

Technical Supplement for: The Triangular Model with Random Coefficients

Stefan Hoderlein *
Boston College

Hajo Holzmann†
Marburg

Alexander Meister‡
Rostock

June 24, 2015

Abstract

We provide additional technical material for the paper Hoderlein, Holzmann and Meister (2015) (HHM in the following).

Keywords: Random Coefficients, Endogeneity, Consumer Demand, Welfare, Nonparametric Identification, Radon Transform.

Consider

$$\begin{aligned} Y &= B_0 + B_1 X, \\ X &= A_0 + A_1 Z, \end{aligned} \tag{B.1}$$

where Y, X, Z are observed random scalars, and $A = (A_0, A_1)'$, $B = (B_0, B_1)'$ are unobserved random coefficients. We impose the following two basic assumptions.

Assumption 1. *The random vector $(A', B)'$ has a continuous Lebesgue density f_{AB} .*

Assumption 2 (Exogeneity). *Z and $(A', B)'$ are independent.*

When analyzing the triangular RC model (B.1), it will often be convenient to pass to the reduced form model by inserting the second equation into the first one.

$$\begin{aligned} Y &= C_0 + C_1 Z, \\ X &= A_0 + A_1 Z. \end{aligned} \tag{B.2}$$

where $C = (C_0, C_1)$, $C_0 = B_0 + B_1 A_0$ and $C_1 = B_1 A_1$.

*Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, Tel. +1-617-552-6042. email: stefan_hoderlein@yahoo.com

†Department of Mathematics and Computer Science, Marburg University, Hans-Meerweinstr., 35032 Marburg, Germany, Tel. +49 -6421-2825454. email: holzmann@mathematik.uni-marburg.de

‡Institute for Mathematics, University of Rostock, 18051 Rostock, Germany, Tel. + 49 - 381-4986620. email: alexander.meister@uni-rostock.de

Assumption 3 (Independence and moment assumption). Suppose that $B = (B_0, B_1)$ and A_1 are independent, and that A_1^{-1} is absolutely integrable.

C Estimation under full support

C.1 Estimation of the scaling constant

Assumption 4. We impose that $f_Z(z) \geq c_Z \max\{1, |z|\}^{-\gamma}$ for all $z \in \mathbb{R}$ where $c_Z > 0$ and $\gamma > 1$ are constants.

Assumption 5. There is a $C_A > 0$, such that for all $\tilde{M} \geq 1$,

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R} \setminus [-\tilde{M}, \tilde{M}]} \int_{\mathbb{R}} f_A(x - a_1 z, a_1) da_1 dz \right)^2 v(x) dx \leq C_A \tilde{M}^{-1},$$

as well as

$$\sup_{x \in \mathbb{R}} \left| \int_{\mathbb{R}} [f_{A_0, A_1}(x - a_1 z, a_1) - f_{A_0, A_1}(x - a_1 w, a_1)] da_1 \right| \leq C_A |z - w| \quad z, w \in \mathbb{R}.$$

Let $(X_{(j)}, Y_{(j)}, Z_{(j)})$, $j = 1, \dots, n$, denote the sorted sample for which $Z_{(1)} < \dots < Z_{(n)}$, and let $v : \mathbb{R} \rightarrow [0, \infty)$ be a smooth probability density, $\tilde{M} > 0$ a truncation parameter. Set

$$\hat{a}_n(\tilde{M}) = \sum_{j=1}^{n-1} v(X_{(j)}) (Z_{(j+1)} - Z_{(j)}) \mathbf{1}_{\{-\tilde{M} \leq Z_{(j)} < Z_{(j+1)} < \tilde{M}\}},$$

Proposition 1. Given Assumptions 4 and 5, for $\tilde{M}_n \approx n^{1/(2\gamma+1)}$, we have that

$$E(\hat{a}_n(\tilde{M}_n) - E|A_1|^{-1})^2 = \mathcal{O}(n^{-1/(2\gamma+1)}).$$

Proof of Proposition 1. Let

$$sc := E|A_1|^{-1} = \int_{\mathbb{R}} v(x) \int_{\mathbb{R}} f_{X|Z}(x|z) dz dx,$$

so that

$$E(\hat{a}_n(\tilde{M}) - sc)^2 = E \text{Var}(\hat{a}_n(\tilde{M})|\sigma_Z) + E[E(\hat{a}_n(\tilde{M})|\sigma_Z) - sc]^2. \quad (\text{C.1})$$

As before,

$$\text{Var}(\hat{a}_n(\tilde{M})|\sigma_Z) \leq \sum_{-\tilde{M} \leq Z_{(j)} < Z_{(j+1)} < \tilde{M}} C_v^2 (Z_{(j+1)} - Z_{(j)})^2,$$

where C_v is a bound for v , and from Lemma F.1,

$$E \text{Var}(\hat{a}_n(\tilde{M})|\sigma_Z) \leq C n^{-1} \tilde{M}^{2\gamma}. \quad (\text{C.2})$$

Further, we have that

$$E(\hat{a}_n(\tilde{M})|\sigma_Z) = \int_{\mathbb{R}} \int_{\mathbb{R}} v(x) \tilde{f}(x, z) dz dx,$$

$$\tilde{f}(x, z) = \sum_{-\tilde{M} \leq Z_{(j)} < Z_{(j+1)} < \tilde{M}} f_{X|Z}(x|Z_{(j)}) \mathbf{1}_{[Z_{(j)}, Z_{(j+1)}]}(z).$$

Using the Cauchy-Schwarz inequality twice, we estimate

$$\begin{aligned}
& E[E(\hat{a}_n(\tilde{M})|\sigma_Z) - sc]^2 \\
& \leq 6\tilde{M} E \sum_{-\tilde{M} \leq Z_{(j)} < Z_{(j+1)} < \tilde{M}} \int_{\mathbb{R}} \int_{Z_{(j)}}^{Z_{(j+1)}} [f_{X|Z}(x|z) - f_{X|Z}(x|Z_{(j)})]^2 dz v(x) dx \\
& \quad + 3 E \left(\int_{\mathbb{R}} \left(\int_{-\infty}^{\max(Z_{(1)}, -\tilde{M})} f_{X|Z}(x|z) dz \right)^2 v(x) dx \right. \\
& \quad \left. + \int_{\mathbb{R}} \left(\int_{\min(Z_{(n)}, \tilde{M})}^{\infty} f_{X|Z}(x|z) dz \right)^2 v(x) dx \right).
\end{aligned}$$

For the first term, we proceed as before and obtain

$$\begin{aligned}
& \tilde{M} E \sum_{-\tilde{M} \leq Z_{(j)} < Z_{(j+1)} < \tilde{M}} \int_{\mathbb{R}} \int_{Z_{(j)}}^{Z_{(j+1)}} [f_{X|Z}(x|z) - f_{X|Z}(x|Z_{(j)})]^2 dz v(x) dx \\
& \leq C n^{-2} \tilde{M}^{3\gamma+1}.
\end{aligned}$$

For the second two terms, using Assumption 5 and Lemma F.1, we estimate

$$\begin{aligned}
& E \int_{\mathbb{R}} \left(\int_{-\infty}^{\max(Z_{(1)}, -\tilde{M})} f_{X|Z}(x|z) dz \right)^2 v(x) dx \\
& \leq C P(Z_{(1)} \geq -\tilde{M}) + \int_{\mathbb{R}} \left(\int_{-\infty}^{-\tilde{M}} f_{X|Z}(x|z) dz \right)^2 v(x) dx \\
& \leq C \left(\exp(-n c_Z \tilde{M}^{1-\gamma}/(\gamma-1)) + \tilde{M}^{-1} \right).
\end{aligned}$$

Summarizing, we obtain

$$E(\hat{a}_n(\tilde{M}) - sc)^2 = C [n^{-1} \tilde{M}^{2\gamma} + \tilde{M}^{-1} + n^{-2} \tilde{M}^{3\gamma+1} + \exp(-n c_Z \tilde{M}^{1-\gamma}/(\gamma-1))].$$

The given choice of \tilde{M} balances the first two terms and gives the rate, the other terms are negligible. \square

C.2 Nonparametric estimation of the density f_B

We now turn to nonparametric estimation of the density f_B itself. By Assumption 2 we can relate the identified conditional characteristic function of (X, Y) given $Z = z$ to $\psi_{A,C}$ via

$$\begin{aligned}
\psi_{X,Y|Z}(t_1, t_2|z) & := E(\exp(it_1 X + it_2 Y)|Z = z) \\
& = E \exp(it_1(A_0 + A_1 z) + it_2(C_0 + C_1 z)) = \psi_{A,C}(t_1, t_1 z, t_2, t_2 z)
\end{aligned} \tag{C.3}$$

where $z \in \text{supp } Z$.

Assumption 6.

$$\int_{\mathbb{R}^3} |t| \left| \int_{\mathbb{R}^3} \exp(it(b_0 + b_1 x)) f_{B,A_0,A_1}(b, x - a_1 z, a_1) da_1 db_0 db_1 \right| dz dt dx < \infty.$$

From part (ii) of Theorem 3 in HHM (and under Assumption 6), we set

$$g(b_0, b_1) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} |t| \exp(-itb_0) \psi_{X,Y|Z}(-tb_1, t|z) dt dz,$$

so that

$$f_B(b_0, b_1) = g(b_0, b_1) \cdot (E|A_1|^{-1})^{-1}.$$

Assumption 7. *a. There exists $\beta > 0$ such that*

$$\int_{[-L, L]} \int_{\mathbb{R}} (1+t^2)^\beta \left| \int \psi_{A, C}(-tb_1, -ub_1, t, u) du \right|^2 dt db_1 < \infty.$$

b. There are $\alpha > 1/2$, $c_D > 0$, so that for all $b_1 \in [-L, L]$, $t \in \mathbb{R}$, $z > 0$ we have that

$$\max \left(\int_z^\infty |\psi_{A, C}(-tb_1, -ub_1, t, u)| du, \int_{-\infty}^{-z} |\psi_{A, C}(-tb_1, -ub_1, t, u)| du \right) \leq c_D \min(z^{-\alpha}, 1).$$

We still require an estimate of $g(b_0, b_1)$. For a weight function $w(t)$ and a $h > 0$, we define the following approximation to $g(b_0, b_1)$:

$$\begin{aligned} \tilde{g}(b_0, b_1; h) &:= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} |t| w(ht) \exp(-itb_0) \psi_{X, Y|Z}(-tb_1, t|z) dt dz \\ &= \int \int \int K(y - b_1x - b_0; h) f_{X, Y|Z}(x, y|z) dx dy dz. \end{aligned} \quad (\text{C.4})$$

where

$$K(x; h) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} |t| \exp(itx) w(ht) dt.$$

Remark. (*Choice of weight function*) For our theoretical developments, we shall choose $w(t) = 1_{[-1, 1]}(t)$. In practice, however, smoother versions of $w(t)$ are preferable, see HKM, where also closed-form formulas for $K(x; h)$ for various weight functions can be found. \diamond

For an additional truncation parameter $M > 0$, we consider the estimator

$$\hat{g}(b_0, b_1; h, M) := \sum_{j=1}^{n-1} K(Y_{(j)} - b_1 X_{(j)} - b_0; h) (Z_{(j+1)} - Z_{(j)}) 1_{\{-M \leq Z_{(j)} < Z_{(j+1)} < M\}}.$$

If $\hat{a}_n(\tilde{M})$ denotes the above estimator of the scaling constant, as an estimator of f_B we finally take

$$\hat{f}_B(b_0, b_1; h, M, \tilde{M}) := \frac{\hat{g}(b_0, b_1; h, M)}{\hat{a}_n(\tilde{M})}.$$

Theorem 2. *Impose Assumptions 1, 2, 3, 4 and 7, and moreover suppose that the support of f_{B_0, B_1} is contained in $\mathbb{R} \times [-L, L]$ for some $L > 0$, as well as $E|A_1| + E|C_1| < \infty$. Then letting $M_n \approx n^{2\beta/(4\gamma\beta+2\beta+3)}$ and $h_n \approx n^{-1/(4\gamma\beta+2\beta+3)}$, we have that*

$$\int_{[-L, L]} \int_{\mathbb{R}} |\hat{f}_B(b_0, b_1; h_n, M_n, \tilde{M}_n) - f_B(b_0, b_1)|^2 db_0 db_1 = \mathcal{O}_P(n^{-\frac{2\beta}{4\gamma\beta+2\beta+3}})$$

Proof of Theorem 2. It suffices to show that

$$MSE(\hat{g}) := E \int_{[-L, L]} \int_{\mathbb{R}} |\hat{g}(b_0, b_1; h_n, M_n) - g(b_0, b_1)|^2 db_0 db_1 = \mathcal{O}_P(n^{-\frac{2\beta}{4\gamma\beta+2\beta+3}}),$$

then the result will follow by using Proposition 1. We have that

$$\begin{aligned} MSE(\hat{g}) &\leq 2E \int_{[-L,L]} \int_{\mathbb{R}} |\hat{g}(b_0, b_1; h) - \tilde{g}(b_0, b_1; h)|^2 db_0 db_1 \\ &\quad + 2 \int_{[-L,L]} \int_{\mathbb{R}} |\tilde{g}(b_0, b_1; h) - g(b_0, b_1)|^2 db_0 db_1 =: 2I_1 + 2I_2. \end{aligned}$$

Using the Parseval equality w.r.t. b_0 , we may estimate the bias term I_2 as

$$\begin{aligned} I_2 &= \frac{1}{2\pi} \int_{[-L,L]} \int_{\mathbb{R}} t^2 (1 - w(ht))^2 \left| \int \psi_{A,C}(-tb_1, -tz b_1, t, tz) dz \right|^2 dt db_1 \\ &= \frac{1}{2\pi} \int_{[-L,L]} \int_{\mathbb{R}} (1 - w(ht))^2 \left| \int \psi_{A,C}(-tb_1, -ub_1, t, u) du \right|^2 dt db_1 \\ &\leq \sup_{t \geq 0} \frac{(1 - w(ht))^2}{(1 + t^2)^\beta} \frac{1}{2\pi} \int_{[-L,L]} \int_{\mathbb{R}} (1 + t^2)^\beta \left| \int \psi_{A,C}(-tb_1, -ub_1, t, u) du \right|^2 dt db_1 \\ &\leq C_{BIAS} h^{2\beta} \end{aligned}$$

for some constant C_{BIAS} , when making use Assumption 7, a. For I_1 , using again the Parseval equality we obtain

$$\begin{aligned} I_1 &= \frac{1}{2\pi} \int_{[-L,L]} \int_{\mathbb{R}} t^2 w(ht)^2 E \left| \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right. \\ &\quad \left. - \sum_{j=1}^{n-1} \exp(itY_{(j)} - itb_1 X_{(j)}) (Z_{(j+1)} - Z_{(j)}) 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M} \right|^2 dt db_1. \end{aligned} \tag{C.5}$$

Let σ_Z again denote the σ -field generated by Z_1, \dots, Z_n . For fixed t, b_1 , consider

$$\begin{aligned} &E \left| \sum_{j=1}^{n-1} \exp(itY_{(j)} - itb_1 X_{(j)}) (Z_{(j+1)} - Z_{(j)}) 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M} - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 \\ &= E \left[\text{Var} \left(\sum_{j=1}^{n-1} \exp(itY_{(j)} - itb_1 X_{(j)}) (Z_{(j+1)} - Z_{(j)}) 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M} \middle| \sigma_Z \right) \right] \\ &\quad + E \left| E \left(\sum_{j=1}^{n-1} \exp(itY_{(j)} - itb_1 X_{(j)}) (Z_{(j+1)} - Z_{(j)}) 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M} \middle| \sigma_Z \right) - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 \\ &= I_3 + I_4. \end{aligned}$$

Since the $(X_{(j)}, Y_{(j)})$ are independent conditional on σ_Z , for the conditional variance term we have that

$$I_3 \leq \sum_{j=1}^{n-1} E (Z_{(j+1)} - Z_{(j)})^2 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M},$$

and its contribution in I_1 in (C.5) is bounded by

$$\leq Ch^{-3} \sum_{j=1}^{n-1} E (Z_{(j+1)} - Z_{(j)})^2 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M}.$$

For the conditional expectation, we have

$$\begin{aligned} &E \left(\sum_{j=1}^{n-1} \exp(itY_{(j)} - itb_1 X_{(j)}) (Z_{(j+1)} - Z_{(j)}) 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M} \middle| \sigma_Z \right) \\ &= \sum_{j=1}^{n-1} \psi_{X,Y|Z}(-tb_1, t|Z_{(j)}) (Z_{(j+1)} - Z_{(j)}) 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M} = \int_{\mathbb{R}} \tilde{\psi}(-tb_1, t; z) dz, \end{aligned}$$

where

$$\tilde{\psi}(-tb_1, t; z) = \sum_{j=1}^{n-1} \psi_{X,Y|Z}(-tb_1, t|Z_{(j)}) \mathbf{1}_{[Z_{(j)}, Z_{(j+1)}](z)} \mathbf{1}_{-M \leq Z_{(j)} < Z_{(j+1)} < M}.$$

Therefore,

$$\begin{aligned} & \left| E \left(\sum_{-M \leq Z_{(j)} < Z_{(j+1)} < M} \exp(itY_{(j)} - itb_1 X_{(j)}) (Z_{(j+1)} - Z_{(j)}) \middle| \sigma_Z \right) - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 \\ & \leq 3 \left| \int_{\min(Z_{(n)}, M)}^{\infty} \psi_{A,C}(-tb_1, -tzb_1, t, tz) dz \right|^2 + 3 \left| \int_{-\infty}^{\max(-M, Z_{(1)})} \psi_{A,C}(-tb_1, -tzb_1, t, tz) dz \right|^2 \\ & \quad + 6M \sum_{-M \leq Z_{(j)} < Z_{(j+1)} < M} \int_{Z_{(j)}}^{Z_{(j+1)}} |\psi_{A,C}(-tb_1, -tzb_1, t, tz) - \psi_{A,C}(-tb_1, -tZ_{(j)}b_1, t, tZ_{(j)})|^2 dz \end{aligned}$$

Let us bound each term. First,

$$\begin{aligned} & \sum_{-M \leq Z_{(j)} < Z_{(j+1)} < M} \int_{Z_{(j)}}^{Z_{(j+1)}} |\psi_{A,C}(-tb_1, -tzb_1, t, tz) - \psi_{A,C}(-tb_1, -tZ_{(j)}b_1, t, tZ_{(j)})|^2 dz \\ & \leq |t|^2 (|b_1|E|A_1| + E|C_1|) \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^3 \mathbf{1}_{-M \leq Z_{(j)} < Z_{(j+1)} < M}, \end{aligned}$$

and the contribution in I_1 in (C.5) is bounded by

$$\leq C h^{-5} M \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^3 \mathbf{1}_{-M \leq Z_{(j)} < Z_{(j+1)} < M}.$$

By Assumption 7 b., for $t \neq 0$

$$\begin{aligned} & \left| \int_{\min(Z_{(n)}, M)}^{\infty} \psi_{A,C}(-tb_1, -tzb_1, t, tz) dz \right| \leq |t|^{-1} c_D \left(\mathbf{1}_{Z_{(n)} > M} \min(|t|M^{-\alpha}, 1) + \mathbf{1}_{Z_{(n)} \leq M} \right), \\ & \left| \int_{-\infty}^{\max(-M, Z_{(1)})} \psi_{A,C}(-tb_1, -tzb_1, t, tz) dz \right| \leq |t|^{-1} c_D \left(\mathbf{1}_{Z_{(1)} < -M} \min(|t|M^{-\alpha}, 1) + \mathbf{1}_{Z_{(1)} \geq -M} \right). \end{aligned}$$

Thus

$$\begin{aligned} & \frac{1}{2\pi} \int_{[-L, L]} \int_{\mathbb{R}} t^2 w(ht)^2 \left| \int_{\min(Z_{(n)}, M)}^{\infty} \psi_{A,C}(-tb_1, -tzb_1, t, tz) dz \right|^2 dt db_1 \\ & \leq C \int_{\mathbb{R}} w(ht)^2 t^2 \left(\mathbf{1}_{Z_{(n)} > M} [\min(|t|M^{-\alpha}, 1)]^2 + \mathbf{1}_{Z_{(n)} \leq M} \right) dt \\ & \leq C \mathbf{1}_{Z_{(n)} > M} \left(\int_{-1/M}^{1/M} 1 dt + M^{-2\alpha} \int_{1/M}^{\infty} |t|^{-2\alpha} dt \right) + h^{-3} \mathbf{1}_{Z_{(n)} \leq M} \\ & \leq C M^{-1} + h^{-3} \mathbf{1}_{Z_{(n)} \leq M}, \end{aligned}$$

and similarly for $Z_{(k)}$. Summarizing, we obtain using Lemma F.1

$$\begin{aligned} I_1 & \leq C h^{-3} E \sum_{-M \leq Z_{(j)} < Z_{(j+1)} < M} (Z_{(j+1)} - Z_{(j)})^2 + C h^{-5} M E \sum_{-M \leq Z_{(j)} < Z_{(j+1)} < M} (Z_{(j+1)} - Z_{(j)})^3 \\ & \quad + C M^{-1} + C h^{-3} \left(P(Z_{(n)} \leq M) + P(Z_{(1)} \geq -M) \right) \\ & \leq C (h^{2\beta} + h^{-3} n^{-1} M^{2\gamma} + h^{-5} n^{-2} M^{3\gamma+1} + M^{-1} + \exp(-nc_Z M^{1-\gamma}/(\gamma-1))). \end{aligned}$$

Choosing M_n and h_n as in the theorem balances the three terms $h^{2\beta}$, $h^{-3} n^{-1} M^{2\gamma}$, and M^{-1} , and gives the rate, the other terms are negligible. \square

D Nonparametric estimation under limited support

Next, we consider nonparametric estimation for compactly supported Z . First, we recall the following lemma from HHM.

Lemma D.1. *Let $(A', B)'$, $A' = (A_0, A_1)$, $B' = (B_0, B_1)$ be a four-dimensional random vector with continuous Lebesgue density, which satisfies Assumptions 3 and 6, and for which $\mathcal{F}_2 f_B$ is integrable. Set $C_0 = B_0 + B_1 A_0$, $C_1 = B_1 A_1$ and $C' = (C_0, C_1)$, and let $\psi_{A,C}$ denote the characteristic function of $(A', C)'$. Then*

$$f_B(b_0, b_1) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \exp(-itb_0) \psi_{A,C}(-tb_1, -tzb_1, t, tz) |t| dt dz (E|A_1|^{-1})^{-1}. \quad (\text{D.1})$$

We always maintain the Assumptions 1, 2, and 3, and as above, let $(X_{(j)}, Y_{(j)}, Z_{(j)})$, $j = 1, \dots, n$, denote the sorted sample for which $Z_{(1)} < \dots < Z_{(n)}$. Consider

Assumption 8. *Assume that Z has support $[-1, 1]$, and a density f_Z with $f_Z(z) \geq cz$ for all $z \in [-1, 1]$ for some $c_Z > 0$.*

Assumption 9. *In model (B.2), all the absolute moments of A_1 and $C_1 = B_1 A_1$ are finite and satisfy*

$$\lim_{k \rightarrow \infty} \frac{d^k}{k!} (E|A_1|^k + E|C_1|^k) = 0,$$

for all fixed $d \in (0, \infty)$.

Under these additional assumptions, we obtain identification as in Theorem 6 in HHM.

For nonparametric estimation, we require an estimator of $\psi_{X,Y|Z}(x, y|z)$ for all $z \in \mathbb{R}$ (or an increasing range of values of z) and not only for those z which are contained in the support $[-1, 1]$ of Z . To this end we use analytic continuation, and first introduce the normalized Legendre polynomials

$$L_k(x) = \frac{\sqrt{k+1/2}}{k!2^k} \frac{d^k}{dx^k} (x^2 - 1)^k, \quad (\text{D.2})$$

which satisfy the following orthogonality relations

$$\int_{-1}^1 L_k(x) L_j(x) dx = \begin{cases} 1, & \text{for } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

Thus the orthogonal projection $\mathcal{P}_J f$ of some $f \in L_2([-1, 1])$ onto the linear span of the polynomials with the degree $\leq J$ can be represented by

$$[\mathcal{P}_J f](x) = \sum_{k=0}^J L_k(x) \int_{-1}^1 f(y) L_k(y) dy. \quad (\text{D.3})$$

Clearly the function $z \mapsto \psi_{X,Y|Z}(x, y|z)$, $z \in [-1, 1]$, is located in $L_2([-1, 1])$ so that this function can take the role of f in the above expansion. The $L_2([-1, 1])$ -inner product of $\psi_{X,Y|Z}(x, y|\cdot)$ and L_k is estimated by

$$\hat{\psi}_k(x, y) = \sum_{j=1}^{n-1} \exp(ixX_{(j)} + iyY_{(j)}) \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(z) dz,$$

As an estimator of $\psi_{X,Y|Z}(x, y|z)$, we consider

$$\hat{Q}_J(x, y|z) = \sum_{k=0}^J L_k(z) \hat{\psi}_k(x, y). \quad (\text{D.4})$$

Note that the domain of $\hat{Q}_J(x, y|z)$ can be continued from $z \in [-1, 1]$ to $z \in \mathbb{R}$ in a unique natural way as the function $z \mapsto \hat{Q}(x, y|z)$ represents a polynomial.

Motivated by Lemma D.1, we define

$$\begin{aligned} \hat{g}(b_0, b_1) &= \frac{1}{(2\pi)^2} \iint \exp(-itb_0) w(th_0) w(zh_1) |t| \hat{Q}_J(-tb_1, t|z) dt dz \\ &= \frac{1}{(2\pi)^2} \int \exp(-itb_0) w(th_0) |t| \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \hat{\psi}_k(-tb_1, t) dt \\ &= \frac{1}{(2\pi)^2} \int \left\{ \exp(-itb_0) \sum_{j=1}^{n-1} w(th_0) |t| \exp(-itb_1 X_{(j)} + itY_{(j)}) \right. \\ &\quad \left. \cdot \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \right\} dt \end{aligned} \quad (\text{D.5})$$

as an estimator of $g(b_0, b_1)$ for a kernel function K and two bandwidth parameters h_0 and h_1 . Therein we allow for the parameters h_1 and J to depend on t .

Assumption 10 (Support). *The support of f_{B_0, B_1} is contained in $[-L, L]^2$ for some $L > 0$.*

The final estimator is then defined by

$$\hat{f}_B(b_0, b_1) = \hat{g}(b_0, b_1) \left(\int_{[-L, L]^2} \hat{g}(b_0, b_1) db_0 db_1 \right)^{-1}.$$

In order to study the asymptotic MISE of this estimator some quantitative conditions are required. With respect to the density of (A, C) we impose that

$$\int \left(\int \sup_{x_0, x_1} |\psi_{A,C}(x_0, x_1, x_2, x_3)| dx_3 \right)^2 dx_2 < \infty. \quad (\text{D.6})$$

Furthermore we assume that

$$\max\{|A_1|, |C_1|\} \leq R \quad \text{a.s.}, \quad (\text{D.7})$$

which strengthens Assumption 9. We provide

Theorem 1. (Consistency) *Impose Assumption 10 as well as conditions (D.6) and (D.7). Fix $h_0 = (\log_2 n)/(c_{h,0} \log n)$ with some constant $c_{h,0} > 0$; and $T \asymp (\log n)^{-\delta}$ for some $\delta > 0$. We select $J = h_1 = 1$ when $|t| \leq T$. On the domain $|t| > T$ we choose $J = C_J(\log n)/\log_2 n$, and $h_1 = |t| \cdot (\log_2 n)/(c_{h,1} \log n)$ with some constants $C_J, c_{h,1} > 0$ where \log_k denotes the k -fold iterated logarithm. Moreover we impose that*

$$C_J < 1/(3 + 3\delta), \quad (\text{D.8})$$

$$c_{h,0} + c_{h,1} < C_J/(4LR \exp(1)). \quad (\text{D.9})$$

Then the above estimator of g_B satisfies

$$\lim_{n \rightarrow \infty} E \int_{[-L, L]^2} |\hat{g}(b_0, b_1) - g(b_0, b_1)|^2 db_1 db_0 = 0,$$

and further,

$$\int_{[-L, L]^2} |\hat{f}_B(b_0, b_1) - f_B(b_0, b_1)|^2 db_1 db_0 = o_P(1).$$

The statement about \hat{f}_B follows from that about \hat{g} , since over a finite domain, L_2 -convergence implies L_1 -convergence and hence convergence of the integrals.

Remark. In order to establish convergence rates in the context of Theorem 1 some smoothness restrictions on the density $f_{A,C}$ are required such as

$$\int (1 + x_2)^\beta \left(\int \sup_{x_0, x_1} |\psi_{A,C}(x_0, x_1, x_2, x_3)| dx_3 \right)^2 dx_2 \leq c_S,$$

for some uniform constant c_S where β denotes the smoothness level, which is very similar to Assumption 7 a. Then our estimator attains the convergence rates $(\log_2 n / \log n)^{2\beta}$ if $\delta > 2\beta$ as an inspection of the proof of Theorem 1 shows. These slow rates compare well to the results of Goldenshluger (2002) and Meister (2007) in related estimation problems.

Proof of Theorem 1.

$$\begin{aligned} E \int_{[-L, L]^2} |\hat{g}(b_0, b_1) - g(b_0, b_1)|^2 db_1 db_0 &\leq E \int_{-L}^L \int_{\mathbb{R}} |\hat{g}(b_0, b_1) - g(b_0, b_1)|^2 db_0 db_1 \\ &= \frac{1}{2\pi} \int_{-L}^L \int_{\mathbb{R}} |t|^2 E \left| \sum_{j=1}^{n-1} w(th_0) \exp(-itb_1 X_{(j)} + itY_{(j)}) \cdot \sum_{k=0}^J \int w(zh_1) L_k(z) dz \right. \\ &\quad \left. \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 dt db_1, \end{aligned} \quad (\text{D.10})$$

by Parseval's identity. For any fixed $t \in \mathbb{R}$ we consider that

$$\begin{aligned}
& E \left| \sum_{j=1}^{n-1} w(th_0) \exp(-itb_1 X_{(j)} + itY_{(j)}) \cdot \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \right. \\
& \qquad \qquad \qquad \left. - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 \\
& = E \operatorname{var} \left(\sum_{j=1}^{n-1} w(th_0) \exp(-itb_1 X_{(j)} + itY_{(j)}) \cdot \sum_{k=0}^J \int w(zh_1) L_k(z) dz \right. \\
& \qquad \qquad \qquad \left. \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \middle| \sigma_Z \right) \\
& + E \left| E \left(\sum_{j=1}^{n-1} w(th_0) \exp(-itb_1 X_{(j)} + itY_{(j)}) \cdot \sum_{k=0}^J \int w(zh_1) L_k(z) dz \right. \right. \\
& \qquad \qquad \qquad \left. \left. \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \middle| \sigma_Z \right) - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 \tag{D.11}
\end{aligned}$$

where we write σ_Z for the σ -field generated by Z_1, \dots, Z_n . We deduce that

$$\begin{aligned}
& E \left(w(th_0) \exp(-itb_1 X_{(j)} + itY_{(j)}) \cdot \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \middle| \sigma_Z \right) \\
& = w(th_0) E \left(\exp(-itb_1 X_{(j)} + itY_{(j)}) \middle| \sigma_Z \right) \sum_{k=0}^J \int w(zh_1) L_k(z) dz \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \\
& = w(th_0) \psi_{X,Y|Z}(-tb_1, t|Z_{(j)}) \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du. \tag{D.12}
\end{aligned}$$

Moreover we obtain that

$$\begin{aligned}
& \operatorname{var} \left(\sum_{j=1}^{n-1} w(th_0) \exp(-itb_1 X_{(j)} + itY_{(j)}) \cdot \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \middle| \sigma_Z \right) \\
& = \sum_{j=1}^{n-1} |w(th_0)|^2 \left| \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{Z_{(j)}}^{Z_{(j+1)}} L_k(u) du \right|^2 \\
& \qquad \qquad \qquad \cdot \operatorname{var} \left(\exp(-itb_1 X_{(j)} + itY_{(j)}) \middle| \sigma_Z \right) \\
& \leq |w(th_0)|^2 \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^2 \cdot \left(\sum_{k=0}^J \left| \int w(zh_1) L_k(z) dz \right| \sup_{|u| \leq 1} |L_k(u)| \right)^2 \\
& \leq (J + 1/2) |w(th_0)|^2 \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^2 \cdot \left(\sum_{k=0}^J \left| \int w(zh_1) L_k(z) dz \right| \right)^2, \tag{D.13}
\end{aligned}$$

almost surely where we have exploited the conditional independence of the $(X_{(j)}, Y_{(j)})$, $j = 1, \dots, n$ given σ_Z .

We use the inequality

$$|L_k(x)| \leq \sqrt{k + 1/2} (2|x| + 2)^k, \quad \forall x \in \mathbb{R},$$

to show that (D.13) is bounded from above by

$$(J + 1/2)^2 |w(th_0)|^2 (2/h_1 + 2)^{2J+2} \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^2. \tag{D.14}$$

Lemma F.1, for $\kappa = 2$, yields that the expectation of term (D.14) is bounded from above by

$$2c_Z^{-2} |w(th_0)|^2 (J + 1/2)^2 (2/h_1 + 2)^{2J+2} \mathcal{O}(1/n). \quad (\text{D.15})$$

Thus, the first addend in (D.11) obeys the uniform asymptotic upper bound (D.15). Let us now focus on the second addend in (D.11). The equality (D.12) yields that this term equals

$$E \left| w(th_0) \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{-1}^1 \tilde{\psi}(t, b_1, u) L_k(u) du - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2. \quad (\text{D.16})$$

where

$$\tilde{\psi}(t, b_1, u) := \sum_{j=1}^{n-1} \psi_{X,Y|Z}(-tb_1, t|Z_{(j)}) \cdot 1_{[Z_{(j)}, Z_{(j+1)})}(u).$$

We consider that

$$\begin{aligned} & E \left| w(th_0) \sum_{k=0}^J \int w(zh_1) L_k(z) dz \cdot \int_{-1}^1 (\tilde{\psi}(t, b_1, u) - \psi_{X,Y|Z}(-tb_1, t|u)) L_k(u) du \right|^2 \\ & \leq |w(th_0)|^2 \left(\sum_{k=0}^J \left| \int w(zh_1) L_k(z) dz \right| \right)^2 E \int_{-1}^1 |\tilde{\psi}(t, b_1, u) - \psi_{X,Y|Z}(-tb_1, t|u)|^2 du \\ & \leq |w(th_0)|^2 (J + 1/2) \cdot (2/h_1 + 2)^{2J+2} \left(E(1 - Z_{(n)}) + E(Z_{(1)} + 1) \right. \\ & \quad \left. + \sum_{j=1}^{n-1} E \int_{Z_{(j)}}^{Z_{(j+1)}} |\tilde{\psi}(t, b_1, u) - \psi_{X,Y|Z}(-tb_1, t|u)|^2 du \right) \\ & \leq |w(th_0)|^2 (J + 1/2) \cdot (2/h_1 + 2)^{2J+2} \left(E(1 - Z_{(n)}) + E(Z_{(1)} + 1) \right. \\ & \quad \left. + \frac{1}{3} |t|^2 \{ |b_1| E|A_1| + E|C_1| \}^2 E \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^3 \right) \\ & \leq |w(th_0)|^2 (J + 1/2) \cdot (2/h_1 + 2)^{2J+2} \left(E(1 - Z_{(n)}) + E(Z_{(1)} + 1) \right. \\ & \quad \left. + 2|t|^2 \{ |b_1| E|A_1| + E|C_1| \}^2 c_Z^{-3} \cdot \mathcal{O}(n^{-2}) \right), \end{aligned}$$

by Lemma F.1 since, for $u \in [Z_{(j)}, Z_{(j+1)})$, we have

$$\begin{aligned} |\tilde{\psi}(t, b_1, u) - \psi_{X,Y|Z}(-tb_1, t|u)| &= |\psi_{A,C}(-tb_1, -tb_1 Z_{(j)}, t, t Z_{(j)}) - \psi_{A,C}(-tb_1, -tb_1 u, t, tu)| \\ &\leq |u - Z_{(j)}|^2 \cdot |t|^2 \{ |b_1| E|A_1| + E|C_1| \}^2. \end{aligned}$$

By Lemma F.1, (D.16) is bounded from above by the sum of

$$4|w(th_0)|^2 (J + 1/2) \cdot (2/h_1 + 2)^{2J+2} (c_Z^{-1} n^{-1} + c_Z^{-3} |t|^2 \{ |b_1| E|A_1| + E|C_1| \}^2 \cdot \mathcal{O}(n^{-2})), \quad (\text{D.17})$$

and the deterministic expression

$$2 \left| w(th_0) \int w(zh_1) [\mathcal{P}_J \psi_{X,Y|Z}(-tb_1, t|\cdot)](z) dz - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2. \quad (\text{D.18})$$

For any $u \in \mathbb{R}$ we apply the decomposition

$$\psi_{X,Y|Z}(-tb_1, t|u) = \psi_{A,C}(-tb_1, -tb_1 u, t, tu) = T_J(t, b_1; u) + R_J(t, b_1; u),$$

with the Taylor polynomial

$$T_J(t, b_1; u) = \sum_{k=0}^J \frac{1}{k!} u^k \frac{d^k \psi_{A,C}}{(dv)^k}(-tb_1, -tb_1 v, t, tv) \Big|_{v=0},$$

and the residual term $R_J(t, b_1; u)$ which satisfies

$$\begin{aligned} |R_J(t, b_1; u)| &\leq \frac{1}{(J+1)!} |u|^{J+1} \sup_{v \in \mathbb{R}} \left| \frac{d^{J+1} \psi_{A,C}}{(dv)^{J+1}}(-tb_1, -tb_1 v, t, tv) \right| \\ &\leq \frac{1}{(J+1)!} |u|^{J+1} (2|t|)^{J+1} (|b_1|^{J+1} E|A_1|^{J+1} + E|C_1|^{J+1}), \end{aligned} \quad (\text{D.19})$$

where we use the same bounding technique as in the proof of Lemma ???. Note that

$$\mathcal{P}_J T_J(t, b_1; \cdot) = T_J(t, b_1; \cdot),$$

since $u \mapsto T_J(t, b_1; u)$ is a polynomial with the degree $\leq J$. Also note that by the definition (D.3) the operator \mathcal{P}_J can be extended to functions with the real line as their domain. Hence term (D.18) equals

$$\begin{aligned} &2 \left| w(th_0) \left(\int w(zh_1) T_J(t, b_1; z) dz + \int w(zh_1) [\mathcal{P}_J R_J(t, b_1; \cdot)](z) dz \right) \right. \\ &\quad \left. - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 \\ &\leq 2 \left| w(th_0) \left(\int w(zh_1) \psi_{X,Y|Z}(-tb_1, t|z) dz - \int w(zh_1) R_J(t, b_1; z) dz \right. \right. \\ &\quad \left. \left. + \int w(zh_1) [\mathcal{P}_J R_J(t, b_1; \cdot)](z) dz \right) - \int \psi_{X,Y|Z}(-tb_1, t|z) dz \right|^2 \\ &\leq 4 \left| \int (w(th_0) w(zh_1/t) - 1) \psi_{A,C}(-tb_1, -b_1 z, t, z) dz \right|^2 / |t|^2 \\ &\quad + 4 |w(th_0)|^2 \left| \int w(zh_1) ([\mathcal{P}_J R_J(t, b_1; \cdot)](z) - R_J(t, b_1; z)) dz \right|^2 \end{aligned} \quad (\text{D.20})$$

The second term in (D.20) has the upper bound

$$\begin{aligned} &8 |w(th_0)|^2 \left(\sum_{k=0}^J \left| \int w(zh_1) L_k(z) dz \right| \right)^2 \int_{-1}^1 |R_J(t, b_1; u)|^2 du \\ &\quad + 8 |w(th_0)|^2 \left(\int |w(zh_1)| |R_J(t, b_1; z)| dz \right)^2 \\ &\leq 16 |w(th_0)|^2 (J+1/2)(2/h_1 + 2)^{2J+2} \frac{(2|t|)^{2J+2}}{(J+1)!^2 (2J+3)} \{ |b_1|^{J+1} E|A_1|^{J+1} + E|C_1|^{J+1} \}^2 \\ &\quad + 32 |w(th_0)|^2 \frac{1}{(J+2)!^2} h_1^{-2J-2} (2|t|)^{2J+2} \{ |b_1|^{J+1} E|A_1|^{J+1} + E|C_1|^{J+1} \}^2, \end{aligned} \quad (\text{D.21})$$

thanks to (D.19).

Now let us piece everything together. According to (D.15), (D.17), (D.21) and the first term in (D.20) The MISE of \hat{g} is bounded from above by a constant times the sum of

$$\int w^{ft}(th_0) |t|^2 J^2 (2/h_1 + 2)^{2J+2} dt \cdot \mathcal{O}(1/n), \quad (\text{D.22})$$

$$\int w^{ft}(th_0) |t|^4 J (2/h_1 + 2)^{2J+2} dt \cdot \mathcal{O}(1/n^2), \quad (\text{D.23})$$

$$\int w^{ft}(th_0) |t|^2 (2/h_1 + 2)^{2J+2} L^{2J} (E|A_1|^{J+1} + E|C_1|^{J+1})^2 (2|t|)^{2J+2} / (J+1)!^2 dt, \quad (\text{D.24})$$

$$\int \left| \int [w^{ft}(th_0) w^{ft}(zh_1/t) - 1] \psi_{A,C}(-tb_1, -b_1 z, t, z) dz \right|^2 dt. \quad (\text{D.25})$$

On the domain $|t| > T$ we choose $J = C_J(\log n)/\log_2 n$ and $h_1 = |t| \cdot (\log_2 n)/(c_{h,1} \log n)$ with some constants $C_J, c_{h,1} > 0$. Condition (D.6) guarantees that the term (D.25) tends to zero by dominated convergence. Moreover condition (D.7) yields that (D.24) obeys the upper bound

$$\mathcal{O}\left(h_0^{-3} \exp\left\{\frac{2C_J \log n}{\log_2 n} \cdot \log\left(\frac{4(c_{h,0} + c_{h,1})LR \exp(1)}{C_J}\right)\right\}\right),$$

where we have used Stirling's formula. Therefore the constraint (D.9) guarantees that the term (D.24) tends to zero with a super-logarithmic rate. The terms (D.22) and (D.23) obey the asymptotic upper bounds

$$\mathcal{O}\left(h_0^{-3} J^2 \exp\left\{3C_J \log n \cdot \left(\frac{\log[3c_{h,1}]}{\log_2 n} - \frac{\log_3 n}{\log_2 n} + 1 + \delta\right) - \log n\right\}\right),$$

and

$$\mathcal{O}\left(h_0^{-5} J \exp\left\{3C_J(\log n) \cdot \left(\frac{\log[3c_{h,1}]}{\log_2 n} - \frac{\log_3 n}{\log_2 n} + 1 + \delta\right) - 2 \log n\right\}\right),$$

respectively, which decays faster than some polynomial rate with respect to n , thanks to (D.8). Finally we conclude that the asymptotic behaviour of the statistical risk is governed by the term (D.25), which completes the proof of the first part of the theorem. \square

E Contrast condition for the normal distribution

In Section 5.2 in HHM, we suppose that $f_B = f_{B,\theta_0}$ belongs to the parametric family of models $\{f_{B,\theta} : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^d$ is a d -dimensional bounded cuboid. For $\theta \in \Theta$ and $t \in \mathbb{R}$ we let

$$\Phi(\theta, t, I) := \int_I (\mathcal{F}f_{B,\theta})(t, tx) dx.$$

Let ν be a probability measure on \mathbb{R} , and let I_1, \dots, I_q be finitely many (distinct) intervals. For bounded functions $\Phi_1(t, I_j)$ and $\Phi_2(t, I_j)$, $t \in \mathbb{R}$, $j = 1, \dots, q$, we set

$$\|\Phi_1(\cdot) - \Phi_2(\cdot)\|_{\nu; q}^2 := \frac{1}{q} \sum_{j=1}^q \int_{\mathbb{R}} |\Phi_1(t, I_j) - \Phi_2(t, I_j)|^2 d\nu(t), \quad (\text{E.1})$$

Assumption 11. *There exist intervals I_1, \dots, I_q and a probability measure ν , such that*

$$c_{\Theta, 0} \|\theta - \theta'\|^2 \leq \|\Phi(\theta, \cdot) - \Phi(\theta', \cdot)\|_{\nu; q}^2 \leq c_{\Theta, 1} \|\theta - \theta'\|^2$$

for all $\theta, \theta' \in \Theta$ with some uniform constants $c_{\Theta, j}$, $j = 0, 1$.

Note that in the context of Section 5.2 in HHM, the intervals I_1, \dots, I_q in addition have to satisfy the support condition (5.1) (in HHM). Further, we included the scaling constant into the constants $c_{\Theta, j}$, $j = 0, 1$.

In the following proposition, we show the claim of the example following the above assumption in HHM, namely that it holds for the bivariate normal distribution.

Proposition 3. *Suppose that $\{f_{B,\theta} : \theta \in \Theta\}$ is the bivariate normal distribution, where the parameter set Θ is restricted to a compact cuboid an the covariance matrices are all positive definite. Given any interval I , there exist three disjoint subintervals I_j , $j = 0, 1, 2$ and points $0 < t_0 < t_1$ such that if ν assigns mass $1/2$ to the points t_0 and t_1 , the resulting contrast function (E.1) satisfies Assumption 11.*

Proof. We parametrize

$$\sigma_\theta = \begin{pmatrix} \theta_0 & \theta_1/2 \\ \theta_1/2 & \theta_2 \end{pmatrix} \quad \text{and} \quad \mu = (\theta_3, \theta_4).$$

For $x = (x_0, x_1, x_2) \in I^3$, we introduce

$$\Psi(\theta, t, x) := \{(\mathcal{F}f_B)(t, tx_k)\}_{k=0,\dots,2} = \{\exp(-\theta_0 t^2 - \theta_1 x_k t^2 - \theta_2 x_k^2 t^2 + i\theta_3 t + ix_k \theta_4 t)\}_{k=0,\dots,2}.$$

We deduce that

$$\Psi(\theta, t, x) - \Psi(\tilde{\theta}, t, x) = D(t, x, \theta, \tilde{\theta}) W(t, x) (\theta - \tilde{\theta}), \quad (\text{E.2})$$

for any $\theta, \tilde{\theta} \in \Theta$ where $D(t, x, \theta, \theta')$ denotes the diagonal 3×3 -matrix with the entries

$$D_{k,k}(t, x, \theta, \tilde{\theta}) = \int_0^1 \Psi_k(\lambda\theta + (1-\lambda)\tilde{\theta}, t, x) d\lambda, \quad k = 0, 1, 2, \quad (\text{E.3})$$

and $W(t, x)$ denotes the 3×5 -matrix

$$W(t, x) = \begin{pmatrix} -t^2 & -x_0 t^2 & -x_0^2 t^2 & it & ix_0 t \\ -t^2 & -x_1 t^2 & -x_1^2 t^2 & it & ix_1 t \\ -t^2 & -x_2 t^2 & -x_2^2 t^2 & it & ix_2 t \end{pmatrix}.$$

Writing $V(t, x, \theta, \tilde{\theta}) := D(t, x, \theta, \tilde{\theta}) W(t, x)$ and $\psi(\theta, \tilde{\theta}, t_0, t_1, x) := \sum_{j=0}^1 |\Psi(\theta, t_j, x) - \Psi(\tilde{\theta}, t_j, x)|^2$, we deduce by (E.2) that

$$\psi(\theta, \tilde{\theta}, s, t, x) = (\theta - \tilde{\theta})' U(t_0, t_1, x, \theta, \tilde{\theta}) (\theta - \tilde{\theta}), \quad (\text{E.4})$$

where the 5×5 -matrix $U(t_0, t_1, x, \theta, \tilde{\theta}) := \sum_{j=0}^1 \{V(t_j, x, \theta, \tilde{\theta})\}^\dagger V(t_j, x, \theta, \tilde{\theta})$ is Hermitian and positive semi-definite.

For a vector $y \in \mathbb{C}^5$ such that $U(t_0, t_1, x, \theta, \tilde{\theta}) y = 0$, it follows then that

$$0 = y^\dagger U(t_0, t_1, x, \theta, \tilde{\theta}) y = \sum_{j=0}^1 |V(t_j, x, \theta, \tilde{\theta}) y|^2,$$

and, hence, $D(t_j, x, \theta, \tilde{\theta}) W(t_j, x) y = 0$, $j = 0, 1$.

Fix some $x_0 < x_1 < x_2$ in I . Then, considering the diagonal entries (E.3), we may choose $T > 0$ sufficiently small such that, for all $t < T$ and $k = 0, 1, 2$, we have $D_{k,k}(t, x, \theta, \tilde{\theta}) \neq 0$. Note that T is uniform with respect to $\theta, \tilde{\theta} \in \Theta$ thanks to the compactness of Θ . Thus the matrices $D(t_j, x, \theta, \tilde{\theta})$, $j = 0, 1$, are invertible and we conclude that

$$W(t_j, x) y = 0 \quad j = 0, 1, \quad (\text{E.5})$$

whenever $t_j \in (0, T)$ for $j = 0, 1$ and $t_0 \neq t_1$. Combining all three equalities from (E.5) for $j = 0$ with the first two equalities for $j = 1$ we obtain that y lies in the kernel of the 5×5 -matrix

$$\begin{pmatrix} -t_0^2 & -x_0 t_0^2 & -x_0^2 t_0^2 & it_0 & ix_0 t_0 \\ -t_0^2 & -x_1 t_0^2 & -x_1^2 t_0^2 & it_0 & ix_1 t_0 \\ -t_0^2 & -x_2 t_0^2 & -x_2^2 t_0^2 & it_0 & ix_2 t_0 \\ -t_1^2 & -x_0 t_1^2 & -x_0^2 t_1^2 & it_1 & ix_0 t_1 \\ -t_1^2 & -x_1 t_1^2 & -x_1^2 t_1^2 & it_1 & ix_1 t_1 \end{pmatrix}. \quad (\text{E.6})$$

Applying appropriate rank-invariant linear transformations the matrix in (E.6) changes to

$$\begin{pmatrix} 1 & x_0 & x_0^2 & 0 & 0 \\ 1 & x_1 & x_1^2 & 0 & 0 \\ 1 & x_2 & x_2^2 & 0 & 0 \\ \tau & x_0\tau & x_0^2\tau & 1-\tau & x_0(1-\tau) \\ \tau & x_1\tau & x_1^2\tau & 1-\tau & x_1(1-\tau) \end{pmatrix},$$

where $\tau = t_1/t_0 \neq 1$. The upper left 3×3 -submatrix represents a Vandermonde matrix, which is known to be invertible since $x_0 < x_1 < x_2$. Also the lower right 2×2 -submatrix has full rank as $x_1 > x_0$ so that invertibility of the matrix (E.6) follows and, hence, $y = 0$. Thus we have shown that the matrix $U(t_0, t_1, x, \theta, \tilde{\theta})$ is invertible and, therefore, (strictly) positive definite under the imposed restrictions on the t_j , $j = 0, 1$, and x_k , $k = 0, 1, 2$. Now we consider the random vector $X = x + hU$ for some $h > 0$ and $U = (U_1, U_2, U_3)$ where the components of U are independent random variables which are uniformly distributed on the interval $[-1, 1]$. Let us define

$$\begin{aligned} \Psi_h(\theta, t) &:= E\Psi(\theta, t, X), \\ \psi_h(\theta, \tilde{\theta}, t_0, t_1) &:= \sum_{j=0}^1 |\Psi_h(\theta, t_j) - \Psi_h(\tilde{\theta}, t_j)|^2, \end{aligned}$$

so that

$$\psi_h(\theta, \tilde{\theta}, s, t) = (\theta - \theta'^{\dagger} U_h(t_0, t_1, \theta, \tilde{\theta})) (\theta - \tilde{\theta}), \quad (\text{E.7})$$

where

$$U_h(t_0, t_1, \theta, \tilde{\theta}) := \sum_{j=0}^1 \{EV(t_j, X, \theta, \tilde{\theta})\}^{\dagger} EV(t_j, X, \theta, \theta'),$$

which follows analogously as (E.4). The matrix $U_h(t_0, t_1, \theta, \tilde{\theta})$ is Hermitian and positive semi-definite. For $j = 0, 1$, the parameterizations $(x, \theta, \tilde{\theta}) \mapsto V(t_j, x, \theta, \tilde{\theta})$ represent continuous mappings with respect to the Frobenius norm $\|\cdot\|_F$ on the codomain. The domain of these parameterizations is restricted to the compact set $(x + [-h, h]^3) \times \Theta^2$. Since Θ is compact the Heine-Cantor theorem yields that

$$\lim_{h \downarrow 0} \sup_{\theta, \tilde{\theta} \in \Theta} \sup_{|y-x| \leq h} \|V(t_j, x, \theta, \tilde{\theta}) - V(t_j, y, \theta, \theta')\|_F = 0.$$

By the submultiplicativity of the Frobenius norm we deduce that

$$\begin{aligned} \|U_h(t_0, t_1, \theta, \tilde{\theta}) - U(t_0, t_1, x, \theta, \tilde{\theta})\|_F &\leq \sum_{j=0}^1 (\|V(t_j, x, \theta, \tilde{\theta})\|_F + E\|V(t_j, X, \theta, \theta')\|_F) \\ &\quad \cdot E\|V(t_j, x, \theta, \tilde{\theta}) - V(t_j, X, \theta, \tilde{\theta})\|_F, \end{aligned}$$

so that

$$\lim_{h \downarrow 0} \sup_{\theta, \tilde{\theta} \in \Theta} \|U_h(t_0, t_1, \theta, \tilde{\theta}) - U(t_0, t_1, x, \theta, \tilde{\theta})\|_F = 0,$$

As the invertible 5×5 -matrices form an open set with respect to the Frobenius norm we can select $h > 0$ sufficiently small – independently of $\theta, \tilde{\theta} \in \Theta$ – such that the matrix $U_h(t_0, t_1, \theta, \tilde{\theta})$ is positive definite. Simultaneously we may arrange $h > 0$ small enough such that $[x_j - h, x_j + h] \subset I$, and that these intervals are

disjoint. Since, for all $\theta_0, \theta'_0, \theta_1, \theta'_1 \in \Theta$, we have

$$\begin{aligned} \|U_h(t_0, t_1, \theta_0, \theta'_0) - U_h(t_0, t_1, \theta_1, \theta'_1)\|_F &\leq \sum_{j=0}^1 \left(\sup_{|y-x| \leq h} \|V(t_j, y, \theta_0, \theta'_0)\|_F + \sup_{|y-x| \leq h} \|V(t_j, y, \theta_1, \theta'_1)\|_F \right) \\ &\quad \cdot \sup_{|y-x| \leq h} \|V(t_j, y, \theta_0, \theta'_0) - V(t_j, y, \theta_1, \theta'_1)\|_F, \end{aligned}$$

the mapping $(\theta, \tilde{\theta}) \mapsto U_h(t_0, t_1, \theta, \tilde{\theta})$ is continuous on the compact domain Θ^2 thanks to the continuity of the parameterization of the matrix $V(t_j, \dots)$ as mentioned above. The smallest and the largest eigenvalues of $U_h(t_0, t_1, \theta, \tilde{\theta})$ are denoted by $r_h(t_0, t_1, \theta, \tilde{\theta})$ and $R_h(t_0, t_1, \theta, \tilde{\theta})$, respectively. As

$$\begin{aligned} |r_h(t_0, t_1, \theta_0, \theta'_0) - r_h(t_0, t_1, \theta_1, \theta'_1)| &\leq \|U_h(t_0, t_1, \theta_0, \theta'_0) - U_h(t_0, t_1, \theta_1, \theta'_1)\|_F, \\ |R_h(t_0, t_1, \theta_0, \theta'_0) - R_h(t_0, t_1, \theta_1, \theta'_1)| &\leq \|U_h(t_0, t_1, \theta_0, \theta'_0) - U_h(t_0, t_1, \theta_1, \theta'_1)\|_F, \end{aligned}$$

the functions r_h and R_h are continuous with respect to $(\theta, \tilde{\theta})$. The compactness of Θ provides that

$$\inf_{\theta, \tilde{\theta} \in \Theta} r_h(t_0, t_1, \theta, \theta') > 0, \quad \sup_{\theta, \tilde{\theta} \in \Theta} R_h(t_0, t_1, \theta, \tilde{\theta}) < \infty.$$

Putting $I_{j+1} = [x_j - h, x_j + h]$, $q = 3$, ν equal to the uniform probability measure on the discrete set $\{t_0, t_1\}$ we obtain that

$$8h^2 \|\Phi(\theta, \cdot) - \Phi(\tilde{\theta}, \cdot)\|_{W; q}^2 = \psi_h(\theta, \tilde{\theta}, t_0, t_1),$$

so that which suffices to verify Assumption 11. □

F Estimation: Technical assumptions and proofs

F.1 Preliminary results on order statistics

Lemma F.1. *Let $\Gamma(x)$ denote the gamma function. For any $\kappa \geq 2$ and $n \geq 2$, we have*

(a) *under Assumption 8 that*

$$\begin{aligned} E \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^\kappa &\leq n(n-1)^{-\kappa} c_Z^{-\kappa} \kappa \Gamma(\kappa), \\ \max \{E(1 - Z_{(n)})^\kappa, E(Z_{(1)} + 1)^\kappa\} &\leq \kappa c_Z^{-\kappa} n^{-\kappa} \Gamma(\kappa), \end{aligned}$$

(b) *under Assumption 4 that for $M > 1$,*

$$\begin{aligned} E \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^\kappa \cdot 1_{-M \leq Z_{(j)} < Z_{(j+1)} < M} &\leq \kappa c_Z^{-\kappa} \Gamma(\kappa) M^{\kappa\gamma} n(n-1)^{-\kappa}, \\ \max \{P(Z_{(n)} \leq M), P(Z_{(1)} \geq -M)\} &\leq \exp(-n c_Z M^{1-\gamma} / (\gamma - 1)). \end{aligned}$$

Proof of Lemma F.1. : First, for the second part of (a),

$$\begin{aligned}
E(1 - Z_{(n)})^\kappa &= \int_0^\infty P[1 - Z_{(n)} > t^{1/\kappa}] dt = \kappa \int_{-1}^1 P[Z_{(n)} < s] (1 - s)^{\kappa-1} ds \\
&= \kappa \int_{-1}^1 \left(1 - \int_s^1 f_Z(u) du\right)^n (1 - s)^{\kappa-1} ds \leq \kappa \int_{-1}^1 \exp(-nc_Z(1 - s)) (1 - s)^{\kappa-1} ds \\
&\leq \kappa \int_0^\infty \exp(-nc_Z t) t^{\kappa-1} dt = \kappa c_Z^{-\kappa} n^{-\kappa} \Gamma(\kappa),
\end{aligned}$$

and the bound for $E(Z_{(1)} + 1)^\kappa$ follows analogously. For the second part of (b), we compute

$$\begin{aligned}
P(Z_{(n)} \leq M) &= \left(1 - \int_M^\infty f_Z(z) dz\right)^n \leq \left(1 - c_Z \int_M^\infty z^{-\gamma} dz\right)^n \\
&= \left(1 - c_Z M^{1-\gamma}/(\gamma - 1)\right)^n \leq \exp(-nc_Z M^{1-\gamma}/(\gamma - 1))
\end{aligned}$$

Second, we consider jointly the first parts of (a) as well as of (b), in case (a) we set $M = 1$. Let $\mathcal{R} = (-M, M)$. By exchanging the order in the sums we deduce that

$$\sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^\kappa \cdot 1_{\mathcal{R}}(Z_{(j)}) \cdot 1_{\mathcal{R}}(Z_{(j+1)}) = \sum_{j=1}^n (Z_j^* - Z_j)^\kappa \cdot 1_{\mathcal{R}}(Z_j) \cdot 1_{\mathcal{R}}(Z_j^*),$$

holds almost surely where

$$Z_j^* := \begin{cases} Z_j, & \text{if } Z_j \geq Z_k, \forall k = 1, \dots, n, \\ \min\{Z_k : Z_k > Z_j\}, & \text{otherwise.} \end{cases}$$

Thus we have

$$\begin{aligned}
E \sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^\kappa \cdot 1_{\mathcal{R}}(Z_{(j)}) \cdot 1_{\mathcal{R}}(Z_{(j+1)}) &= \sum_{j=1}^n E 1_{\mathcal{R}}(Z_j) E(1_{\mathcal{R}}(Z_j^*) \cdot (Z_j^* - Z_j)^\kappa \mid Z_j) \\
&= \sum_{j=1}^n E 1_{\mathcal{R}}(Z_j) \int_0^{(M-Z_j)^\kappa} P[Z_j^* > Z_j + t^{1/\kappa} \mid Z_j] dt \\
&\leq \sum_{j=1}^n E 1_{\mathcal{R}}(Z_j) \int_0^{(M-Z_j)^\kappa} P[Z_k \notin (Z_j, Z_j + t^{1/\kappa}), \forall k \neq j \mid Z_j] dt \\
&\leq n \int_{-M}^M \int_0^{M-z} \left(1 - \int_z^{z+s} f_Z(x) dx\right)^{n-1} \kappa s^{\kappa-1} ds f_Z(z) dz \\
&\leq n \int_{-M}^M \int_0^{M-z} \exp(-(n-1)c_Z M^{-\gamma} s) \kappa s^{\kappa-1} ds f_Z(z) dz \\
&\leq n \int_{-M}^M \int_0^\infty \exp(-(n-1)c_Z M^{-\gamma} s) \kappa s^{\kappa-1} ds f_Z(z) dz \\
&\leq n \int_0^\infty \exp(-(n-1)c_Z M^{-\gamma} s) \kappa s^{\kappa-1} ds \\
&= \kappa c_Z^{-\kappa} \Gamma(\kappa) M^{\kappa\gamma} n(n-1)^{-\kappa}.
\end{aligned}$$

□

G Linear instrumental variables

In this subsection, we extend the result for IV estimation to bivariate W .

Consider

$$\begin{aligned} Y &= B_0 + B_1 X + B_{2,1} W_1 + B_{2,2} W_2, \\ X &= A_0 + A_1 Z + A_{2,1} W_1 + A_{2,2} W_2. \end{aligned} \tag{G.1}$$

Further we assume the regressors to be normalized:

$$EZ = EX = EW_j = 0, \quad j = 1, 2, \quad EW_1 W_2 = 0. \tag{G.2}$$

Assumption 12 (*Independence and moment assumption*). Suppose that $B = (B_0, B_1, B_2)'$ and A_1 are independent, and that A_1^{-1} is absolutely integrable.

Proposition 4. Assume that (Y, X, Z, W_1, W_2) follow model (G.1), for which we maintain the Assumptions 2 (exogeneity) and 12 (independence assumption). If the random coefficients and the covariates Z, W_1, W_2 have finite second moments, the covariates are standardized as in (G.2), and

$$EA_1 (EZ^2 EW_1^2 EW_2^2 - EW_1^2 (EZW_2)^2 - EW_2^2 (EZW_1)^2) \neq 0,$$

then the linear IV estimator based on an i.i.d. sample from (Y, X, Z, W_1, W_2) converges at rate $n^{-1/2}$ to

$$\mu_{IV} = \begin{pmatrix} EB_0 + EB_1 A_0 \\ EB_1 \\ EB_{2,1} + EB_1 A_{2,1} - EB_1 EA_{2,1} \\ EB_{2,2} + EB_1 A_{2,2} - EB_1 EA_{2,2} \end{pmatrix}$$

In particular, the linear IV estimate is consistent for EB_1 , and it is consistent for $EB_{2,2}$ if the coefficient $A_{2,2}$ of W_2 in the first equation is deterministic.

In (G.2) is not yet satisfied, we pass to

$$\begin{aligned} \tilde{X} &= X - EX, \quad \tilde{Z} = Z - EZ, \quad \tilde{W}_1 = W_1 - EW_1, \\ \tilde{W}_2 &= \bar{W}_2 - \frac{E(\tilde{W}_1 \bar{W}_2)}{E(\tilde{W}_1)^2} \tilde{W}_1, \quad \text{where } \bar{W}_2 = W_2 - EW_2, \end{aligned}$$

which satisfies

$$\begin{aligned} Y &= \tilde{B}_0 + B_1 \tilde{X} + \tilde{B}_{2,1} \tilde{W}_1 + B_{2,2} \tilde{W}_2, \\ X &= \tilde{A}_0 + A_1 \tilde{Z} + \tilde{A}_{2,1} \tilde{W}_1 + A_{2,2} \tilde{W}_2, \end{aligned}$$

where

$$\begin{aligned} \tilde{B}_0 &= B_0 + B_1(EX) + B_{2,1}(EW_1) + B_{2,2}(EW_2), & \tilde{B}_{2,1} &= B_{2,1} + \frac{E(\tilde{W}_1 \bar{W}_2)}{E(\tilde{W}_1)^2} B_{2,2}, \\ \tilde{A}_0 &= A_0 + A_1(EZ) + A_{2,1}(EW_1) + A_{2,2}(EW_2), & \tilde{A}_{2,1} &= A_{2,1} + \frac{E(\tilde{W}_1 \bar{W}_2)}{E(\tilde{W}_1)^2} A_{2,2}. \end{aligned}$$

This system still satisfies Assumptions 2 and 12, and the coefficients $B_1, A_1, B_{2,2}$ and $A_{2,2}$ remain unchanged, so that the above consistency still applies.

Proof of Proposition 4. Assume (G.2), and let $V = E(Y, YZ, YW_1, YW_2)'$ and

$$M = E(1, Z, W_1, W_2)'(1, X, W_1, W_2) = E \begin{pmatrix} 1 & X & W_1 & W_2 \\ Z & ZX & ZW_1 & ZW_2 \\ W_1 & W_1X & W_1^2 & W_1W_2 \\ W_2 & W_2X & W_1W_2 & W_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \widetilde{M} \end{pmatrix},$$

where

$$\widetilde{M} = \begin{pmatrix} EXZ & EZW_1 & EZW_2 \\ EW_1X & EW_1^2 & 0 \\ EW_2X & 0 & EW_2^2 \end{pmatrix}. \quad (\text{G.3})$$

Under our assumptions, linear IV consistently estimates at rate $n^{-1/2}$ the vector $M^{-1}V$, provided the matrix M is invertible. Using exogeneity and normalization, we immediately compute

$$V = \begin{pmatrix} EB_0 + EB_1X \\ EB_1XZ + EB_{2,1}EZW_1 + EB_{2,2}EZW_2 \\ EB_1XW_1 + EB_{2,1}EW_1^2 \\ EB_1XW_2 + EB_{2,2}EW_2^2 \end{pmatrix} \quad (\text{G.4})$$

Now, using the independence assumption,

$$\begin{aligned} EB_1X &= EB_1A_0, \\ EB_1XZ &= EA_1EB_1EZ^2 + EB_1A_{2,1}EZW_1 + EB_1A_{2,2}EZW_2, \\ EB_1XW_1 &= EA_1EB_1EZW_1 + EB_1A_{2,1}EW_1^2, \\ EB_1XW_2 &= EA_1EB_1EZW_2 + EB_1A_{2,2}EW_2^2. \end{aligned} \quad (\text{G.5})$$

Similarly, for \widetilde{M} we compute

$$\begin{aligned} EXZ &= EA_1EZ^2 + EA_{2,1}EZW_1 + EA_{2,2}EZW_2, \\ EXW_1 &= EA_1EZW_1 + EA_{2,1}EW_1^2, \\ EXW_2 &= EA_1EZW_2 + EA_{2,2}EW_2^2. \end{aligned} \quad (\text{G.6})$$

Inserting (G.6) into (G.3) we compute

$$\det \widetilde{M} = d = EA_1 (EZ^2EW_1^2EW_2^2 - EW_1^2(EZW_2)^2 - EW_2^2(EZW_1)^2),$$

so that \widetilde{M} is invertible and

$$\widetilde{M}^{-1} = d^{-1}.$$

$$\begin{pmatrix} EW_1^2EW_2^2 & -EW_1ZEW_2^2 & -EW_2ZEW_1^2 \\ -EW_2^2(EA_1EZW_1 + EA_{2,1}EW_1^2) & EA_1(EZ^2EW_2^2 - (EZW_2)^2) + EA_{2,1}EZW_1EW_2^2 & EW_2Z(EA_1EZW_1 + EA_{2,1}EW_1^2) \\ -EW_1^2(EA_1EZW_2 + EA_{2,2}EW_2^2) & EW_1Z(EA_1EZW_2 + EA_{2,2}EW_2^2) & EA_1(EZ^2EW_1^2 - (EZW_1)^2) + EA_{2,2}EZW_2EW_1^2 \end{pmatrix}$$

Further, inserting (G.5) into (G.4), $M^{-1}V$ is simplified after some calculations into μ_{IV} . \square

H Estimation including exogenous regressors

Finally, we discuss how to extend the estimators from the previous sections to the model

$$\begin{aligned} Y &= B_0 + B_1 X + B_2 W, \\ X &= A_0 + A_1 Z + A_2 W. \end{aligned} \tag{H.1}$$

Assumption 13 (*Exogeneity*). (Z, W) and $(A', B)'$ are independent.

We maintain the identification Assumptions 13 and 12.

H.0.1 Nonparametric estimation

First consider the nonparametric setting as in Section C. Following HKM, it is quite straightforward to extend a Nadaraya-Watson approach to the case of exogenous covariates W , as follows:

$$\begin{aligned} \check{g}(b_0, b_1, b_2; h, M) &= \sum_{j=1}^n K_2(Y_j - b_0 - b_1 X_j - b_2 W_j; h) \lambda_{j,NW}, \\ \lambda_{j,NW} &= 1_{M^{-1} \leq \hat{f}_{Z,W}(Z_j, W_j)} (n \hat{f}_{Z,W}(Z_j, W_j))^{-1} \end{aligned} \tag{H.2}$$

where $\hat{f}_{Z,W}$ is a density estimator and $M \rightarrow \infty$ a trimming parameter, and

$$K_2(x; h) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}} |t|^2 \exp(itx) w(ht) dt.$$

The estimator for the scaling constant also needs to be modified in a similar fashion. Although conceptually simple, the Nadaraya-Watson weights $\lambda_{j,NW}$ require additional smoothing and trimming parameters and will be quite unstable. At least for semiparametric estimation and bounded support, we rather recommend extensions of the Priestly-Chao weights, as discussed below.

H.0.2 Semiparametric estimation and bounded support

Assume that the support $\mathcal{S}_{Z,w} = I_Z$ of Z given $W = w$ is a compact interval, independent of w (the conditional distribution itself may depend on w), and that the support of W is the compact interval I_W . Further, impose the support restriction for a rectangle $I_X \times I_W \subseteq \text{supp}(X, W)$, i.e.

$$\text{supp} \left(\frac{x - A_0 - w A_2}{A_1} \right) \subseteq \text{supp}(Z|W = w) = I_Z, \quad \forall (x, w)' \in I_X \times I_W. \tag{H.3}$$

Moreover, for the joint density of (Z, W) we assume that $f_{Z,W}(z, w) \geq c > 0$ for all $z \in [-1, 1]$, $w \in I_W$, for some $c > 0$.

H.0.3 Estimation of the scaling constant

From HHM, Theorem 5, for all $(x, w)' \in I_X \times I_W$,

$$E|A_1|^{-1} = \int_{-1}^1 f_{X|Z,W}(x|z, w) dz.$$

Choose a bounded weight function $v : \mathbb{R}^2 \rightarrow (0, \infty)$ with $\text{supp } v \subseteq I_X \times I_W$ and $\int v = 1$, then we have that

$$E|A_1|^{-1} = \int_{I_X \times I_W} v(x, w) \int_{-1}^1 f_{X|Z,W}(x|z, w) dz dx dw.$$

The Nadaraya-Watson type estimator of this expression is

$$\hat{a}_{I_X \times I_W, n} = \sum_{j=1}^n v(X_j, W_j) \lambda_{j, NW}.$$

Note that we do not need the trimming parameter $M \rightarrow \infty$ for the weights $\lambda_{j, NW}$ in (H.2), at least asymptotically, since we assume that the density $f_{Z,W}$ is bounded away from zero. In the bounded support context we recommend the use of the following Priestly-Chao type weights

$$\lambda_{j, PC} = \text{Area} \left\{ (z, w) \in I_Z \times I_W : |(z, w) - (Z_j, W_j)| \leq |(z, w) - (Z_k, W_k)|, \forall k = 1, \dots, n \right\},$$

$j = 1, \dots, n$. In the univariate situation of Section C without W , this corresponds to the weight $\lambda_{j, PC} = (Z_{(j+1)} - Z_{(j-1)})/2$, which gives the same results asymptotically as $Z_{(j+1)} - Z_{(j)}$ as chosen in that section. In the multivariate situation it is hard to compute the $\lambda_{j, PC}$ analytically. However, it is straightforward to approximate them using Monte Carlo: for given $N \in \mathbb{N}$ (we use $N = 200$ in the simulation section), generate i.i.d. $U_1, \dots, U_{N \cdot n}$, uniform on $I_Z \times I_W$, and take $\lambda_{j, PC}$ as the proportion of all of those $U_1, \dots, U_{N \cdot n}$ closest to (Z_j, W_j) , multiplied by $\text{Area}(I_Z \times I_W)$, where Area denotes the Lebesgue area. This requires $N \cdot n^2$ comparisons. The resulting estimator of the scaling constant is

$$\hat{a}_{I_X \times I_W, n} = \sum_{j=1}^n v(X_j, W_j) \lambda_{j, PC}. \quad (\text{H.4})$$

H.0.4 Semiparametric estimator

To proceed, let $I_{k,X} \subset I_X$ and $I_{l,W} \subset I_W$ be subintervals. From Theorem 5 in HHM we obtain

$$\begin{aligned} & \int_{-1}^1 \int \exp(it y) \int_{I_{k,X}} \int_{I_{l,W}} f_{Y,X|Z,W}(y, x|z, w) dx dw dy dz \\ &= E|A_1|^{-1} \cdot \int_{I_{k,X}} \int_{I_{l,W}} (\mathcal{F}f_B)(t, tx, tw) dx dw =: \Phi(\theta, t, I_{k,X}, I_{l,W}). \end{aligned} \quad (\text{H.5})$$

We estimate the left hand side of (H.5) by

$$\hat{\Phi}_n(t, I_{k,X}, I_{l,W}) := \sum_{j=1}^n \exp(it Y_j) 1_{I_{k,X}}(X_j) 1_{I_{l,W}}(W_j) \lambda_{j, PC}.$$

Given intervals $I_{1,X}, \dots, I_{p,X} \subset I_X$ and $I_{1,W}, \dots, I_{q,W} \subset I_W$ and a probability measure ν on \mathbb{R} , we define the contrast

$$\| \hat{\Phi}(\cdot) - \Phi(\theta, \cdot) \|_{\nu; q, p}^2 := \sum_{k=1}^q \sum_{l=1}^p \int_{\mathbb{R}} | \hat{\Phi}_n(t, I_{k,X}, I_{l,W}) - \hat{a}_n \cdot \Phi(\theta, t, I_{k,X}, I_{l,W}) |^2 d\nu(t), \quad (\text{H.6})$$

where $\hat{a}_n = \hat{a}_{I_X \times I_W, n}$. The minimizer of this criterion is taken as estimator $\hat{\theta}_n$ for θ . For $B \sim N(\mu, \Sigma)$ multivariate normal, and ν the normal $N(0, s^2)$ -distribution, after dropping terms not depending on the parameters, we have to minimize

$$\begin{aligned} M(\mu, \Sigma) &= \sum_{k=1}^p \sum_{l=1}^q \left(\hat{a}_n \int_{I_{k,X} \times I_{k,X}} \int_{I_{l,W} \times I_{l,W}} \varphi(0; \mu_1(x_2 - x_1) + \mu_2(w_2 - w_1), ((1, x_1, w_1)\Sigma(1, x_1, w_1)' \right. \\ &\quad \left. + (1, x_2, w_2)\Sigma(1, x_2, w_2)' + s^{-2})^{-1}) dx_1 dx_2 dw_1 dw_2 \right. \\ &\quad \left. - 2 \sum_{j=1}^n \int_{I_{k,X}} \int_{I_{l,W}} \varphi(0; Y_j - \mu_0 - \mu_1 x - \mu_2 w, ((1, x, w)\Sigma(1, x, w)' + s^{-2})^{-1}) \right. \\ &\quad \left. 1_{I_{k,X}}(X_j) 1_{I_{l,W}}(W_j) \lambda_{j,PC} dx dw \right). \end{aligned}$$