}(u,v)>C_I$$

holds for all $(u,v)\in I^2$. Thous almost surely for n large enough, we have

$$\frac{1}{\log n}<\frac{C_I}{2}\quad\text{and}\quad\sup_{u,v\in I}|\varphi_{(X_1,Y_1)_1^n}(u,v)-\varphi_{(X_1,Y_1)}(u,v)|\leq\frac{C_I}{2}.$$

This implies that with probability one the inequality $\min_{u,v\in I}\varphi_{(X^{(1)},Y^{(1)})_1^n}(u,v)\geq\frac{1}{\log(n)}$ holds on I for n large enough. Since I was arbitrary this completes the proof. □

Lemma 8. *Assume $a_3\neq 0$ and $b_3\neq 0$ and that $X^{(1)},\ldots,Y^{(l)}$ are subgaussian. Furthermore assume $\mathbf{E}Z_1^2>0$, $\mathbf{E}Z_2^2>0$ and choose the parameters $\hat{K}_n$ and $\hat{L}_n^*$ as in Theorem 1. Then*

$$\frac{\sqrt{n}}{\log(n)^2}|\hat{a}_2-a_2|\to 0\quad\textit{and}\quad\frac{\sqrt{n}}{\log(n)^2}|\hat{b}_2-b_2|\to 0$$

almost surely.

Proof. Since $X^{(1)}, X^{(2)}$ and $X^{(3)}$ are subgaussian, the products $X^{(1)} \cdot X^{(2)}$ and $X^{(1)} \cdot X^{(3)}$ are square integrable. By the law of the iterated logarithm (c.f. i.e. Theorem 33.1 in Bauer (1996)) we conclude that with probability one

$$\frac{\sqrt{n}}{\sqrt{2\log(\log(n))}} \Big| \frac{1}{n} \sum_{i=1}^{n} X_i^{(1)} \cdot X_i^{(3)} - \mathbf{E}\{X^{(1)} \cdot X^{(3)}\} \Big|$$

and

$$\frac{\sqrt{n}}{\sqrt{2\log(\log(n))}} \Big| \frac{1}{n} \sum_{i=1}^{n} X_i^{(2)} \cdot X_i^{(3)} - \mathbf{E}\{X^{(2)} \cdot X^{(3)}\} \Big|$$

are bounded. This implies

$$\frac{\sqrt{n}}{\log(n)^2} \Big| \frac{1}{n} \sum_{i=1}^{n} X_i^{(1)} \cdot X_i^{(3)} - \mathbf{E}\{X^{(1)} \cdot X^{(3)}\} \Big| \to 0$$

and

$$\frac{\sqrt{n}}{\log(n)^2} \Big| \frac{1}{n} \sum_{i=1}^{n} X_i^{(2)} \cdot X_i^{(3)} - \mathbf{E}\{X^{(2)} \cdot X^{(3)}\} \Big| \to 0$$

almost surely.

Recall that $\hat{a}$ is defined as

$$\hat{a}_2 = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^{(2)} \cdot X_i^{(3)}}{\frac{1}{n}\sum_{i=1}^{n} X_i^{(1)} \cdot X_i^{(3)}}$$

and that

$$a_2 = \frac{\mathbf{E}\{X^{(2)} \cdot X^{(3)}\}}{\mathbf{E}\{X^{(1)} \cdot X^{(3)}\}}$$

holds. Now, by using the elementary inequality

$$\left| \frac{x_n}{y_n} - \frac{x}{y} \right| \leq \frac{|x||y_n - y|}{|yy_n|} + \frac{|y||x_n - x|}{|yy_n|}$$

we obtain

$$\begin{aligned}
&\frac{\sqrt{n}}{\log(n)^2} |\hat{a}_2 - a_2| \\
&\leq \frac{|\mathbf{E}\{X^{(2)} \cdot X^{(3)}\}| \cdot \frac{\sqrt{n}}{\log(n)^2} |\frac{1}{n}\sum_{i=1}^{n} X_i^{(1)} \cdot X_i^{(3)} - \mathbf{E}\{X_i^{(1)} \cdot X_i^{(3)}\}|}{|\frac{1}{n}\sum_{i=1}^{n} X_i^{(1)} \cdot X_i^{(3)} \cdot \mathbf{E}\{X_i^{(1)} \cdot X_i^{(3)}\}|} \\
&\quad + \frac{|\mathbf{E}\{X^{(1)} \cdot X^{(3)}\}| \cdot \frac{\sqrt{n}}{\log(n)^2} |\frac{1}{n}\sum_{i=1}^{n} X_i^{(2)} \cdot X_i^{(3)} - \mathbf{E}\{X_i^{(2)} \cdot X_i^{(3)}\}|}{|\frac{1}{n}\sum_{i=1}^{n} X_i^{(1)} \cdot X_i^{(3)} \cdot \mathbf{E}\{X_i^{(1)} \cdot X_i^{(3)}\}|}.
\end{aligned}$$

For both of these fractions, the nominator tends towards $\mathbf{E}\{X^{(1)} \cdot X^{(3)}\}^2 = a_3^2 \neq 0$ almost surely and the second factor of the denominator converges to 0 almost surely. This yields

$$\frac{\sqrt{n}}{\log(n)^2} |\hat{a}_2 - a_2| \to 0 \quad a.s.$$

The second assertion follows in the same way. □

Lemma 9. *Assume that assumptions A1) to A6) on our model* (5) *are fulfilled. Assume furthermore that Z_1 and Z_2 as well as $\epsilon_1, \ldots, \delta_l$ are subgaussian. Define T_n as in* (7)*. Then*

$$\frac{\sqrt{n}}{\log(n)^2} T_n \to 0$$

almost surely.

Proof. Let us recall the definition of T_n. We have

$$\begin{aligned}
&T_n(\hat{z}_1, \hat{z}_2) := \\
&\sup_{u_1, v_1 \in [-2\hat{K}_n, 2\hat{K}_n]} \Bigg\{ \Big| \varphi_{(X^{(1)}, Y^{(1)})_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big) - \varphi_{(\hat{z}_1, \hat{z}_2)_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}\big(\frac{\pi u_1}{\hat{L}_n^*}\big) \cdot \varphi_{\hat{\delta}_{1,1}^n}\big(\frac{\pi v_1}{\hat{L}_n^*}\big) \Big| \\
&+ \Big| \frac{\partial}{\partial u_2} \varphi_{(X,Y)_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, 0, \ldots, 0, \frac{\pi v_1}{\hat{L}_n^*}, 0, \ldots, 0\big) \\
&\qquad - \hat{a}_2 \cdot \frac{\partial}{\partial u_1} \varphi_{(\hat{z}_1, \hat{z}_2)_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}\big(\frac{\pi u_1}{\hat{L}_n^*}\big) \cdot \varphi_{\hat{\delta}_{1,1}^n}\big(\frac{\pi v_1}{\hat{L}_n^*}\big) \Big| \\
&+ \Big| \frac{\partial}{\partial v_2} \varphi_{(X,Y)_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, 0, \ldots, 0, \frac{\pi v_1}{\hat{L}_n^*}, 0, \ldots, 0\big) \\
&\qquad - \hat{b}_2 \cdot \frac{\partial}{\partial v_1} \varphi_{(\hat{z}_1, \hat{z}_2)_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}\big(\frac{\pi u_1}{\hat{L}_n^*}\big) \cdot \varphi_{\hat{\delta}_{1,1}^n}\big(\frac{\pi v_1}{\hat{L}_n^*}\big) \Big| \Bigg\}.
\end{aligned}$$

Here $\hat{z}_1$ and $\hat{z}_2$ are chosen in such a way that they minimize T_n under the constraint $\frac{1}{n}\sum_{j=1}^n \hat{z}_{2,j}^8 \leq \sqrt{\hat{K}_n}$. We define $\tilde{T}_n$ by replacing $\hat{z}_{1,j}$ with $Z_{1,j}$, replacing $\hat{z}_{2,j}$ with $Z_{2,j}$, replacing $\hat{\epsilon}_{1,j} = X_j^{(1)} - \hat{z}_{1,j}$ with $\tilde{\epsilon}_{1,j} := X_j^{(1)} - Z_{1,j}$ and replacing $\hat{\delta}_{1,j} = Y_j^{(1)} - \hat{z}_{2,j}$ with $\tilde{\delta}_{1,j} := Y_j^{(1)} - Z_{2,j}$. Since Z_2 is subgaussian, $\mathbf{E}Z_2^8 < \infty$ and $\frac{1}{n}\sum_{j=1}^n \hat{Z}_{2,j}^8$ converges to $\mathbf{E}Z_2^8$ almost surely. By Lemma 7, $\hat{K}_n$ tends to ∞ almost surely, thus with probability one

$$T_n \leq \tilde{T}_n$$

holds for n large enough. Using the relations between the characteristic functions of (X, Y) and (Z_1, Z_2) from the first part of Lemma 1, we have

$$\begin{aligned}
\tilde{T}_n \leq &\sup_{u_1, v_1 \in [-2\hat{K}_n, 2\hat{K}_n]} \Big\{ |\varphi_{(X^{(1)}, Y^{(1)})_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big) - \varphi_{(X^{(1)}, Y^{(1)})}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big)| \\
&+ |\varphi_{(Z_1, Z_2)}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big) \cdot \varphi_{\epsilon_1}\big(\frac{\pi u_1}{\hat{L}_n^*}\big) \cdot \varphi_{\delta_1}\big(\frac{\pi v_1}{\hat{L}_n^*}\big) \\
&\qquad - \varphi_{(Z_1, Z_2)_1^n}\big(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}\big) \cdot \varphi_{\epsilon_{1,1}^n}\big(\frac{\pi u_1}{\hat{L}_n^*}\big) \cdot \varphi_{\delta_{1,1}^n}\big(\frac{\pi v_1}{\hat{L}_n^*}\big)|
\end{aligned}$$

$$+ |\frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, 0, \ldots, \frac{\pi v_1}{\hat{L}_n^*}, 0, \ldots) - \frac{\partial}{\partial u_2}\varphi_{(X,Y)}(\frac{\pi u_1}{\hat{L}_n^*}, 0, \ldots, \frac{\pi v_1}{\hat{L}_n^*}, 0, \ldots)|$$
$$+ |a_2\frac{\partial}{\partial u}\varphi_{(Z_1,Z_2)}(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}) \cdot \varphi_{\epsilon_1}(\frac{\pi u_1}{\hat{L}_n^*}) \cdot \varphi_{\delta_1}(\frac{\pi v_1}{\hat{L}_n^*})$$
$$- \hat{a}_2\frac{\partial}{\partial u}\varphi_{(Z_1,Z_2)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}) \cdot \varphi_{\epsilon_{1,1}^n}(\frac{\pi u_1}{\hat{L}_n^*}) \cdot \varphi_{\delta_{1,1}^n}(\frac{\pi v_1}{\hat{L}_n^*})|$$
$$+ |\frac{\partial}{\partial v_2}\varphi_{(X,Y)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, 0, \ldots, \frac{\pi v_1}{\hat{L}_n^*}, 0, \ldots) - \frac{\partial}{\partial v_2}\varphi_{(X,Y)}(\frac{\pi u_1}{\hat{L}_n^*}, 0, \ldots, \frac{\pi v_1}{\hat{L}_n^*}, 0, \ldots)|$$
$$+ |b_2\frac{\partial}{\partial v}\varphi_{(Z_1,Z_2)}(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}) \cdot \varphi_{\epsilon_1}(\frac{\pi u_1}{\hat{L}_n^*}) \cdot \varphi_{\delta_1}(\frac{\pi v_1}{\hat{L}_n^*})$$
$$- \hat{b}_2\frac{\partial}{\partial v}\varphi_{(Z_1,Z_2)_1^n}(\frac{\pi u_1}{\hat{L}_n^*}, \frac{\pi v_1}{\hat{L}_n^*}) \cdot \varphi_{\epsilon_{1,1}^n}(\frac{\pi u_1}{\hat{L}_n^*}) \cdot \varphi_{\delta_{1,1}^n}(\frac{\pi v_1}{\hat{L}_n^*})|\Big\}$$

Applying Lemmas 5 to 8 to the above expressions yields

$$\frac{\sqrt{n}}{\log(n)^2}T_n \to 0$$

almost surely. This completes the proof. □

Proof of Lemma 4. As a first step, let us verify that the conditions of Lemmas $5-9$ are satisfied under the assumptions of Lemma 4. Under the assumptions of Theorem 1 all the occurring random variables are subgaussian. Thus the assumptions of Lemmas 5, 6 and 8 are satisfied. Under the assumptions of Theorem 1 the characteristic function of $(X^{(1)}, Y^{(1)})$ vanishes nowhere. This shows that the assumptions of Lemma 7 are satisfied. Additionally the assumptions concerning our model that are needed in Lemma 9 are also part of the assumptions of Theorem 1. Thus Lemmas $5-9$ are applicable under the assumptions of Lemma 4.

By Lemma 9 almost surely

$$\sup_{u,v\in[-2\hat{K}_n,2\hat{K}_n]} \left|\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*}) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}(\frac{u\pi}{\hat{L}_n^*}) \cdot \varphi_{\hat{\delta}_{1,1}^n}(\frac{v\pi}{\hat{L}_n^*}) - \varphi_{(X^1,Y^1)_1^n}(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*})\right|$$
$$\leq \frac{\log(n)^2}{\sqrt{n}} \leq \frac{1}{2\log(n)}$$

for n large enough. On the other hand

$$\inf_{|u|\leq 2\hat{K}_n, |v|\leq 2\hat{K}_n} |\varphi_{(X^1,Y^1)_1^n}(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*})| \geq \frac{1}{\log(n)}$$

holds by our definition of $\hat{K}_n$ in (6). Thus almost surely

$$\inf_{u,v\in[-2\hat{K}_n,2\hat{K}_n]} |\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*}) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}(\frac{u\pi}{\hat{L}_n^*}) \cdot \varphi_{\hat{\delta}_{1,1}^n}(\frac{v\pi}{\hat{L}_n^*})| \geq \frac{1}{2\log(n)}$$

for n large enough. This implies that is suffices to show

$$(\hat{K}_n)^3\beta_n^2 \sup_{|u|\leq 2\hat{K}_n, |v|\leq 2\hat{K}_n} \left| \varphi_{(\hat{z}_1,\hat{z}_2)_1^n}\Big(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*}\Big) - \varphi_{(Z_1,Z_2)}\Big(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*}\Big) \right| \to 0 \quad (a.s.)$$

under the additional assumption that $|\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(\frac{u\pi}{\hat{L}_n^*}, \frac{v\pi}{\hat{L}_n^*}) \cdot \varphi_{\hat{\epsilon}_{1,1}^n}(\frac{u\pi}{\hat{L}_n^*}) \cdot \varphi_{\hat{\delta}_{1,1}^n}(\frac{v\pi}{\hat{L}_n^*})|$ is bounded away from zero by $\frac{1}{2\log(n)}$ on $[-2\hat{K}_n, 2\hat{K}_n]^2$.

As in the previous proofs, we will use

$$u \in [-2\hat{K}_n, 2\hat{K}_n] \Leftrightarrow \frac{u\pi}{\hat{L}_n^*} \in \left[-\pi\sqrt{\hat{K}_n}, \pi\sqrt{\hat{K}_n}\right]$$

to simplify the notation. We will also need some auxiliary result from complex analysis. It is easy to see that the functions

$$\varphi_{v,n}(u):\ u \mapsto \varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,v) \quad \text{and} \quad \varphi_{u,n}(v):\ v \mapsto \varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,v)$$

are continuously differentiable for any fixed $v \in \mathbb{R}$ or $u \in \mathbb{R}$ respectively. Furthermore, it is a well know fact from probability theory (c.f. e.g. Theorem 25.2 in Bauer (1996)) that the characteristic function of a random variable is continuously differentiable if moments of the appropriate order exist. Since Z_1 and Z_2 are subgaussian and thus square integrable, we see that the functions

$$\varphi_v(u):\ u \mapsto \varphi_{(Z_1,Z_2)}(u,v) \quad \text{and} \quad \varphi_u(v): v \mapsto \varphi_{(Z_1,Z_2)}(u,v)$$

are also continuously differentiable for fixed v and u respectively. If $f : \mathbb{R} \to \mathbb{C}$ is continuously differentiable and does not vanish on an interval $I \subset \mathbb{R}$ we define $L_f(t)$ for any $t \in I$ by

$$L_f(t) := w_0 + \int_{t_0}^t \frac{f'(\tau)}{f(\tau)}\, d\tau$$

where w_0 is a complex number such that $\exp(w_0) = f(t_0)$. Now $L_f(t)$ defines a logarithm of $f(t)$, i.e. $\exp(L_f(t)) = f(t)$ (cf., e.g., Section III, §6 in Lang (1999)). The above functions do not vanish on $[-\pi\sqrt{\hat{K}_n}, \pi\sqrt{\hat{K}_n}]^2$. Thus

$$w_0 = 0 + \int_0^u \frac{\frac{\partial}{\partial u}\varphi_{(Z_1,Z_2)}(z,0)}{\varphi_{(Z_1,Z_2)}(z,0)}\, dz$$

is a logarithm of $\varphi_{(Z_1,Z_2)}(u,0)$ by $\ln(\varphi_{(Z_1,Z_2)}(0,0)) = \ln(1) = 0$. Thus

$$\begin{aligned} L_{\varphi_u}(v) &= \int_0^u \frac{\frac{\partial}{\partial u}\varphi_{(Z_1,Z_2)}(z,0)}{\varphi_{(Z_1,Z_2)}(z,0)}\, dz + \int_0^v \frac{\frac{\partial}{\partial v}\varphi_{(Z_1,Z_2)}(u,z)}{\varphi_{(Z_1,Z_2)}(u,z)}\, dz \\ &= w_0 + \int_0^v \frac{\frac{\partial}{\partial v}\varphi_{(Z_1,Z_2)}(u,z)}{\varphi_{(Z_1,Z_2)}(u,z)}\, dz \end{aligned}$$

statisfies $\exp(L_{\varphi_u}(v)) = \varphi_{(Z_1,Z_2)}(u,v)$ for $(u,v) \in [-2\hat{K}_n, 2\hat{K}_n]^2$. An analogous expression is obtained for a logarithm $L_{\varphi_{u,n}}(v)$ of $\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,v)$. To simplify notation in the next step, let us denote $L_{\varphi_{u,n}}(v)$ as b_n and $L_{\varphi_u}(v)$ as b. This yields

$$\begin{aligned}
&\left|\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,v) - \varphi_{(Z_1,Z_2)}(u,v)\right| = \left|\exp\bigl(L_{\varphi_u}(v)\bigr) - \exp\bigl(L_{\varphi_{u,n}}(v)\bigr)\right| \\
&= \left|\exp(b_n)\right| \cdot \left|\exp(b - b_n) - 1\right| \\
&= \left|\exp(b_n)\right| \cdot \left|\exp(Re(b - b_n)) \cdot \exp(i \cdot Im(b - b_n)) - 1\right| \\
&= \left|\exp(b_n)\right| \cdot \left|\bigl(\exp(Re(b - b_n)) - 1\bigr) \cdot \exp(i \cdot Im(b - b_n)) + \exp(i \cdot Im(b - b_n)) - 1\right|.
\end{aligned}$$

We have

$$|\exp(b_n)| = |\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,v)| \leq 1.$$

Therefore, with $\exp(0) = 1$ and $\xi \in \mathbb{R}$ between 0 and $Re(b - b_n)$, using the mean value theorem shows that the right hand side is bounded by

$$\begin{aligned}
&\left|\exp(\xi) \cdot |Re(b - b_n)|\right| \cdot 1 + \left|\cos\bigl(Im(b - b_n)\bigr) + i \sin\bigl(Im(b - b_n)\bigr) - 1\right| \\
&\quad \leq |\exp(|Re(b - b_n)|)| \cdot |b - b_n| + 2|b - b_n|.
\end{aligned}$$

Here we have used the monotonicity of the exponential function and the Lipschitz continuity of sin and cos. If b_n converges to b almost surely,

$$\exp(|Re(b - b_n)|) \to 1 \quad a.s.$$

Combining this with the above upper bound for $\left|\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,v) - \varphi_{(Z_1,Z_2)}(u,v)\right|$, we see that

$$(\hat{K}_n)^3 \beta_n^2 \sup_{|u| \leq \pi\sqrt{\hat{K}_n}, |v| \leq \pi\sqrt{\hat{K}_n}} |L_{\varphi_{u,n}}(v) - L_{\varphi_u}(v)| \to 0 \quad a.s. \tag{18}$$

implies the assertion of the lemma.

Let us denote the vectors $(u, 0, \ldots, 0, s, 0, \ldots, 0)$ and $(t, 0, \ldots, 0, 0, 0, \ldots, 0)$ by (u_0, s_0) and $(t_0, 0)$ respectively. Using again the relations between the characteristic functions of (X, Y), Z_1, Z_2 and their derivatives from Lemma 1 yields

$$\begin{aligned}
|b - b_n| &= \Bigl| \int_0^u \frac{\frac{\partial}{\partial u}\varphi_{(Z_1,Z_2)}(t,0)}{\varphi_{(Z_1,Z_2)}(t,0)}\, dt + \int_0^v \frac{\frac{\partial}{\partial v}\varphi_{(Z_1,Z_2)}(u,s)}{\varphi_{(Z_1,Z_2)}(u,s)}\, ds \\
&\quad - \int_0^u \frac{\frac{\partial}{\partial u}\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(t,0)}{\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(t,0)}\, dt - \int_0^v \frac{\frac{\partial}{\partial v}\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,s)}{\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,s)}\, ds \Bigr| \\
&= \Bigl| \int_0^u \frac{1}{a_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(t_0,0)}{\varphi_{(X,Y)}(t_0,0)} dt + \int_0^v \frac{1}{b_2} \cdot \frac{\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u_0,s_0)}{\varphi_{(X,Y)}(u_0,s_0)} ds \\
&\quad - \int_0^u \frac{\frac{\partial}{\partial u}\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(t,0)}{\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(t,0)}\, dt - \int_0^v \frac{\frac{\partial}{\partial v}\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,s)}{\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,s)}\, ds \Bigr|
\end{aligned}$$

$$
\begin{aligned}
&\leq \Big| \int_0^u \frac{1}{a_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(t_0,0)}{\varphi_{(X,Y)}(t_0,0)} dt - \int_0^u \frac{1}{\hat{a}_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(t_0,0)}{\varphi_{(X,Y)_1^n}(t_0,0)} dt \Big| \\
&\quad + \Big| \int_0^u \frac{1}{\hat{a}_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(t_0,0)}{\varphi_{(X,Y)_1^n}(t_0,0)} dt - \int_0^u \frac{\frac{\partial}{\partial u_2}\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(t,0)}{\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(t,0)} dt \Big| \\
&\quad + \Big| \int_0^v \frac{1}{b_2} \cdot \frac{\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u_0,s_0)}{\varphi_{(X,Y)}(u_0,s_0)} ds - \int_0^v \frac{1}{\hat{b}_2} \cdot \frac{\frac{\partial}{\partial v_2}\varphi_{(X,Y)_1^n}(u_0,s_0)}{\varphi_{(X,Y)_1^n}(u_0,s_0)} ds \Big| \\
&\quad + \Big| \int_0^v \frac{1}{\hat{b}_2} \cdot \frac{\frac{\partial}{\partial v_2}\varphi_{(X,Y)_1^n}(u_0,s_0)}{\varphi_{(X,Y)_1^n}(u_0,s_0)} ds - \int_0^v \frac{\frac{\partial}{\partial v_2}\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,s)}{\varphi_{(\hat{z}_1,\hat{z}_2)_1^n}(u,s)} ds \Big| \\
&= S_{1,n} + S_{2,n} + S_{3,n} + S_{4,n}.
\end{aligned}
$$

We have

$$
\sup_{|u|,|v| \leq \pi\sqrt{\hat{K}_n}} S_{1,n} \leq \pi\sqrt{\hat{K}_n} \sup_{|u|,|v| \leq \pi\sqrt{\hat{K}_n}} \Big| \frac{1}{a_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(u_0,v_0)}{\varphi_{(X,Y)}(u_0,v_0)} - \frac{1}{\hat{a}_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(u_0,v_0)}{\varphi_{(X,Y)_1^n}(u_0,v_0)} \Big|.
$$

For two sequences $a_n \in \mathbb{R}$, $b_n \in \mathbb{R} \setminus \{0\}$ with limits $a \in \mathbb{R}$ and $b \in \mathbb{R} \setminus \{0\}$, the elementary inequality

$$
\left| \frac{a}{b} - \frac{a_n}{b_n} \right| \leq \left| \frac{(a-a_n)b}{bb_n} \right| + \left| \frac{a(b-b_n)}{bb_n} \right|
$$

holds. Thus we obtain

$$
\begin{aligned}
&\left| \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(u_0,v_0)}{a_2\varphi_{(X,Y)}(u_0,v_0)} - \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(u_0,v_0)}{\hat{a}_2\varphi_{(X,Y)_1^n}(u_0,v_0)} \right| \\
&\leq \left| \frac{\left(\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(u_0,v_0) - \frac{\partial}{\partial u_2}\varphi_{(X,Y)_1^n}(u_0,v_0)\right) \cdot a_2\varphi_{(X,Y)}(u_0,v_0)}{a_2\varphi_{(X,Y)}(u_0,v_0) \cdot \hat{a}_2\varphi_{(X,Y)_1^n}(u_0,v_0)} \right| \\
&\quad + \left| \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(u_0,v_0) \cdot \left(a_2\varphi_{(X,Y)}(u_0,v_0) - \hat{a}_2\varphi_{(X,Y)_1^n}(u_0,v_0)\right)}{a_2\varphi_{(X,Y)}(u_0,v_0) \cdot \hat{a}_2\varphi_{(X,Y)_1^n}(u_0,v_0)} \right| \\
&= S_{1,n}^1 + S_{1,n}^2.
\end{aligned}
$$

By the definition of $\hat{K}_n$, we have $|\varphi_{(X,Y)_1^n}(u_0,s_0)| \geq \frac{1}{\log(n)}$ on $[-\pi\sqrt{\hat{K}_n}, \pi\sqrt{\hat{K}_n}]^2$. Now $\varphi_{(X,Y)_1^n}(u_0,s_0)$ converges to $\varphi_{(X,Y)}(u_0,s_0)$ with rate $\frac{\log(n)}{\sqrt{n}}$ almost surely by Lemma 5 and $\hat{a}_2$ converges almost surely to $a_2 \neq 0$ by Lemma 8. Thus there exists a $C > 0$ depending only on a such that with probability one the denominators are bounded away from zero by $C\frac{1}{\log(n)^2}$ for n sufficiently large. The absolute value of the characteristic function is bounded by one. Thus, by Lemma 6

$$
\frac{\sqrt{n}}{\log(n)^4} \sup_{|u|,|v| \leq \pi\sqrt{\hat{K}_n}} S_{1,n}^1
$$

is bounded for all $n \in \mathbb{N}$ with probability one.

To obtain analogous bounds for $S^2_{1,n}$, we observe that $\frac{\partial}{\partial u_2}\varphi_{(X_{M_n},Y_{M_n})}(u_0,v_0)$ is bounded by some constant since all moments of $X^{(2)}$ are finite. Thus we can use Lemma 5 and Lemma 8 again. Thus

$$\frac{\sqrt{n}}{\log(n)^4}\sup_{|u|,|v|\leq\pi\sqrt{\hat{K}_n}} S^2_{1,n}$$

is bounded for all $n\in\mathbb{N}$ with probability one.

The same arguments as for $S_{1,n}$ are applicable to $S_{3,n}$ and yield the same bounds.

To obtain these bounds for $S_{2,n}$ and $S_{4,n}$, we may use an analogous argument. As we have seen at the beginning of the proof we can assume that $|\varphi_{(\hat{z}_1,\hat{z}_2)^n_1}(u,v)\cdot\varphi_{\hat{\epsilon}^n_{1,1}}(u)\cdot\varphi_{\hat{\delta}^n_{1,1}}(v)|$ is bounded away from zero by $\frac{1}{2\log(n)}$. Furthermore

$$|\varphi_{(X,Y)^n_1}(t_0,0)| = \Big|\frac{1}{n}\sum_{l=1}^{n}\exp\Big(i\cdot(t\cdot X^{(1)}+0\cdot Y^{(1)}\Big)\Big|$$

is clearly bounded by one and

$$\begin{aligned}|\frac{\partial}{\partial u_2}\varphi_{(X,Y)^n_1}(t_0,0)| &= |\frac{1}{n}\sum_{j=1}^{n} iX^{(2)}_j e^{i(tX^{(1)}_j+0Y^{(1)}_j)}| \\ &\leq \frac{1}{n}\sum_{j=1}^{n}|X^{(2)}_j| \to \mathbf{E}\{|X^{(2)}|\} \quad a.s.\end{aligned}$$

implies that $\frac{\partial}{\partial u_2}\varphi_{(X,Y)^n_1}(t_0,0)$ is almost surely bounded. Thus using the same decomposition for S_2 and S_4 as for S_1 and applying Lemma 9 instead of Lemma 5 and Lemma 6 yields the desired bounds.

Returning to (18), we use the above bounds as well as $\beta_n = n^{\frac{1}{11}}$ and $\hat{K}_n \leq n^{\frac{1}{11}}$. With probability one, there exists a random $C>0$ such that

$$\begin{aligned}&(\hat{K}_n)^3\beta_n^2 \sup_{|u|\leq\pi\sqrt{\hat{K}_n},|v|\leq\pi\sqrt{\hat{K}_n}} |L_{\varphi_{u,n}}(v)-L_{\varphi_u}(v)| \\ &\leq n^{\frac{5}{11}} \sup_{|u|\leq\pi\sqrt{\hat{K}_n},|v|\leq\pi\sqrt{\hat{K}_n}} (S_{1,n}+S_{2,n}+S_{3,n}+S_{4,n}) \\ &\leq C\cdot n^{\frac{5}{11}}\cdot\frac{\log(n)^4}{\sqrt{n}}\end{aligned}$$

holds. Since the right hand side tends to 0 the assertion follows. □

Proof of Theorem 1. Under the assumptions of Theorem 1, Lemma 4 is applicable. Thus we can conclude the assertion of Theorem 1 from Lemma 3. □

References

[1] Bauer, H. (1996). *Probability Theory.* Walter de Gruyter & Co., Berlin.

[2] Beirlant, J. and Györfi, L. (1998). On the asymptotic L_2-error in partitioning regression estimation. *Journal of Statistical Planning and Inference*, **71**, pp. 93–107.

[3] Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, **53**, pp. 605-634.

[4] Connes, D., Ronchetti, E., and Victoria-Feser, M.-P. (2010). Goodness of fit for generalized linear latent variables models. *Journal of the American Statistical Association*, **105**, pp. 1126-1134.

[5] Devroye, L., Györfi, L., Krzyżak, A. and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics*, **22**, pp. 1371–1385.

[6] Devroye, L. and Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference*, **23**, pp. 71–82.

[7] Devroye, L. and Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics*, **8**, pp. 231–239.

[8] Györfi, L. (1981). Recent results on nonparametric regression estimate and multiple classification. *Problems of Control and Information Theory*, **10**, pp. 43–52.

[9] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression.* Springer Verlag, New York.

[10] Jackson, D. (1930). *The Theory of Approximation.* American Mathematical Society, New York.

[11] Kelava, A., Kohler, M., Krzyżak, A. and Weinbender, D. (2012). *Nonparametric Estimation of a Latent Variable Model.* Submitted for publication.

[12] Kohler, M. and Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Transactions on Information Theory*, **47**, pp. 3054-3058.

[13] Lang, S. (1999). *Complex Analysis.* Springer Verlag, New York.

[14] Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Transactions on Information Theory*, **41**, pp. 677–687.

[15] Mack, Y. P. (1981). Local properties of k–nearest neighbor regression estimates. *SIAM Journal on Algebraic and Discrete Methods*, **2**, pp. 311–323.

[16] Marsh, H. W., Wen, Z. and Hau, K.-T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R.O. Mueller (Eds.) *Struc-*

tural Equation Modeling: A Second Course, pp. 225-265. Information Age Publishing, Charlotte, NC.

[17] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, pp. 141–142.

[18] Nadaraya, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves. *Theory of Probability and its Applications*, **15**, pp. 134–137.

[19] Nobel, A. (1996). Histogram regression estimation using data-dependent partitions. *Annals of Statistics*, **24**, pp. 1084–1105.

[20] Paul, D., Bair, E., Hastie, T. and Tibshirani, R. (2008). "Preconditioning" for feature selection and regression in high-dimensional problems. *Annals of Statistics*, **36**, pp. 1595-1618.

[21] Schumacker, R. and Marcoulides, G. (1998). *Interaction and Nonlinear Effects in Structural Equation Modeling.* Lawrence Erlbaum Associates, Hillsdale, NJ.

[22] Skrondal, A. and Rabe-Hesketh, S. (2007). Latent variable modelling: A survey. *Scandinavian Journal of Statistics*, **34**, pp. 712-745.

[23] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics*, **5**, pp. 595–645.

[24] Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A*, **26**, pp. 359–372.

[25] Zhao, L. C. (1987). Exponential bounds of mean error for the nearest neighbor estimates of regression functions. *Journal of Multivariate Analysis*, **21**, pp. 168–178.

Supplementary material for the referees, not for publication

Proof of Lemma 1 The first assertion follows from some elementary calculations. Using (5) and the independence assumption we see that the characteristic function $\varphi_{(X,Y)}$ of (X,Y) is given by

$$\begin{aligned}
&\varphi_{(X,Y)}(u_1,\dots,u_d,v_1,\dots,v_l)\\
&= \mathbf{E}\left\{\exp\left(i\cdot\sum_{j=1}^{d} u_j\cdot X^{(j)} + i\cdot\sum_{k=1}^{l} v_k\cdot Y^{(k)}\right)\right\}\\
&= \mathbf{E}\left\{\exp\left(i\cdot\sum_{j=1}^{d} u_j\cdot(a_j\cdot Z_1+\epsilon_j) + i\cdot\sum_{k=1}^{l} v_k\cdot(b_k\cdot Z_2+\delta_k)\right)\right\}\\
&= \mathbf{E}\Bigg\{\exp\left(i\cdot(\sum_{j=1}^{d} u_j\cdot a_j\cdot Z_1 + \sum_{k=1}^{l} v_k\cdot b_k\cdot Z_2)\right)\cdot\prod_{j=1}^{d}\exp\left(i\cdot u_j\cdot\epsilon_j\right)\\
&\qquad\qquad\cdot\prod_{k=1}^{l}\exp\left(i\cdot v_k\cdot\delta_k\right)\Bigg\}\\
&= \varphi_{(Z_1,Z_2)}\left(\sum_{j=1}^{d} u_j\cdot a_j\ ,\ \sum_{k=1}^{l} v_k\cdot b_k\right)\cdot\prod_{j=1}^{d}\varphi_{\epsilon_j}(u_j)\cdot\prod_{k=1}^{l}\varphi_{\delta_k}(v_k).
\end{aligned}$$

Furthermore, using

$$\varphi_{\epsilon_j}(0)=\varphi_{\delta_k}(0)=1 \quad (j=2,\dots,d,k=2,\dots,l)$$

and

$$\varphi'_{\epsilon_2}(0)=i\cdot\mathbf{E}\epsilon_2=0=\varphi'_{\delta_2}(0)$$

we get

$$\varphi_{(X,Y)}(u_1,0,\dots,0,v_1,0,\dots,0)=\varphi_{(Z_1,Z_2)}(u_1,v_1)\cdot\varphi_{\epsilon_1}(u_1)\cdot\varphi_{\delta_1}(v_1)$$

as well as

$$\begin{aligned}
&\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(u_1,0,\dots,0,v_1,0,\dots,0)\\
&= a_2\cdot\frac{\partial}{\partial u_1}\varphi_{(Z_1,Z_2)}(u_1,v_1)\cdot\varphi_{\epsilon_1}(u_1)\cdot\varphi_{\delta_1}(v_1)\\
&\quad+\varphi_{(Z_1,Z_2)}(u_1,v_1)\cdot\varphi_{\epsilon_1}(u_1)\cdot\varphi_{\delta_1}(v_1)\cdot\varphi'_{\epsilon_2}(0)\\
&= a_2\cdot\frac{\partial}{\partial u_1}\varphi_{(Z_1,Z_2)}(u_1,v_1)\cdot\varphi_{\epsilon_1}(u_1)\cdot\varphi_{\delta_1}(v_1)
\end{aligned}$$

and

$$\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u_1,0,\dots,0,v_1,0,\dots,0)=b_2\cdot\frac{\partial}{\partial v_1}\varphi_{(Z_1,Z_2)}(u_1,v_1)\cdot\varphi_{\epsilon_1}(u_1)\cdot\varphi_{\delta_1}(v_1).$$

To show the second assertion, assume that in addition to (5)

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(d)} \\ Y^{(1)} \\ Y^{(2)} \\ \vdots \\ Y^{(l)} \end{pmatrix} = \begin{pmatrix} 1 \cdot \tilde{Z}_1 \\ \tilde{a}_2 \cdot \tilde{Z}_1 \\ \vdots \\ \tilde{a}_d \cdot \tilde{Z}_1 \\ 1 \cdot \tilde{Z}_2 \\ \tilde{b}_2 \cdot \tilde{Z}_2 \\ \vdots \\ \tilde{b}_l \cdot \tilde{Z}_2 \end{pmatrix} + \begin{pmatrix} \tilde{\epsilon}_1 \\ \tilde{\epsilon}_2 \\ \vdots \\ \tilde{\epsilon}_d \\ \tilde{\delta}_1 \\ \tilde{\delta}_2 \\ \vdots \\ \tilde{\delta}_l \end{pmatrix}.$$

also holds in distribution and that the appearing random variables are square integrable and satisfy assumptions $A4$) and $A5$). Set $a_1 = \tilde{a}_1 = b_1 = \tilde{b}_1 = 1$. By independence, we have for $i, j \in \{1, \dots, d\}$

$$a_j a_i \mathbf{E}\{Z_1^2\} = \mathbf{E}\{X^{(j)} \cdot X^{(i)}\} = \tilde{a}_j \tilde{a}_i \mathbf{E}\{\tilde{Z}_1^2\}$$

for $i, j \in \{1, \dots, d\}$. Since the left hand side does not vanish, every factor on the right hand side shares this property. This yields

$$a_2 = \frac{\mathbf{E}\{X^{(2)} \cdot X^{(3)}\}}{\mathbf{E}\{X^{(1)} \cdot X^{(3)}\}} = \tilde{a}_2.$$

Similiar arguments show $a_j = \tilde{a}_j$ and $b_i = \tilde{b}_i$ for all $j \in \{1, \dots, d\}$ and all $i \in \{1, \dots, l\}$. Using the same argument as in the proof of Lemma 3, we have

$$\begin{aligned} &\varphi_{(Z_1,Z_2)} = \\ &\exp\Big(\int_0^v \frac{1}{b_2} \cdot \frac{\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u,0,\dots,0,s,0,\dots,0)}{\varphi_{(X,Y)}(u,0,\dots,0,s,0,\dots,0)} ds\Big) \\ &\quad \cdot \exp\Big(\int_0^u \frac{1}{a_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(t,0,\dots,0)}{\varphi_{(X,Y)}(t,0,\dots,0)} dt\Big) \end{aligned}$$

and

$$\begin{aligned} &\varphi_{(\tilde{Z}_1,\tilde{Z}_2)} = \\ &\exp\Big(\int_0^v \frac{1}{\tilde{b}_2} \cdot \frac{\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u,0,\dots,0,s,0,\dots,0)}{\varphi_{(X,Y)}(u,0,\dots,0,s,0,\dots,0)} ds\Big) \\ &\quad \cdot \exp\Big(\int_0^u \frac{1}{\tilde{a}_2} \cdot \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(t,0,\dots,0)}{\varphi_{(X,Y)}(t,0,\dots,0)} dt\Big). \end{aligned}$$

With $\tilde{a}_2 = a_2$ and $\tilde{b}_2 = b_2$ we conclude $\varphi_{(\tilde{Z}_1,\tilde{Z}_2)} = \varphi_{(Z_1,Z_2)}$. Combining this with

$$\varphi_{(\tilde{Z}_1,\tilde{Z}_2)}(\tilde{a}_j \cdot u_j, 0) \cdot \varphi_{\tilde{\epsilon}_j}(u_j) = \varphi_{(X,Y)}(0,\dots,0,u_j,0,\dots,0) = \varphi_{(Z_1,Z_2)}(a_j \cdot u_j, 0) \cdot \varphi_{\epsilon_j}(u_j)$$

and similar relations yields

$$\varphi_{\epsilon_j} = \varphi_{\tilde{\epsilon}_j} \quad \text{and} \quad \varphi_{\delta_i} = \varphi_{\tilde{\delta}_i}$$

for all $j \in \{1, \ldots, d\}$ and all $i \in \{1, \ldots, l\}$. Together with

$$\varphi_{(\tilde{Z}_1, \tilde{Z}_2)} = \varphi_{(Z_1, Z_2)}$$

this implies the second assertion.