Online Appendix: Estimation in a Generalization of Bivariate Probit Models with Dummy Endogenous Regressors

Sukjin Han
Department of Economics
University of Texas at Austin
sukjin.han@austin.utexas.edu

Sungwon Lee
Global Asia Institute
National University of Singapore
gails@nus.edu.sg

This Draft: March 4, 2019

Abstract
This online appendix contains all the proofs of the main results in Han and Lee (2018), technical assumptions, and additional simulation results. Section A contains the proofs of the identification/non-identification results. Section B collects the proofs in the asymptotic theory of the sieve ML estimators, related technical assumptions, and other expressions. Finally, Section C shows more simulation results.

A Proofs of Results in Section 2
A.1 Proof of Theorem 2.6
Recall
\[ q_0 = F_{\nu}(\mu_{\nu}/\sigma_{\nu}), \quad q_1 = F_{\nu}((1 - \mu_{\nu})/\sigma_{\nu}), \]
\[ t_0 = F_{\varepsilon}(\mu_{\varepsilon}/\sigma_{\varepsilon}), \quad t_1 = F_{\varepsilon}((1 - \mu_{\varepsilon})/\sigma_{\varepsilon}), \]
and
\[ \hat{p}_{11,0} = C(F_{\varepsilon}(F_{\varepsilon}^{-1}(t_0) + \delta_1), q_0; \rho), \]
\[ \hat{p}_{11,1} = C(F_{\varepsilon}(F_{\varepsilon}^{-1}(t_1) + \delta_1), q_1; \rho), \]
\[ \hat{p}_{10,0} = t_0 - C(t_0, q_0; \rho), \]
\[ \hat{p}_{10,1} = t_1 - C(t_1, q_1; \rho), \]
\[ \hat{p}_{00,0} = 1 - t_0 - q_0 + C(t_0, q_0; \rho), \]
\[ \hat{p}_{00,1} = 1 - t_1 - q_1 + C(t_1, q_1; \rho), \]
where \( \tilde{p}_{yd,x} \equiv \Pr[Y = y, D = d|X_1 = x] \). Again, we want to show that, given \((q_0, q_1)\) which are identified from the reduced-form equation, there are two distinct sets of parameter values \((t_0, t_1, \delta_1, \rho)\) and \((t_0^*, t_1^*, \delta_1^*, \rho^*)\) (with \((t_0, t_1, \delta_1, \rho) \neq (t_0^*, t_1^*, \delta_1^*, \rho^*)\)) that generate the same observed fitted probabilities \( \tilde{p}_{yd,0} \) and \( \tilde{p}_{yd,1} \) for all \((y, d) \in \{0, 1\}^2\). In showing this, the following lemma is useful:

**Lemma A.1.** Assumption 5 implies that, for any \((u_1, u_2) \in (0, 1)^2\) and \(\rho \in \Omega\),

\[
C_p(u_1, u_2; \rho) > 0. \tag{A.1}
\]

The proof of this lemma can be found below.

Now fix \((q_0, q_1) \in (0, 1)^2\). First, consider the fitted probability \( \tilde{p}_{10,0} \). Given \(t_0 \in (0, 1)\) and \(\rho \in \Omega\), note that, for \(\rho^* > \rho\),\(^1\) there exists a solution \(t_0^* = t_0^*(t_0, q_0, \rho, \rho^*)\) such that

\[
t_0 - C(t_0, q_0; \rho) = \Pr[u_1 \leq t_0, u_2 \geq q_0; \rho] \tag{A.2}
\]

\[
= \Pr[u_1 \leq t_0^*, u_2 \geq q_0; \rho^*] \tag{A.3}
\]

\[
= t_0^* - C(t_0^*, q_0; \rho^*),
\]

and note that by Assumption 5 and a variant of Lemma A.1, we have that \(t_0^* > t_0\). Here, \((t_0, q_0, \rho)\) and \((t_0^*, q_0, \rho^*)\) result in the same observed probability \( \tilde{p}_{10,0} = t_0 - C(t_0, q_0; \rho) = t_0^* - C(t_0^*, q_0; \rho^*) \).

Now consider the fitted probability \( \tilde{p}_{11,0} \). Choose \(\delta_1 = 0\). Also let \(F_{\tilde{x}} \sim \text{Unif}(0, 1)\) only for simplicity, which is relaxed later. Then there exists a solution \(t_0^1 = t_0^1(t_0, q_0, \rho, \rho^*)\) such that

\[
C(t_0, q_0; \rho) = \Pr[u_1 \leq t_0, u_2 \leq q_0; \rho] \tag{A.4}
\]

\[
= \Pr[u_1 \leq t_0^1, u_2 \leq q_0; \rho^*] \tag{A.5}
\]

\[
= C(t_0^1, q_0; \rho^*),
\]

and note that \(t_0^1 < t_0\) by Assumption 5 and Lemma A.1. Then, by letting \(\delta_1^* = t_0^1 - t_0^1, (t_0, q_0, \delta_1, \rho)\) and \((t_0^*, q_0, \delta_1^*, \rho^*)\) satisfy \( \tilde{p}_{11,0} = C(t_0 + 0, q_0; \rho) = C(t_0^* + \delta_1^*, q_0; \rho^*) \). Lastly, note that \(\tilde{p}_{00,0} = 1 - q_0 - \tilde{p}_{10,0}\) and \(\tilde{p}_{01,0} = q_0 - \tilde{p}_{11,0}\), and so \((t_0, \delta_1, \rho)\) and \((t_0^*, \delta_1^*, \rho^*)\) above will also result in the same values of \(\tilde{p}_{00,0}\) and \(\tilde{p}_{01,0}\).

It is tempting to have a parallel argument for \(\tilde{p}_{10,1}, \tilde{p}_{11,1}, \tilde{p}_{00,1}, \) and \(\tilde{p}_{01,1}\), but there is a complication. Although other parameters are not, \(\delta_1\) and \(\rho\) are common in both sets of probabilities. Therefore, we proceed as follows. First, consider \(\tilde{p}_{10,1}\). Given \(t_1 \in (0, 1)\) and the above choice of

\(^1\)The inequality here and other inequalities implied from this (e.g., \(t_0^* > t_0\), and etc.) are assumed only for concreteness.
\( \rho^* \in \Omega \), note that there exists a solution \( t_1^* = t_1^*(t_1, q_1, \rho, \rho^*) \) such that
\[
t_1 - C(t_1, q_1; \rho) = \Pr[u_1 \leq t_1, u_2 \geq q_1; \rho] = \Pr[u_1 \leq t_1^*, u_2 \geq q_1; \rho^*] = t_1^* - C(t_1^*, q_1; \rho^*),
\]
and similarly as before, we have \( t_1^* > t_1 \). Here, \( (t_1, q_1, \rho) \) and \( (t_1^*, q_1, \rho^*) \) result in the same observed probability \( \tilde{p}_{10,1} = t_1 - C(t_1, q_1; \rho) = t_1^* - C(t_1^*, q_1; \rho^*) \). Now consider \( \tilde{p}_{11,1} \). Recall \( \delta_1 = 0 \) and \( F_\varepsilon \sim Unif(0,1) \). Then there exists a solution \( t_1^\dagger = t_1^\dagger(t_1, q_1, \rho, \rho^*) \) such that
\[
C(t_1, q_1; \rho) = \Pr[u_1 \leq t_1, u_2 \leq q_1; \rho] = \Pr[u_1 \leq t_1^\dagger, u_2 \leq q_1; \rho^*] = C(t_1^\dagger, q_1; \rho^*),
\]
and thus \( t_1^\dagger < t_1 \). Then, if we can show that
\[
t_1^\dagger = t_1^* + \delta_1^*,
\]
where \( t_1^* \) and \( \delta_1^* \) are the values already determined above, then \( (t_1, q_1, \delta_1, \rho) \) and \( (t_1^*, q_1, \delta_1^*, \rho^*) \) result in \( \tilde{p}_{11,1} = C(t_1 + 0, q_1; \rho) = C(t_1^* + \delta_1^*, q_1; \rho^*) \). Then similar as before, the two sets of parameters will generate the same values of \( \tilde{p}_{00,1} = 1 - q_1 - \tilde{p}_{10,1} \) and \( \tilde{p}_{01,1} = q_1 - \tilde{p}_{11,1} \). Consequently, \( (t_0, t_1, q_0, q_1, \delta, \rho) \) and \( (t_0^*, t_1^*, q_0, q_1, \delta_1^*, \rho^*) \) generate the same entire observed fitted probabilities. The remaining question is whether we can find \( (t_0, t_1, \delta_1, \rho) \) and \( (t_0^*, t_1^*, \delta_1^*, \rho^*) \) such that (A.10) holds.

To show this, we choose further specifications. We assume a normal copula.\(^2\) We choose \( \rho = 0, \rho^* = 1, q_0 = t_0 = 1/3, \) and \( q_1 = t_1 = 2/3 \). Since \( (U_1, U_2) \) are jointly uniform, note that when \( \rho = 0 \), the probability of the quadrant in \([0,1]^2\) specified by each of (A.2), (A.4), (A.6), and (A.8) equals the volume of the quadrant. When \( \rho^* = 1 \), all the probability mass lies on the 45 degree line in \([0,1]^2\) and no where else, so the probability of a quadrant specified by each of (A.3), (A.5), (A.7), and (A.9) equals the length of the 45 line which intersects with that quadrant.

\(^2\)This choice is not critical except that we can have \( \rho \) reach to 1.
Suppose that the following observational equivalence holds:

\[
\begin{align*}
\Pr[u_1 \leq t_0, u_2 \geq q_0; \rho] &= \Pr[u_1 \leq t_0^*, u_2 \geq q_0; \rho^*] = 2/9, \\
\Pr[u_1 \leq t_0, u_2 \leq q_0; \rho] &= \Pr[u_1 \leq t_0^\dagger, u_2 \leq q_0; \rho^*] = 1/9, \\
\Pr[u_1 \leq t_1, u_2 \geq q_1; \rho] &= \Pr[u_1 \leq t_1^*, u_2 \geq q_1; \rho^*] = 2/9, \\
\Pr[u_1 \leq t_1, u_2 \leq q_1; \rho] &= \Pr[u_1 \leq t_1^\dagger, u_2 \leq q_1; \rho^*] = 4/9.
\end{align*}
\]

One can easily show that these equations yield that \(t_0^* = 5/9, t_0^\dagger = 1/9, t_1^* = 8/9, \) and \(t_1^\dagger = 4/9.\)

Consider the equation (A.10), which can be rewritten as \(t_1^\dagger = t_1^* + t_0^\dagger - t_0^*\) or \(t_1^\dagger - t_1^* = t_0^\dagger - t_0^*.\)

Then, note that we have \(t_1^\dagger - t_1^* = t_0^\dagger - t_0^* = -4/9,\) which is, in fact, the value of \(\delta_1^*.\) In sum, the values of parameters that give the observationally equivalent fitted probabilities are

\[
(t_0, t_1, q_0, q_1, \delta_1, \rho) = \left(\frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, 0, 0\right), \quad (A.11)
\]

\[
(t_0^*, t_1^*, q_0, q_1, \delta_1^*, \rho^*) = \left(\frac{5}{9}, \frac{8}{9}, \frac{1}{3}, \frac{2}{3}, \frac{4}{9}, 1\right). \quad (A.12)
\]

This argument can be made slightly more general, and thus the counterexample more realistic, by relaxing \(F_{\xi} \sim \text{Unif}(0, 1)\) and \(\rho^* = 1.\) We show that a similar argument goes through with \(F_{\xi}\) being a general distribution function with a symmetric density function, and \(-1 \leq \rho^* \leq 1\) as long as the copula density is symmetric around \(u_2 = u_1\) (i.e., the 45 degree line) and \(u_2 = 1 - u_1.\)

Let \(F \equiv F_{\xi}\) be a general distribution whose density function is symmetric. Then there exists a solution \(s_0^\dagger = s_0^\dagger(t_0, q_0, \rho, \rho^*)\) such that

\[
C(F(F^{-1}(t_0) + 0), q_0; \rho) = \Pr[u_1 \leq t_0, u_2 \leq q_0; \rho]
\]

\[
= \Pr[u_1 \leq s_0^\dagger, u_2 \leq q_0; \rho^*]
\]

\[
= C(s_0^\dagger, q_0; \rho^*).
\]

Then, by letting \(\delta_1^* = F^{-1}(s_0^\dagger) - F^{-1}(t_0^\dagger),\) we have \(s_0^\dagger = F(F^{-1}(t_0^\dagger) + \delta_1^*)\) and therefore \((t_0, q_0, \delta_1, \rho)\) and \((t_0^*, q_0^*, \delta_1^*, \rho^*)\) result in \(p_{11,x} = C(F(F^{-1}(t_0) + 0), q_0; \rho) = C(F(F^{-1}(t_0^*) + \delta_1^*), q_0; \rho^*).\) Suppose that \(\delta_1 = 0.\) Then there exists a solution \(s_1^\dagger = s_1^\dagger(t_1, q_1, \rho, \rho^*)\) such that

\[
C(F(F^{-1}(t_1) + 0), q_1; \rho) = \Pr[u_1 \leq t_1, u_2 \leq q_1; \rho]
\]

\[
= \Pr[u_1 \leq s_1^\dagger, u_2 \leq q_1; \rho^*]
\]

\[
= C(s_1^\dagger, q_1; \rho^*).
\]

Then, if we can show that

\[
F^{-1}(s_1^\dagger) = F^{-1}(t_1^*) + \delta_1^*,
\]


then \( s^1_1 = F(F^{-1}(t^*_1) + \delta^*_1) \) and therefore \((t_1, q_1, \delta_1, \rho)\) and \((t^*_1, q_1, \delta^*_1, \rho^*)\) result in \( \bar{p}_{11,1} = C(F(F^{-1}(t_1) + 0), q_1; \rho) \). Note \( F^{-1}(s^1_1) = F^{-1}(t^*_1) + \delta^*_1 \) can be rewritten as \( F^{-1}(s^1_1) = F^{-1}(t^*_1) + F^{-1}(s^1_0) - F^{-1}(t^*_0) \) or

\[
F^{-1}(s^1_1) - F^{-1}(t^*_1) = F^{-1}(s^1_0) - F^{-1}(t^*_0). \tag{A.13}
\]

But note that since the density of \( F \) is symmetric, any two values \( s \) and \( \bar{s} \) in \((0, 1)\) that are symmetric around \( u_1 = 1/2 \) will satisfy \( F^{-1}(s) = -F^{-1}(\bar{s}) \). Therefore, since in our example \( s^1_0 \) and \( t^*_1 \) are symmetric around \( u_1 = 1/2 \), and so are \( s^1_1 \) and \( t^*_0 \), we have the desired result (A.13), and the counterexample (A.11)–(A.12) remains valid. Note that the symmetry of the density function of \( F \) plays a key role here; the uniform distribution trivially satisfies the condition as does the normal distribution.

The above counter-example to identification involves a parameter on the boundary of the parameter space \((\rho^* = 1)\), while the identification results in the paper assume that the parameter space is open and thus that \( \rho \in (-1, 1) \). We now show that the key idea of the argument remains the same with \(-1 < \rho^* < 1\). Suppose that the copula density is symmetric around \( u_2 = u_1 \) and \( u_2 = 1 - u_1 \). The normal copula satisfies this condition for any \( \rho \in (-1, 1) \). Because of this condition, the symmetry of \( s^1_0 \) and \( t^*_1 \) (and of \( s^1_1 \) and \( t^*_0 \)) around \( u_1 = 1/2 \) does not break at a different value of \( \rho^* \), even though the values of \( s^1_0, t^*_1, s^1_1 \), and \( t^*_0 \) themselves change. Therefore, (A.13) continues to hold with \( \rho^* \neq 1 \).

### A.2 Proof of Lemma A.1

The proof of Lemma A.1 is a slight modification of the proof of Theorem 2.14 of Joe (1997, p. 44). Suppose \( C_{2|1} \prec_S \hat{C}_{2|1} \). Let \((U_1, U_2) \sim C, (\bar{U}_1, \bar{U}_2) \sim \hat{C}\), with \( U_j \overset{d}{=} \bar{U}_j, j = 1, 2 \). By Theorem 2.9 of Joe (1997, p. 40), \((U_1, U_2) \overset{d}{=} (\bar{U}_1, \psi(U_1, U_2)) \) with \( \psi(u_1, u_2) = \hat{C}^{-1}_{2|1}(C_{2|1}(u_2|u_1)|u_1) \). Since \( C_{2|1} \prec_S \hat{C}_{2|1} \), \( \psi \) is increasing in \( u_1 \) and \( u_2 \). We consider two cases:

- **Case 1:** Suppose that \( u_1 \) and \( u_2 \) are such that \( \psi(u_1, u_2) \leq u_2 \). Then

\[
\hat{C}(u_1, u_2) = \Pr[\bar{U}_1 \leq u_1, \bar{U}_2 \leq u_2] = \Pr[\bar{U}_1 < u_1, \bar{U}_2 < u_2] \\
= \Pr[U_1 < u_1, \psi(U_1, U_2) < u_2] \geq \Pr[U_1 < u_1, \psi(U_1, U_2) < u_2] \\
> \Pr[U_1 < u_1, U_2 < u_2] = C(u_1, u_2)
\]

where the strict inequality holds since \( U_2 < u_2 \) implies \( \psi(u_1, U_2) \leq \psi(u_1, u_2) \leq u_2 \) (but not vice versa since \( \psi(u_1, U_2) \leq u_2 \) and \( \psi(u_1, u_2) \leq u_2 \) does not necessarily imply \( U_2 < u_2 \) and \( \Pr[\psi(u_1, u_2) < \psi(u_1, U_2)] = \Pr[u_2 < U_2] \neq 0 \)), and the second last inequality holds since, given \( U_1 < u_1 \), \( \psi(U_1, U_2) \leq \psi(u_1, U_2) < u_2 \).
• Case 2: Suppose that \( u_1 \) and \( u_2 \) are such that \( \psi(u_1, u_2) > u_2 \). Then

\[
u_2 - C(u_1, u_2) = \Pr[U_1 > u_1, U_2 < u_2] > \Pr[U_1 > u_1, \psi(u_1, U_2) \leq u_2] \\
\geq \Pr[U_1 > u_1, \psi(U_1, U_2) \leq u_2] = \Pr[\tilde{U}_1 > u_1, \tilde{U}_2 < u_2] = u_2 - \tilde{C}(u_1, u_2)
\]

where the strict inequality holds since \( U_2 > u_2 \) implies \( \psi(u_1, U_2) \geq \psi(u_1, u_2) > u_2 \) or \( \psi(u_1, U_2) \leq u_2 \) implies \( U_2 \leq u_2 \) (but not vice versa).

Therefore in both cases, \( C(u_1, u_2) < \tilde{C}(u_1, u_2) \) for any \( u_1 \) and \( u_2 \).

B Proofs of Results in Section 4

B.1 Identification under Transformation of Marginal Distribution Functions

Recall that we consider the following specification of the marginal distribution functions to derive the asymptotic theory for the sieve ML estimator:

\[
F_{\epsilon_0}(x) = H_{\epsilon_0}(G(x)), \quad F_{\nu_0}(x) = H_{\nu_0}(G_{\nu}(x)),
\]

(B.1)

where \( G: \mathbb{R} \rightarrow [0, 1] \) is a strictly increasing function with its derivative \( g(x) \equiv \frac{dG(x)}{dx} \) and \( g(x) \) is bounded away from zero on \( \mathbb{R} \).

We first verify that there exist \( H_{\epsilon_0} \) and \( H_{\nu_0} \) that satisfy (B.1) for given \( F_{\epsilon_0}, F_{\nu_0} \), and \( G \). Since \( G \) is assumed to be strictly increasing, there exists an inverse function \( G^{-1} \). Letting \( H_{\epsilon_0}(\cdot) = F_{\epsilon_0}(G^{-1}(\cdot)) \) and \( H_{\nu_0}(\cdot) = F_{\nu_0}(G^{-1}(\cdot)) \), it is straightforward to show that \( H_{\epsilon_0} \) and \( H_{\nu_0} \) are mappings from \([0, 1]\) to \([0, 1]\) and that satisfy the relations in (B.1). Note too that this transformation does not change the identification results. That is, \( F_0 \) is identified on \( \mathbb{R} \) if and only if \( H_0 \) is identified on \([0, 1]\). Assuming that \( g \) is bounded away from zero on \( \mathbb{R} \) and bounded above, the unknown density function \( h_{j0} \) can be written as \( h_{j0}(x) = \frac{f_{\epsilon_0}(G^{-1}(x))}{g(G^{-1}(x))} \) for each \( j \in \{\epsilon, \nu\} \), which is well-defined on \([0, 1]\). In addition, we can see that \( h_{\epsilon_0} \) and \( h_{\nu_0} \) are identified if and only if the unknown marginal density functions \( f_{\epsilon_0} \) and \( f_{\nu_0} \) are identified.

We note that the choice of \( G \) depends on the tail behavior of \( f_{\epsilon_0} \) and \( f_{\nu_0} \). If researchers believe that the unknown marginal density functions have fat tails, then they should choose a distribution function with fat tails for \( G \). On the other hand, one can choose the logistic or the standard normal distribution function for \( G \) when \( f_{\epsilon_0} \) and \( f_{\nu_0} \) are likely to have thin tails. This is because Assumption 11 implicitly requires that the unknown marginal density functions and \( g \) decay at the same rate at the tails. Specifically, we observe that

\[
h_{\epsilon_0}(0) = \lim_{x \rightarrow 0^+} h_{\epsilon_0}(x) = \lim_{t \rightarrow -\infty} \frac{f_{\epsilon_0}(t)}{g(t)},
\]

Electronic copy available at: https://ssrn.com/abstract=3346834
and the limit exists if the decaying rates are of the same order. We also provide simulation results to examine how the performance of our semiparametric estimator varies across the choice of $G$ when the marginal density functions have fat tails (see Section (C.3)).

### B.2 Technical Expressions

#### B.2.1 Hölder Norm and Hölder Class

Let $C^m(X)$ be the space of $m$-times continuously differentiable real-valued functions on $X$. Let $\zeta \in (0, 1]$ and, given a $d$-tuple $\omega$, let $[\omega] = \omega_1 + \ldots + \omega_d$. Denote the differential operator by $\mathcal{D}$ and let $\mathcal{D}^{\omega} = \partial^{[\omega]}_{x_1^1 \ldots x_d^\omega}$. Letting $p = m + \zeta$, the Hölder norm of $h \in C^m(X)$ is defined as follows:

$$||h||_{\Lambda^p} \equiv \sup_{[\omega] \leq m, x} |D^\omega h(x)| + \sup_{[\omega] = m, x, y \in X, ||x - y||_E \neq 0} \frac{|D^\omega h(x) - D^\omega h(y)|}{||x - y||_E},$$

where $\zeta$ is the Hölder exponent.

A Hölder class with smoothness $p > 0$, denoted by $\Lambda^p(X)$, is defined as $\Lambda^p(X) \equiv \{h \in C^m(X) : ||h||_{\Lambda^p} < \infty\}$. A Hölder ball with radius $R$, $\Lambda^p_R(X)$, is defined as $\Lambda^p_R(X) \equiv \{h \in \Lambda^p(X) : ||h||_{\Lambda^p} \leq R < \infty\}$.

#### B.2.2 Sup-norm and Pseudo-metric $d_c$

For any $h \in H_\epsilon$ (or $H_\nu$), define the sup-norm on $H_\epsilon$ (or $H_\nu$) as follows:

$$||h||_\infty \equiv \sup_{t \in [0, 1]} |h(t)|.$$

Let $\theta = (\psi', h_\epsilon, h_\nu)' \in \Theta$ be given. We define the consistency norm $|| \cdot ||_c$ as follows:

$$||\theta||_c \equiv ||\psi||_E + ||h_\epsilon||_\infty + ||h_\nu||_\infty,$$

where $|| \cdot ||_E$ is the Euclidean norm. The pseudo-metric $d_c(\cdot, \cdot) : \Theta \times \Theta \rightarrow [0, \infty)$, which induced by the consistency norm $|| \cdot ||_c$, is defined as

$$d_c(\theta_1, \theta_2) = ||\theta_1 - \theta_2||_c.$$

#### B.2.3 $L^2$-norm

$$||\theta - \theta_0||_2 \equiv ||\psi - \psi_0||_E + ||h_\epsilon - h_{\epsilon 0}||_2 + ||h_\nu - h_{\nu 0}||_2,$$

where $||h - \tilde{h}||_2^2 \equiv \int_0^1 (h(t) - \tilde{h}(t))^2 dt$ for any $h, \tilde{h} \in H$. It is straightforward to show that $||\theta - \theta_0||_2 \leq d_c(\theta, \theta_0)$, where $d_c(\theta, \theta_0) = ||\psi - \psi_0||_E + ||h_\epsilon - h_{\epsilon 0}||_\infty + ||h_\nu - h_{\nu 0}||_\infty$. 

Electronic copy available at: https://ssrn.com/abstract=3346834
B.2.4 Fisher inner product and Fisher norm

Recall that \( V \) is the linear span of \( \Theta - \{ \theta_0 \} \). Define the Fisher inner product on the space \( V \) as

\[
<v, \tilde{v} > \equiv E \left[ \left( \frac{\partial l(\theta_0, W)}{\partial \theta} [v] \right) \left( \frac{\partial l(\theta_0, W)}{\partial \theta} [\tilde{v}] \right) \right]
\]

for given \( v, \tilde{v} \in V \). Then, the Fisher norm for \( v \in V \) is defined as

\[
||v||^2 \equiv <v, v >.
\]

B.2.5 Relationship between the Fisher norm and \( L^2 \)-norm

Note that for any \( \theta_1, \theta_2 \in \Theta \), we have

\[
||\theta_1 - \theta_2||^2 = E \left[ \left( \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\theta_1 - \theta_2] \right)^2 \right]
\]

\[
\leq B \left\{ E \left[ \left( \frac{\partial l(\theta_0, W_i)}{\partial \psi} (\psi_1 - \psi_2) \right)^2 \right] + E \left[ \left( \frac{\partial l(\theta_0, W_i)}{\partial h} [h_{\epsilon 1} - h_{\epsilon 2}] \right)^2 \right] + E \left[ \left( \frac{\partial l(\theta_0, W_i)}{\partial h} [h_{\nu 1} - h_{\nu 2}] \right)^2 \right] \right\}
\]

\[
\leq \tilde{B} ||\theta_1 - \theta_2||^2\tag{B.3}
\]

for some \( B, \tilde{B} > 0 \) under Assumptions 10, 11, and 13. From equation (B.3), it is straightforward to see that the convergence rate of the sieve ML estimator with respect to the Fisher norm \( || \cdot || \) is at least as fast as the convergence rate with respect to the \( L^2 \)-norm.

B.2.6 Directional derivatives of the log-likelihood function

Let \( r_{10} = F_{e0}(x' \beta_0 + \delta_{10}), r_{00} = F_{e0}(x' \beta_0), \) and \( s_0 = F_{a0}(x' \alpha_0 + z' \gamma_0) \). For given \( v = (v_\psi, v_\epsilon, v_\nu)' \in V \), we have

\[
\frac{\partial l(\theta_0, w)}{\partial \psi} v_\psi = \sum_{\tilde{g}, \tilde{d} \in \{0,1\}} \left( 1_{\tilde{g}, \tilde{d}} \cdot \frac{1}{p_{\tilde{g}, \tilde{d}, xz}(\theta_0)} \frac{\partial p_{\tilde{g}, \tilde{d}, xz}(\theta_0)}{\partial \psi} \right) v_\psi, \tag{B.4}
\]
Let \( v \). \( B.2.7 \) Directional derivative of the ATE

Following conditions hold:

Proposition B.1. Let \( \frac{\partial l(\theta_0, w)}{\partial h_{\nu}}[v] = 1_{11}(y, d) \times \left[ \frac{1}{p_{11, xz}(\theta_0)} C_1(r_{10}, s_0; \rho_0) \int_{\theta}^{G(x' \beta_0 + \delta_{10})} v_{\epsilon}(t) dt \right] 
\]

\[ + 1_{10}(y, d) \times \left[ \frac{1}{p_{10, xz}(\theta_0)} \left( 1 - C_1(r_{00}, s_0; \rho_0) \right) \int_{\theta}^{G(x' \beta_0)} v_{\epsilon}(t) dt \right] 
\]

\[ + 1_{01}(y, d) \times \left[ \frac{1}{p_{01, xz}(\theta_0)} \left( -C_1(r_{10}, s_0; \rho_0) \right) \int_{\theta}^{G(x' \beta_0 + \delta_{10})} v_{\epsilon}(t) dt \right] 
\]

\[ + 1_{00}(y, d) \times \left[ \frac{1}{p_{00, xz}(\theta_0)} \left( 1 - C_1(r_{00}, s_0; \rho_0) \right) \int_{\theta}^{G(x' \beta_0)} v_{\epsilon}(t) dt \right], \tag{B.5} \]

and

\[ \frac{\partial l(\theta_0, w)}{\partial h_{\nu}}[v_{\epsilon}] = \left\{ 1_{11}(y, d) \times \frac{1}{p_{11, xz}(\theta_0)} C_2(r_{10}, s_0; \rho_0) + 1_{10}(y, d) \times \frac{1}{p_{10, xz}(\theta_0)} (-C_2(r_{00}, s_0; \rho_0)) \right. 
\]

\[ + 1_{01}(y, d) \times \frac{1}{p_{01, xz}(\theta_0)} (1 - C_2(r_{10}, s_0; \rho_0)) + 1_{00}(y, d) \times \frac{1}{p_{00, xz}(\theta_0)} (1 - C_2(r_{00}, s_0; \rho_0)) \left\} \right. 
\]

\[ \times \int_{\theta}^{G(x' \alpha_0 + \gamma_0)} v_{\epsilon}(t) dt \right. \tag{B.6} \]

\[ \text{B.2.7 Directional derivative of the ATE} \]

Let \( v = (v', v_{\epsilon}, v_{\nu})' \in \mathbb{V} \). Then,

\[ \frac{\partial \text{ATE}(\theta_0; x)}{\partial \theta'}[v] = \left\{ f_{\epsilon 0}(x' \beta_0 + \delta_{10})(x' \nu_{\beta} + \nu_{\beta}) - f_{\epsilon 0}(x' \beta_0) x' \nu_{\beta} \right\} + \int_{G(x' \beta_0)}^{G(x' \beta_0 + \delta_{10})} v_{\epsilon}(t) dt, \tag{B.7} \]

where \( f_{\epsilon 0}(x) = h_{\epsilon 0}(G(x))g(x) \).

\[ \text{B.3 Proof of Theorem 4.1} \]

Define \( Q_0(\theta) \equiv E[l(\theta, W_i)] \). The following proposition is a modification of Theorem 3.1 in Chen (2007) and establishes the consistency of sieve M-estimator.\(^3\)

**Proposition B.1.** Let \( \hat{\theta}_n \) be the sieve extremum estimator defined in (4.2). Suppose that the following conditions hold:

(i) \( Q_0(\theta) \) is uniquely maximized at \( \theta_0 \) in \( \Theta \) and \( Q_0(\theta_0) > -\infty \);

(ii) \( \Theta \) is compact under \( d_c(\cdot, \cdot) \), and \( Q_0(\theta) \) is upper semicontinuous on \( \Theta \) under \( d_c(\cdot, \cdot) \);

(iii) The sieve spaces, \( \Theta_n \), are compact under \( d_c(\cdot, \cdot) \);

(iv) \( \Theta_k \subseteq \Theta_{k+1} \subseteq \Theta \) for all \( k \geq 1 \), and there exists a sequence \( \pi_k \theta_0 \in \Theta_k \) such that \( d_c(\theta_0, \pi_k \theta_0) \to 0 \) as \( k \to \infty \);\(^3\)

\(^3\)See also Remark 3.3 in Chen (2007).
(v) For all \( k \geq 1 \), \( p \lim_{n \to \infty} \sup_{\theta \in \Theta_k} |Q_n(\theta) - Q_0(\theta)| = 0 \).

Then, \( d_c(\hat{\theta}_n, \theta_0) = o_p(1) \).

We show that the conditions in Theorem 4.1 imply those in this proposition to prove consistency of the sieve estimator. We first need to verify that (i) the true parameter \( \theta_0 \) is the unique maximizer of \( Q_0(\cdot) \) over \( \Theta \) and that (ii) the sample log-likelihood function \( Q_n(\cdot) \) uniformly converges to \( Q_0(\cdot) \) over the sieve space in probability to establish the consistency of the sieve ML estimator. The following lemma shows that if the model with unknown marginal distributions are identified and some additional conditions are satisfied, then the true parameter \( \theta_0 \) is the unique maximizer of \( Q_0(\cdot) \) over \( \Theta \).

**Lemma B.1.** Suppose that Assumptions 1–5, 7, 8 and 9 are satisfied. Then the condition (i) in Proposition B.1 is satisfied.

**Proof.** By Theorem 2.3, the model parameter is identified. Under Assumption 9, we can see that for any \( \theta \in \Theta, |Q_0(\theta)| \leq E|l(\theta, W_i)| \leq \sum_{y,d \in \{0,1\}} E|\log(p_{yd,xz}(\theta))| < \infty \), and thus the function \( Q_0(\theta) \) is well-defined on \( \Theta \) and \( Q_0(\theta) > -\infty \) for all \( \theta \in \Theta \); hence \( Q_0(\theta_0) > -\infty \). Since the model is identified, it implies that for \( \theta \neq \theta_0 \), there exists a set \( E \subset \text{supp}(X, Z) \) such that \( \int_E dP_{XZ} > 0 \) and for some \( y, d \in \{0,1\} \), \( \frac{p_{yd,xz}(\theta)}{p_{yd,xz}(\theta_0)} \neq 1 \) on \( E \), where \( P_{XZ} \) is the distribution function of \( (X, Z) \).

Thus, we have

\[
Q_0(\theta) - Q_0(\theta_0) = \int \sum_{y,d \in \{0,1\}} p_{yd,xz}(\theta_0) \log \left( \frac{p_{yd,xz}(\theta)}{p_{yd,xz}(\theta_0)} \right) dP_{XZ} \leq 0,
\]

where the strict inequality holds by the fact that \( p_{yd,xz}(\theta) \neq p_{yd,xz}(\theta_0) \) on \( E \) and Jensen’s inequality. Hence, \( \theta_0 \) is the unique maximizer of \( Q_0(\cdot) \). \( \blacksquare \)

For any \( \omega > 0 \), let \( N(\omega; \Theta_n, d_c) \) be the covering numbers without bracketing of \( \Theta_n \) with respect to the pseudo-metric \( d_c \). We now establish the uniform convergence of \( Q_n(\cdot) \) to \( Q_0 \) over the sieve space.

**Lemma B.2.** Suppose that Assumptions 1–5, 7 are satisfied. If Assumptions 8 through 13 hold, then \( \sup_{\theta \in \Theta_n} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0 \) for all \( n \geq 1 \).

**Proof.** We verify Condition 3.5M in Chen (2007). Let \( B \) stand for a generic constant and it can be different in each place. By Assumptions 9 and 10, the first condition in Condition 3.5M is satisfied. Let \( n \geq 1 \) be a natural number and \( \theta, \tilde{\theta} \in \Theta_n \). Define \( R_1(\theta) = F_\epsilon(X' \beta + \delta_1), R_0(\theta) = F_\epsilon(X' \beta), \) and \( S(\theta) = F_\nu(X' \alpha + Z' \gamma) \). Similarly, we define \( R_1(\tilde{\theta}) = \tilde{F}_\epsilon(X' \tilde{\beta} + \tilde{\delta}_1), R_0(\tilde{\theta}) = \tilde{F}_\epsilon(X' \tilde{\beta}), \) and \( S(\tilde{\theta}) = \tilde{F}_\nu(X' \tilde{\alpha} + Z' \gamma) \). For the simplicity of the notations, we write \( R_j = R_j(\theta), \tilde{R}_j = R_j(\tilde{\theta}), \)
\[ S = S(\theta), \text{ and } \tilde{S} = S(\tilde{\theta}) \text{ for all } j = 0, 1. \] Observe that

\[
|p_{11,XZ}(\theta) - p_{11,XZ}(\tilde{\theta})| = |C(R_1, S; \rho) - C(R_1, \tilde{S}; \tilde{\rho})|
\leq |C(R_1, S; \rho) - C(R_1, \tilde{S}; \rho)| + |C(R_1, \tilde{S}; \rho) - C(R_1, \tilde{S}; \tilde{\rho})|
\leq |R_1 - \tilde{R}_1| + |S - \tilde{S}| + |C_\rho(R_1, \tilde{S}; \rho)||\rho - \tilde{\rho}|
\leq |R_1 - \tilde{R}_1| + |S - \tilde{S}| + B|\rho - \tilde{\rho}|,
\]

where \( C_\rho(\cdot, \cdot; \cdot) \) is the partial derivative of \( C(\cdot, \cdot; \cdot) \) with respect to \( \rho \) and \( \tilde{\rho} \) is between \( \rho \) and \( \tilde{\rho} \) and \( B < \infty \). Note that the last inequality holds due to a generic property of copulas (see, e.g., Theorem 2.2.4 in Nelsen (1999)) and the mean value theorem. We also have

\[
|R_1 - \tilde{R}_1| = \left| F_\epsilon(X' + \delta_1) - \tilde{F}_\epsilon(X' + \tilde{\delta}_1) \right|
\leq \left| F_\epsilon(X' + \delta_1) - F_\epsilon(X' + \tilde{\delta}_1) \right| + \left| F_\epsilon(X' + \tilde{\delta}_1) - \tilde{F}_\epsilon(X' + \tilde{\delta}_1) \right|
\leq \left| f_\epsilon(X' + \delta_1) \right| \cdot \left| X'(\beta - \tilde{\beta}) + (\delta_1 - \tilde{\delta}_1) \right| + \int_0^{G(X' + \tilde{\delta}_1)} \left| h_\epsilon(t) - \tilde{h}_\epsilon(t) \right| dt
\leq \sup_{x \in \mathbb{R}} \left| h_\epsilon(G(x))g(x) \right| \times \| (X', 1)' ||_E \cdot \| \psi - \tilde{\psi} ||_E + \| h_\epsilon - \tilde{h}_\epsilon ||_\infty
\leq B \times \| (X', 1)' ||_E \times \| (\beta', \delta_1)' - (\tilde{\beta}', \tilde{\delta}_1)' ||_E + \| h_\epsilon - \tilde{h}_\epsilon ||_\infty, \quad (B.8)
\]

for some constant \( B < \infty \). Similarly, we can show that

\[
|R_0 - \tilde{R}_0| \leq B \times \| X ||_E \times \| \beta - \tilde{\beta} ||_E + \| h_\epsilon - \tilde{h}_\epsilon ||_\infty \tag{B.9}
\]

and

\[
|S - \tilde{S}| \leq B \times \| (X', Z)' ||_E \times \| (\alpha', \gamma)' - (\tilde{\alpha}', \tilde{\gamma}') ||_E + \| h_\nu - \tilde{h}_\nu ||_\infty. \tag{B.10}
\]

Note that, for any comparable subvectors \( \psi_s \) and \( \tilde{\psi}_s \) of \( \psi \) and \( \tilde{\psi} \), respectively, we have \( \| \psi_s - \tilde{\psi}_s ||_E \leq \| \psi - \tilde{\psi} ||_E \) and that, for any subvector \( W_s \) of \( W \), we have \( \| W_s ||_E \leq \| W ||_E \) a.s. Thus we have

\[
|p_{11,XZ}(\theta) - p_{11,XZ}(\tilde{\theta})| \leq B \| (X', 1)' ||_E \cdot \| \psi - \tilde{\psi} ||_E + \| h_\epsilon - \tilde{h}_\epsilon ||_\infty
\leq B \| (X', 1)' ||_E d_\epsilon(\theta, \tilde{\theta}).
\]
Consequently, it follows that

\[
|p_{10,XZ}(\theta) - p_{10,XZ}(\tilde{\theta})| \leq |R_0 - \tilde{R}_0| + |C(R_0, S; \rho) - C(\tilde{R}_0, \tilde{S}; \tilde{\rho})|
\]

\[
\leq 2|R_0 - \tilde{R}_0| + |S - \tilde{S}| + B|\rho - \tilde{\rho}|
\]

\[
\leq B\{\|X\|_E||\beta - \tilde{\beta}||_E + \|X^\prime, Z^\prime\|_E||\alpha^\prime, \gamma^\prime\|_E - (\tilde{\alpha}^\prime, \tilde{\gamma}^\prime)|_E
\]

\[
+ \|h_e - \tilde{h}_e\|_\infty + \|h_\nu - \tilde{h}_\nu\|_\infty + |\rho - \tilde{\rho}|\}
\]

\[
\leq B \cdot ||(X^\prime, Z^\prime, 1)'||_E d_e(\theta, \tilde{\theta}),
\]

\[
|p_{01,XZ}(\theta) - p_{01,XZ}(\tilde{\theta})| \leq |S - \tilde{S}| + |R_1 - \tilde{R}_1| + B|\rho - \tilde{\rho}|
\]

\[
\leq B \cdot ||(X^\prime, Z^\prime, 1)'||_E d_e(\theta, \tilde{\theta}),
\]

\[
|p_{00,XZ}(\theta) - p_{00,XZ}(\tilde{\theta})| \leq |p_{11,XZ}(\theta) - p_{11,XZ}(\tilde{\theta})| + |p_{10,XZ}(\theta) - p_{10,XZ}(\tilde{\theta})| + |p_{01,XZ}(\theta) - p_{01,XZ}(\tilde{\theta})|
\]

\[
\leq B \cdot ||(X^\prime, Z^\prime, 1)'||_E d_e(\theta, \tilde{\theta}).
\]

In all, we have

\[
|l(\theta, W_i) - l(\tilde{\theta}, W_i)| \leq \sum_{y.d=0,1} \mathbf{1}_{y.d}(Y_i, D_i) \cdot |\log p_{yd}(X_i, Z_i; \theta) - \log p_{yd}(X_i, Z_i; \tilde{\theta})|
\]

\[
\leq \frac{1}{p(X_i, Z_i)} \sum_{y.d=0,1} \mathbf{1}_{y.d}(Y_i, D_i) \left| p_{yd}(X_i, Z_i; \theta) - p_{yd}(X_i, Z_i; \tilde{\theta}) \right|
\]

\[
\leq \frac{B}{p(X_i, Z_i)} \cdot ||(X^\prime, Z^\prime, 1)'||_E d_e(\theta, \tilde{\theta})
\]

\[
\equiv U(W_i) d_e(\theta, \tilde{\theta}), \quad \text{(B.11)}
\]

where $E[U(W_i)^2] < \infty$ by Assumptions 9 and 10. This results in

\[
\sup_{\theta, \tilde{\theta} \in \Theta_n, d_e(\theta, \tilde{\theta}) \leq \epsilon_0} \left| l(\theta, W_i) - l(\tilde{\theta}, W_i) \right| \leq U(W_i) \epsilon_0 \quad \text{(B.12)}
\]

and thus the second condition in Condition 3.5M is satisfied with $s = 1$.

For the last condition in Condition 3.5M, note that for any $\omega > 0$, we have

\[
N(\omega, \Theta_n, d_e) \leq N\left(\frac{\omega}{2}, \Psi, \cdot ||E\right) \cdot N\left(\frac{\omega}{4}, \mathcal{H}_{cn}, \cdot ||\cdot ||_\infty\right) \cdot N\left(\frac{\omega}{4}, \mathcal{H}_{cn}, \cdot ||\cdot ||_\infty\right).
\]

By Lemma 2.5 in van de Geer (2000), we have $\log N\left(\frac{\omega}{4}, \mathcal{H}_{cn}, \cdot ||\cdot ||_\infty\right) \leq k_n \log \left(1 + \frac{32R}{\omega}\right)$ under Assumption 12-(i); and hence

\[
\log N(\omega, \Theta_n, d_e) \leq \text{const.} \times k_n \times \log \left(1 + \frac{32R}{\omega}\right) = o(n)
\]

if $k_n/n \to 0$. Since the condition $k_n/n = o(1)$ is imposed by Assumption 12-(i), the last condition
in Condition 3.5M is also satisfied. In all, we have the uniform convergence of \( Q_n \) to \( Q_0 \) over \( \Theta_k \).

To finish proving Theorem 4.1, we verify the conditions in Proposition B.1. By Lemmas B.1 and B.2, the conditions (i) and (v) in Proposition B.1 are satisfied. Using (B.11) and Jensen’s inequality, we can see that, for any \( \theta, \tilde{\theta} \in \Theta \),

\[
|Q_0(\theta) - Q_0(\tilde{\theta})| \leq E[l(\theta, W_i) - l(\tilde{\theta}, W_i)] \leq E[U(W_i)]d_c(\theta, \tilde{\theta}) = B \cdot d_c(\theta, \tilde{\theta})
\]

for some \( B < \infty \). Thus, \( Q_0(\cdot) \) is continuous with respect to \( d_c \). Note that since the parameter space of the finite-dimensional parameter \( \psi, \Psi \), is assumed to be compact in Assumption 8, the original parameter space \( \Theta \) is compact under the \( d_c \), by Theorems 1 and 2 in Freyberger and Masten (2015), and thus the conditions (ii) and (iii) are satisfied with the specified parameter space and the norm. Since the condition (iv) is directly imposed, we have \( d(\hat{\theta}_n, \theta_0) = o_p(1) \) by Proposition B.1.

### B.4 Proof of Theorem 4.2

To establish the convergence rate with respect to the norm \( || \cdot ||_2 \), we consider the following assumption:

**Assumption B.1.** Let \( K(\theta_0, \theta) \equiv E[l(\theta_0, W_i) - l(\theta, W_i)] \). Then, there exist \( B_1, B_2 > 0 \) such that

\[
B_1K(\theta_0, \theta) \leq ||\theta - \theta_0||_2^2 \leq B_2K(\theta_0, \theta)
\]

for all \( \theta \in \Theta_n \) with \( d_c(\theta, \theta_0) = o(1) \).

Assumption B.1 implies that the \( L^2 \)-norm \( || \cdot ||_2 \) and the square-root of the KL divergence are equivalent.

We derive the convergence rate of the sieve M-estimator with respect to the norm \( || \cdot ||_2 \) by checking the conditions in Theorem 3.2 in Chen (2007). Since \( \{W_i\}_{i=1}^n \) is assumed to be i.i.d by Assumption 10, Condition 3.6 in Chen (2007) is satisfied. For Condition 3.7 in Chen (2007), we note that for a small \( \epsilon_1 > 0 \) and for any \( \theta \in \Theta_n \) such that \( ||\theta - \theta_0||_2 \leq \epsilon_1 \), we have

\[
Var(l(\theta, W_i) - l(\theta_0, W_i)) \leq E[l(\theta, W_i) - l(\theta_0, W_i)]^2
\]

\[
\leq E \left[ \frac{1}{p(X_i, Z_i)^2} \sum_{y,d=0,1} 1_{yd}(Y_i, D_i)|p_{yd}(X_i, Z_i; \theta) - p_{yd}(X_i, Z_i; \theta_0)|^2 \right]
\]

\[
\leq E \left[ \frac{1}{p(X_i, Z_i)^2} \sum_{y,d \in \{0,1\}} |p_{yd}(X_i, Z_i; \theta) - p_{yd}(X_i, Z_i; \theta_0)|^2 \right].
\]
By the same logic in (B.11), we have

\[ \text{Var} (l(\theta, W_i) - l(\theta_0, W_i)) \leq E \left[ U(W_i)^2 \right] d_c(\theta, \theta_0)^2. \]

Note that

\[
d_c(\theta, \theta_0)^2 = (||\psi - \psi_0||_E + ||h_\epsilon - h_{\epsilon 0}||_\infty + ||h_\nu - h_{\nu 0}||_\infty)^2
\leq 4(||\psi - \psi_0||_E^2 + ||h_\epsilon - h_{\epsilon 0}||_\infty^2 + ||h_\nu - h_{\nu 0}||_\infty^2).
\]

By Lemma 2 in Chen and Shen (1998), we have

\[
||h_j - h_{j 0}||_\infty^2 \leq ||h_j - h_{j 0}||^2_2 \tag{B.13}
\]

for all \( j \in \{\epsilon, \nu\}. \) Since \( \frac{4p}{2p+1} > 1 \) under Assumption 11, we can show that

\[
\sup_{\{\theta \in \Theta_n: ||\theta - \theta_0||_2 \leq \epsilon_1\}} \text{Var} (l(\theta, W_i) - l(\theta_0, W_i)) \leq B_1 \epsilon_1^2
\]

with \( \epsilon_1 \leq 1 \) and some constant \( B_1, \) and thus Condition 3.7 in Chen (2007) is satisfied.

We recall equation (B.11) to verify Condition 3.8 in Chen (2007). Let \( \epsilon_2 > 0 \) be given and consider

\[
|l(\theta, W_i) - l(\theta_0, W_i)| \leq U(W_i) d_c(\theta, \theta_0)
= U(W_i) \left\{ ||\psi - \psi_0||_E + ||h_\epsilon - h_{\epsilon 0}||_\infty + ||h_\nu - h_{\nu 0}||_\infty \right\}
\leq U(W_i) \left\{ ||\psi - \psi_0||_E + ||h_\epsilon - h_{\epsilon 0}||_2^{\frac{2p}{2p+1}} + ||h_\nu - h_{\nu 0}||_2^{\frac{2p}{2p+1}} \right\}
\leq U(W_i) \left\{ ||\psi - \psi_0||_E^{\frac{2p+1}{2p}} + ||h_\epsilon - h_{\epsilon 0}||_2 + ||h_\nu - h_{\nu 0}||_2 \right\}^{\frac{2p}{2p+1}}
\leq \bar{U}(W_i) \left\{ ||\psi - \psi_0||_E + ||h_\epsilon - h_{\epsilon 0}||_2 + ||h_\nu - h_{\nu 0}||_2 \right\}^{\frac{2p}{2p+1}}, \tag{B.14}
\]

where \( \bar{U}(W_i) = \max\{1, (\sup_{\psi \in \Psi} ||\psi||_E + ||\psi_0||_E)^{\frac{1}{2p}} \times U(W_i). \) Since the parameter space for \( \psi, \Psi, \) is compact under Assumption 8, \( E[\bar{U}(W_i)^2] < \infty. \) Thus, we have

\[
\sup_{\{\theta \in \Theta_n: ||\theta - \theta_0||_2 \leq \epsilon_2\}} |l(\theta, W_i) - l(\theta_0, W_i)| \leq \epsilon_2^{\frac{2p}{2p+1}} \bar{U}(W_i)
\]

with \( E[\bar{U}(W_i)^2] < \infty \) and this implies that, under Assumption 11, Condition 3.8 in Chen (2007)
is satisfied with \( s = \frac{2p}{2p+1} \in (0, 2) \) and \( \gamma = 2 \).

Let \( \mathcal{L}_n = \{ l(\theta_0, W_i) - l(\theta, W_i) : \theta \in \Theta_n, ||\theta - \theta_0||_2 \leq \epsilon_2 \} \). For given \( \omega > 0 \), let \( N[\omega, \mathcal{L}_n, || \cdot ||_{L^2}] \) be the covering number with bracketing of \( \mathcal{L}_n \) with respect to the norm \( || \cdot ||_{L^2} \). We now need to calculate \( \kappa_n \), which is defined as

\[
\kappa_n = \inf \left\{ \kappa \in (0, 1) : \frac{1}{\sqrt{n\kappa^2}} \int_{b\kappa^2}^{\kappa} \sqrt{H[\omega, \mathcal{L}_n, || \cdot ||_{L^2}]} \, d\omega \leq \text{const.} \right\},
\]

where, for \( f \in \mathcal{L}_n, ||f(\theta, W_i)||_{L^2}^2 \equiv E[f(\theta, W_i)^2] \) is the \( L^2 \)-norm on \( \mathcal{L}_n \) and \( H[\omega, \mathcal{L}_n, || \cdot ||_{L^2}] \) is the \( L^2 \)-metric entropy with bracketing of the class \( \mathcal{L}_n \) (see van der Vaart and Wellner (1996) or van de Geer (2000) for the definition of \( L_2 \)-metric entropy with bracketing). Let \( B_0 = E[U(W_i)^2] \), where \( U(W_i) \) is the same to the one in (B.11). By Theorem 2.7.11 in van der Vaart and Wellner (1996) and equation (B.11), we can show that

\[
N[\omega, \mathcal{L}_n, || \cdot ||_{L^2}] \leq N \left( \frac{\omega}{2B_0}, \Theta_n, d_c \right) \\
\leq N \left( \frac{\omega}{4B_0}, \Psi, || \cdot ||_E \right) \cdot N \left( \frac{\omega}{8B_0}, \mathcal{H}_{\kappa n}, || \cdot ||_\infty \right) \cdot N \left( \frac{\omega}{8B_0}, \mathcal{H}_{\kappa n}, || \cdot ||_\infty \right),
\]

and this leads to

\[
H[\omega, \mathcal{L}_n, || \cdot ||_{L^2}] = \log \left( N[\omega, \mathcal{L}_n, || \cdot ||_{L^2}] \right) \leq \text{const.} \times \kappa_n \times \log \left( 1 + \frac{64B_0R}{\omega} \right).
\]

In all, \( \kappa_n \) solves

\[
\frac{1}{\sqrt{n\kappa_n^2}} \int_{b\kappa_n^2}^{\kappa_n} \sqrt{H[\omega, \mathcal{L}_n, || \cdot ||_{L^2}]} \, d\omega \leq \text{const.} \cdot \frac{1}{\sqrt{n\kappa_n^2}} \int_{b\kappa_n^2}^{\kappa_n} k_n \cdot \log \left( 1 + \frac{64B_0R}{\omega} \right) \, d\omega \\
\leq \text{const.} \cdot \frac{1}{\sqrt{n\kappa_n^2}} \int_{b\kappa_n^2}^{\kappa_n} \frac{1}{\omega} \, d\omega \leq \text{const.} \times \frac{1}{\sqrt{n\kappa_n}} \sqrt{k_n \kappa_n} \leq \text{const.,}
\]

and thus \( \kappa_n \propto \sqrt{\frac{k_n}{n}} \).

Lastly, since \( ||\theta_0 - \pi_n \theta_0||_2 \leq ||\theta_0 - \pi_n \theta_0||_c = O(k_n^{-p}) \) by Lorentz (1966), we have

\[
||\hat{\theta}_n - \theta_0||_2 = O_p \left( \max \left\{ \sqrt{\frac{k_n}{n}}, k_n^{-p} \right\} \right)
\]

by Theorem 3.2 in Chen (2007). By choosing \( k_n \propto n^{-\frac{1}{2p+1}} \), we have

\[
||\hat{\theta}_n - \theta_0||_2 = O_p \left( n^{-\frac{p}{2p+1}} \right).
\]
B.5 Proof of Proposition 4.1

We first provide some technical assumptions for the asymptotic normality. Let $\mu_n(g) = \frac{1}{n} \sum_{i=1}^{n} \{g(W_i) - E[g(W_i)]\}$ be the empirical process indexed by $g$. Let the convergence rate of the sieve estimator be $\delta_n$ (i.e., $||\hat{\theta}_n - \theta_0|| = O_p(\delta_n)$).

Assumption B.2. There exist $\xi_1 > 0$ and $\xi_2 > 0$ with $2\xi_1 + \xi_2 < 1$ and a constant $K$, such that $(\delta_n)^{3-(2\xi_1+\xi_2)} = o(n^{-1})$. In addition, the following hold for all $\hat{\theta} \in \Theta_n$ with $||\hat{\theta} - \theta_0|| \leq \delta_n$, and all $v \in \mathbb{V}$ with $||v|| \leq \delta_n$:

(i) $\left| E \left[ \frac{\partial^2 l(\theta, W)}{\partial \theta \partial \psi} - \frac{\partial^2 l(\theta_0, W)}{\partial \theta \partial \psi} \right] \right| < K \left| \hat{\theta} - \theta_0 \right|^{1-\xi_2}$;

(ii) $\left| E \left[ \sum_{j \in \{\epsilon, \nu\}} \left( \frac{\partial^2 l(\theta, W)}{\partial \theta \partial h_j} [v_j] - \frac{\partial^2 l(\theta_0, W)}{\partial \theta \partial h_j} [v_j] \right) \right] \right| \leq K \left| v \right|^{1-\xi_1} \left| \hat{\theta} - \theta_0 \right|^{1-\xi_2}$;

(iii) $\left| E \left[ \sum_{i,j \in \{\epsilon, \nu\}} \left( \frac{\partial^2 l(\theta, W)}{\partial \theta \partial \psi} [v, v] - \frac{\partial^2 l(\theta_0, W)}{\partial \theta \partial \psi} [v, v] \right) \right] \right| \leq K \left| v \right|^{2(1-\xi_1)} \left| \hat{\theta} - \theta_0 \right|^{1-\xi_2}$.

Assumption B.3. The following hold:

(i) $\sup_{\theta \in \Theta_n; ||\theta - \theta_0|| = O(\delta_n)} \mu_n \left( \frac{\partial l(\theta, W)}{\partial \theta} - \frac{\partial l(\theta_0, W)}{\partial \theta} \right) = o_p \left( n^{-\frac{1}{2}} \right)$;

(ii) For all $j \in \{\epsilon, \nu\}$, $\sup_{\theta \in \Theta_n; ||\theta - \theta_0|| = O(\delta_n)} \mu_n \left( \frac{\partial l(\theta, W)}{\partial \theta} [\pi_n v_j^*] - \frac{\partial l(\theta_0, W)}{\partial \theta} [\pi_n v_j^*] \right) = o_p \left( n^{-\frac{1}{2}} \right)$.

Assumptions B.2 and B.3 are modifications of Assumptions 5 and 6 in CFT06, which are needed to control for the second-order expansion of the log-likelihood function $l(\theta, W)$. Under Assumption 14, these conditions require that the unknown marginal density functions be sufficiently smooth. For example, the sieve estimator needs to converge at a faster rate than $1/(3 - (2\xi_1 + \xi_2))$ to satisfy $(\delta_n)^{3-(2\xi_1+\xi_2)} = o(n^{-1})$. Usually, the convergence rate depends positively on the smoothness parameter $p$ in Assumption 11 and thus the class of models should be restricted to that in which the density functions are sufficiently smooth.

Note that since the sieve ML estimator $\hat{\theta}_n$ is consistent with respect to the pseudo-metric $d_c$ by Theorem 4.1, it is consistent with respect to the norm $|| \cdot ||_2$ and thus with respect to the Fisher norm by equation (B.3). We also point out that $||\hat{\theta}_n - \theta_0|| = O_p(n^{-\frac{p}{2(p+\tau)}})$ by equation (B.3) and Theorem 4.2 under the given set of Assumptions. We follow the proof of Theorem 1 in CFT06. Assumptions 1 and 2 in CFT06 are implied by Assumption 1-5, 7-9, and 14. The first two parts in Assumption 15 correspond to Assumption 3 in CFT06. Since $p > 1/2$ by Assumption 11, $||\hat{\theta}_n - \theta_0|| = o_p(n^{-1/4})$ by Theorem 4.2 and this implies that $||\hat{\theta}_n - \theta_0|| \times ||\pi_n v^* - v^*|| = o(n^{-1/2})$ under Assumption 16. In addition, since $w > 1 + \frac{1}{2p}$, $\delta_n^w = o(n^{-1/2})$ by that $||\hat{\theta}_n - \theta_0|| = O_p(n^{-\frac{p}{2(p+\tau)}})$. Hence, Assumptions 3 and 4 in CFT06 are satisfied.

Define $r[\theta, \theta_0, W_i] \equiv l(\theta, W_i) - l(\theta_0, Z_i) - \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\theta - \theta_0]$ and $\xi_0 = 2\xi_1 + \xi_2$. Let $\zeta_n$ be a
positive sequence with \( \zeta_n = o(n^{-1/2}) \) and \((\delta_n)^{3-(2\xi_1+\xi_2)} = \zeta_n o(n^{-1/2}) \). Then we have

\[
0 \leq \frac{1}{n} \sum_{i=1}^{n} l(\hat{\theta}_n, W_i) - l(\hat{\theta}_n \pm \zeta_n \pi_n v^*, W_i) \leq \mp \zeta_n \frac{1}{n} \sum_{i=1}^{n} \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\pi_n v^*] \\
+ \mu_n (r[\hat{\theta}_n, \theta_0, W_i] - r[\hat{\theta}_n \pm \zeta_n \pi_n v^*, \theta_0, W_i]) + E[r[\hat{\theta}_n, \theta_0, W_i] - r[\hat{\theta}_n \pm \zeta_n \pi_n v^*, \theta_0, W_i]]. \tag{B.15}
\]

We first note that, by Assumption 16,

\[
E \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\pi_n v^* - v^*] \right]^2 \leq \frac{1}{n} E \left[ \left\{ \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\pi_n v^* - v^*] \right\}^2 \right] \\
= \frac{1}{n} \| \pi_n v^* - v^* \|^2 = o(n^{-1}), \tag{B.16}
\]

and hence \( \frac{1}{n} \sum_{i=1}^{n} \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\pi_n v^* - v^*] = o_p(n^{-1/2}) \).

Observe that, by the mean value theorem,

\[
E [r[\theta, \theta_0, W_i]] = E \left[ l(\theta, W_i) - l(\theta_0, W_i) - \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\theta - \theta_0] \right] \\
= E \left[ \frac{1}{2} \frac{\partial^2 l(\theta_0, W_i)}{\partial \theta^2} [\theta - \theta_0, \theta - \theta_0] \right] \\
+ \frac{1}{2} E \left[ \frac{\partial^2 l(\hat{\theta}_i, W_i)}{\partial \theta^2} [\theta - \theta_0, \theta - \theta_0] - \frac{\partial^2 l(\theta_0, W_i)}{\partial \theta^2} [\theta - \theta_0, \theta - \theta_0] \right], \tag{B.17}
\]

where \( \theta, \hat{\theta} \in \Theta_n \) and \( \hat{\theta} \) is between \( \theta \) and \( \theta_0 \). In addition, for any \( v = (v^*_\psi, v_\epsilon, v_\nu) \in \mathcal{V} \) and \( \hat{\theta} \in \Theta_n \) with \( \| \hat{\theta} - \theta_0 \| = O(\delta_n) \), we have

\[
E \left[ \frac{\partial^2 l(\hat{\theta}, W_i)}{\partial \theta^2} [v, v] - \frac{\partial^2 l(\theta_0, W_i)}{\partial \theta^2} [v, v] \right] = v^*_\psi E \left[ \frac{\partial^2 l(\hat{\theta}, W_i)}{\partial \psi \partial \psi'} - \frac{\partial^2 l(\theta_0, W_i)}{\partial \psi \partial \psi'} \right] v^*_\psi \\
+ \sum_{j \in \{\epsilon, \nu\}} 2v^*_j E \left[ \frac{\partial^2 l(\hat{\theta}, W_i)}{\partial \psi \partial h_j} [v_j] - \frac{\partial^2 l(\theta_0, W_i)}{\partial \psi \partial h_j} [v_j] \right] \\
+ \sum_{k \in \{\epsilon, \nu\}} \sum_{j \in \{\epsilon, \nu\}} E \left[ \frac{\partial^2 l(\hat{\theta}, W_i)}{\partial h_k \partial h_j} [v_k, v_j] - \frac{\partial^2 l(\theta_0, W_i)}{\partial h_k \partial h_j} [v_k, v_j] \right],
\]

and this term can be controlled under Assumption B.2 in the same way of CFT06. This leads us to that

\[
E[r[\hat{\theta}_n, \theta_0, W_i] - r[\hat{\theta}_n \pm \zeta_n \pi_n v^*, \theta_0, W_i]] = -\frac{1}{2} (||\hat{\theta}_n - \theta_0||^2 - ||\hat{\theta}_n \pm \zeta_n \pi_n v^* - \theta_0||^2) + \zeta_n o(n^{-1/2}) \\
= \pm \zeta_n \times <\hat{\theta}_n - \theta_0, v^*> + \zeta_n o(n^{-1/2}) \tag{B.18}
\]
because we have $<\hat{\theta}_n - \theta_0, \pi_n v^* - v^*> = o_p(n^{-1/2})$ and $||\pi_n v^*||^2 \to ||v^*||^2 < \infty$.

We also have that

$$
\mu_n \left( r[\hat{\theta}_n, \theta_0, W_i] - r[\hat{\theta}_n \pm \zeta_n \pi_n v^*, \theta_0, W_i] \right)
= \mu_n \left( l(\hat{\theta}_n, W_i) - l(\hat{\theta}_n \pm \zeta_n \pi_n v^*, W_i) - \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\mp \zeta_n \pi_n v^*] \right)
= \mp \zeta_n \cdot \mu_n \left( \frac{\partial l(\hat{\theta}, W_i)}{\partial \theta} [\pi_n v^*] - \frac{\partial l(\theta_0, W_i)}{\partial \theta} [\pi_n v^*] \right),
$$

where $\tilde{\theta} \in \Theta_n$ is between $\hat{\theta}_n$ and $\hat{\theta}_n \pm \zeta_n \pi_n v^*$. By Assumption B.3, we have

$$
\mu_n \left( r[\hat{\theta}_n, \theta_0, W_i] - r[\hat{\theta}_n \pm \zeta_n \pi_n v^*, \theta_0, W_i] \right) = o_p(\zeta_n n^{-1/2}).
$$

Combining equations (B.15) through (B.19) with the fact that $E \left[ \frac{\partial l(\theta_0, W_i)}{\partial \theta} [v^*] \right] = 0$, we have

$$
0 \leq \frac{1}{n} \sum_{i=1}^{n} l(\hat{\theta}_n, W_i) - l(\hat{\theta}_n \pm \zeta_n \pi_n v^*, W_i)
= \mp \zeta_n \cdot \mu_n \left( \frac{\partial l(\theta_0, W_i)}{\partial \theta} [v^*] \right) \pm \zeta_n < \hat{\theta}_n - \theta_0, v^* > + \zeta_n \cdot o_p(n^{-1/2}),
$$

and this results in that

$$
\sqrt{n} < \hat{\theta}_n - \theta_0, v^* > = \sqrt{n} \mu_n \left( \frac{\partial l(\theta_0, W_i)}{\partial \theta} [v^*] \right) + o_p(1) \xrightarrow{d} \mathcal{N}(0, ||v^*||^2).
$$

By Assumption 15, we have

$$
\sqrt{n} \left( T(\hat{\theta}_n) - T(\theta_0) \right) = \sqrt{n} < \hat{\theta}_n - \theta_0, v^* > \xrightarrow{d} \mathcal{N}(0, ||v^*||^2)
$$

by the same way in CFT06.

**B.6 Proof of Theorem 4.3**

Define

$$
S_{\psi_0}' = \frac{\partial l(\theta_0, W)}{\partial \psi'} - \left\{ \frac{\partial l(\theta_0, W)}{\partial h_e} [b^*_e] + \frac{\partial l(\theta_0, W)}{\partial h_\nu} [b^*_\nu] \right\},
$$

(B.20)
where \( b^*_\epsilon = (b^*_{\epsilon 1}, ..., b^*_{\epsilon d_\nu}) \in \Pi_{k=1}^{d_\nu} (\mathcal{H}_\epsilon - \{ h_{\epsilon 0} \}) \) and \( b^*_\nu = (b^*_{\nu 1}, ..., b^*_{\nu d_\nu}) \in \Pi_{k=1}^{d_\nu} (\mathcal{H}_\nu - \{ h_{\nu 0} \}) \) are the solutions to the following optimization problems for \( k = 1, 2, ..., d_\nu \):

\[
\inf_{(b_{\epsilon k}, b_{\nu k}) \in \mathcal{V}_\epsilon \times \mathcal{V}_\nu} E \left[ \left( \frac{\partial l(\theta_0, W)}{\partial \theta_k} - \left\{ \frac{\partial l(\theta_0, W)}{\partial h_\epsilon} [b_{\epsilon k}] + \frac{\partial l(\theta_0, W)}{\partial h_\nu} [b_{\nu k}] \right\} \right)^2 \right].
\]

We consider the following assumption to establish the asymptotic normality for \( \psi_0 \).

**Assumption B.4.** \( \mathcal{I}_*(\psi_0) \equiv E[\mathcal{S}_{\psi_0} \mathcal{S}_{\psi_0}'] \) is non-singular.

To prove Theorem 4.3, take any arbitrary \( \lambda \in \mathbb{R}^{d_\psi} - \{0\} \) with \( |\lambda| \in (0, \infty) \) and let \( T : \Theta \to \mathbb{R} \) be a functional of the form \( T(\theta) = \lambda' \psi \). Then, for any \( v \in \mathcal{V} \), we have \( \partial T(\theta_0) [v] = \lambda' v_\psi \) and there exist a small \( \eta > 0 \) such that \( ||v|| \leq \eta \) and a constant \( \tilde{c} > 0 \) such that

\[
|T(\theta_0 + v) - T(\theta_0) - \partial T(\theta_0) [v]| \leq \tilde{c}||v||^w
\]

with \( w = \infty \). Therefore, Assumption 15-(i) is satisfied with \( w = \infty \) in this case. In addition, we have

\[
\sup_{v \in \mathcal{V} : ||v|| > 0} \frac{|\lambda' v_\psi|^2}{||v||^2} = \sup_{v \in \mathcal{V} : ||v|| > 0} \frac{|\lambda' v_\psi|^2}{E \left[ \left( \frac{\partial l(\theta_0, W)}{\partial \psi'} v_\psi + \sum_{j \in \{\epsilon, \nu\}} \frac{\partial l(\theta_0, W)}{\partial v_j} [v_j] \right)^2 \right]} = \lambda' E[\mathcal{S}_{\psi_0} \mathcal{S}_{\psi_0}']^{-1} \lambda = \lambda' \mathcal{I}_*(\theta_0)^{-1} \lambda.
\]

Note that the Riesz representer \( v^* \) exists if and only if \( \lambda' E[\mathcal{S}_{\psi_0} \mathcal{S}_{\psi_0}']^{-1} \lambda \) is finite. Since Assumption B.4 implies that \( \lambda' E[\mathcal{S}_{\psi_0} \mathcal{S}_{\psi_0}']^{-1} \lambda \) is finite, Assumption 15-(ii) holds. Hence, by Proposition 4.1, we have

\[
\sqrt{n} \left( \lambda' \hat{\psi}_n - \lambda' \psi_0 \right) \overset{d}{\to} \mathcal{N} \left( 0, \lambda' \mathcal{I}_*(\psi_0)^{-1} \lambda \right).
\]

Since \( \lambda \) was arbitrary, we obtain the result by Cramér-Wold device.

### B.7 Hölder ball

Suppose that \( h \in L^m_p(0,1) \), where \( p = m + \zeta, \) \( m \geq 0 \) is an integer and \( \zeta \in (0,1) \) is the Hölder exponent. We want to show that \( h^2 \in L^{p}_{\tilde{R}}(0,1) \), where \( \tilde{R} = R^{2^m+1} \). Recall that \( \mathcal{D} \) is the differential operator. We note that \( ||h||_{\infty} \leq R \) and thus \( \sup_{x} |\mathcal{D}^\omega h(x)| \leq R \) for all \( \omega \leq m \). By Leibniz’s formula, we have

\[
|\mathcal{D}^\omega h^2(x)| = \sum_{\iota \leq \omega} \binom{\omega}{\iota} |\mathcal{D}^{\iota} h \mathcal{D}^{\omega-\iota} h| \leq R^2 \sum_{\iota \leq \omega} \binom{\omega}{\iota} = R^2 2^\omega \leq K^2 2^m < \infty
\]
for all $\omega \leq m$. Observe that, by Leibniz’s formula, for any $x, y \in [0, 1]$ with $x \neq y$,

$$
|D^m h^2(x) - D^m h^2(y)| = \left| \sum_{\omega \leq m} \binom{m}{\omega} D^\omega h(x) D^{m-\omega} h(x) - \sum_{\omega \leq m} \binom{m}{\omega} D^\omega h(y) D^{m-\omega} h(y) \right|
$$

$$
\leq \left| \sum_{\omega \leq m} \binom{m}{\omega} D^\omega h(x) D^{m-\omega} h(x) - \sum_{\omega \leq m} \binom{m}{\omega} D^\omega h(y) D^{m-\omega} h(x) \right|
$$

$$
+ \left| \sum_{\omega \leq m} \binom{m}{\omega} D^\omega h(y) D^{m-\omega} h(x) - \sum_{\omega \leq m} \binom{m}{\omega} D^\omega h(y) D^{m-\omega} h(y) \right|
$$

$$
\leq 2 \times \left\{ \sup_{\omega \leq m} \sup_x |D^\omega h(x)| \right\} \times \left| \sum_{\omega \leq m} \binom{m}{\omega} \{D^\omega h(x) - D^\omega h(y)\} \right|
$$

$$
\leq 2R \sum_{\omega \leq m} \binom{m}{\omega} |D^\omega h(x) - D^\omega h(y)|.
$$

We also have that, for all $\omega < m$,

$$
\frac{|D^\omega h(x) - D^\omega h(y)|}{|x - y|^{1-\zeta}} = \frac{|D^\omega h(x) - D^\omega h(y)|}{|x - y|^{1-\zeta}} = |D^{\omega+1} h(\tilde{x})||x - y|^{1-\zeta} \leq R,
$$

where $\tilde{x}$ is between $x$ and $y$. Note that $\zeta \in (0, 1]$ and thus $|x - y|^{1-\zeta} \leq 1$ for all $x, y \in [0, 1]$. Since $h \in \Lambda^p_R([0, 1])$, we have $\frac{|D^m h(x) - D^m h(y)|}{|x - y|^\zeta} \leq R$. Hence,

$$
\frac{|D^m h^2(x) - D^m h^2(y)|}{|x - y|^{2\zeta}} \leq 2R \sum_{\omega \leq m} \binom{m}{\omega} \frac{|D^\omega h(x) - D^\omega h(y)|}{|x - y|^{\zeta}} \leq 2R^2 \sum_{\omega \leq m} \binom{m}{\omega} = R^2 2^{m+1} < \infty,
$$

and this implies that $h^2 \in \Lambda^p_{\tilde{R}}([0, 1])$ with $\tilde{R} = R^2 2^{m+1}$.

## C Additional Simulation Results

### C.1 A Larger Sample Size

Tables C.1 and C.2 show the simulation results with a larger sample size ($n = 1000$). We can see that the main findings in the main text remain the same even with this larger sample size.

### C.2 Copula and Marginal Misspecification

We consider the simulation results when both the copula and the marginal distributions are misspecified, reported in Tables C.3–C.6 and C.7–C.10. If both the copula and the marginal
distributions are misspecified, the performance of the parametric ML estimators are comparable to, or slightly worse than that under marginal misspecification. Consider, for example, the case where the true copula function is the Frank copula and the sample size is 500. The estimators of \( \psi \) under both the copula and marginal misspecification (Table C.4) have slightly larger root mean squared errors (RMSEs) than the corresponding estimators under the marginal misspecification (Table 2). On the other hand, the performance of the estimators of the ATE varies across copula specifications. In particular, when the true data generating process (DGP) is based on the Gumbel copula, the copula and marginal misspecification has a significant effect on the performance of the parametric estimators of the ATE. The RMSEs of the estimators of the ATE under the copula and marginal misspecification (Table C.6) are larger than those under the marginal misspecification (Table 2). Specifically, the RMSE of the parametric estimator of the ATE under the marginal misspecification is 0.1637 (Table 2), whereas the RMSEs of the corresponding estimators under both the copula and marginal misspecification are 0.1835, 0.2178, and 0.2732 when the Gaussian, Frank, and Clayton copulas are used, respectively (Table C.6). On the other hand, there is no clear evidence that the performance of the sieve ML estimators under both the copula and marginal misspecification is worse than that under misspecification of the marginal distributions. For example, when the true copula belongs to the Frank family but the copula is specified as the Gaussian or Gumbel copula, we can see that the RMSEs of the sieve ML estimators of the finite-dimensional parameters other than \( \gamma \) and the ATE under the copula and marginal misspecification (Table C.4) are lower than those under the marginal misspecification (Table 2). In contrast, we can see from the same tables that the Clayton copula specification draws the opposite conclusion when the true copula is the Frank. In general, no matter whether the copula is misspecified, we find that the sieve ML estimators outperform the parametric estimators in terms of the RMSE when the marginal distributions are misspecified.

C.3 Unknown Marginal Density Functions with Fat Tails

We examine the finite sample performance of the sieve ML estimator of \( \theta_0 \) when the unknown marginal density functions \( f_{\epsilon_0} \) and \( f_{\nu_0} \) have fat tails. We consider the \( t \) distribution with 3 degree of freedom as the true marginal distributions. While the marginal distributions in the parametric models are specified by normal distributions, we consider two specifications for the semiparametric models. These specifications differ in the choice of \( G \): we choose the standard normal distribution and the distribution function of \( t(3) \) for \( G \) in the first and second specifications, respectively. All simulation results are obtained with 500 observations and 2000 simulation iterations.

Table C.11 presents simulation results. While the parametric estimates have larger standard deviations, the biases of the semiparametric estimates are larger than those of the parametric estimates. However, the resulting RMSEs of the semiparametric estimates are slightly larger.
than those of the parametric estimates. This is because the semiparametric specification does not satisfy the assumptions required for the asymptotic theory.

Table C.12 shows simulation results where $G$ is the distribution function of $t(3)$. The performance of semiparametric estimator is comparable to that of parametric estimator in terms of the RMSE. The biases of semiparametric estimates in Table C.12 are much smaller than those in Table C.11, and the standard deviations of semiparametric estimates are very similar to those of parametric estimates.

The simulation results in Tables C.11 and C.12 suggest that if a researcher has a prior belief about the tail behavior of the unknown marginal density functions, it should be reflected in the choice of $G$ for semiparametric models. If it is believed that the marginal density functions have fat tails, one may choose a distribution function with fat tails for $G$, such as the distribution function of $t(3)$.

### C.4 Different Degrees of Dependence

Tables C.13 through C.18 provide simulation results across various degrees of dependence between $\epsilon$ and $\nu$. The dependence measure is unified into the Spearman’s $\rho$, and we consider cases of $\rho_{sp} \in \{-0.5, 0.2, 0.7\}$. We find that regardless of degrees of dependence, the results in our main paper remains the same: (i) the performance of the semiparametric estimator is comparable to that of the parametric estimator under correct specification, (ii) the semiparametric estimators outperform the parametric estimators under misspecification of the marginals.

### C.5 Coverage Probabilities of Bootstrap Confidence Intervals

We conduct simulations to investigate coverage probabilities of bootstrap confidence intervals (CIs). We consider the following design:

$$Y_i = 1\{-X_{1i} + X_{2i}\beta + D_i\delta \geq \epsilon_i\}, \quad D_i = 1\{-X_{1i} + X_{2i}\alpha + Z_i\gamma \geq \nu_i\},$$

where $(\alpha, \gamma, \beta, \delta) = (0.5, 0.8, 0.8, 1.1)$ and $(\epsilon, \nu)$ are generated from the Gaussian copula and normal marginals with $\rho_{sp} = 0.5$. $(X_{1i}, X_{2i}, Z_i)$ is drawn from a multivariate normal distribution. Note that the coefficients on $X_{1i}$ are fixed for scale normalization. The sample size, number of bootstrap iterations, and number of simulations are 500, 200, and 200, respectively. We consider two types of CIs: (i) CIs using the normal approximation, (ii) the percentile bootstrap CIs.

Table C.19 presents the coverage probabilities of both CIs. We find that the bootstrap percentile CIs performs better than the CIs based on the normal approximation and that their coverage probabilities are close to the nominal level (95%).

---

*Note that we only consider the Gaussian and Frank copulas for $\rho_{sp} = -0.5$ as the Clayton or the Gumbel copula does not allow for negative dependence.*
Table C.1: Correct Specification \((n = 1,000)\) (True marginal: normal)

| Parameter Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|--------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 |
| Estimate | 0.8025 | 1.1165 | 0.4996 | 0.3632 | Estimate | 0.8026 | 1.1205 | 0.5031 | 0.3596 |
| S.D | 0.0654 | 0.2737 | 0.1081 | 0.0656 | S.D | 0.0655 | 0.2939 | 0.1092 | 0.0668 |
| Bias | 0.0025 | 0.0165 | -0.0004 | -0.0011 | Bias | 0.0026 | 0.0205 | 0.0031 | -0.0048 |
| RMSE | 0.0655 | 0.2742 | 0.1081 | 0.0656 | RMSE | 0.0655 | 0.2946 | 0.1092 | 0.0670 |

| Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula |
|--------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 |
| Estimate | 0.8017 | 1.1188 | 0.5010 | 0.3635 | Estimate | 0.8007 | 1.1164 | 0.5042 | 0.3594 |
| S.D | 0.0658 | 0.2605 | 0.1023 | 0.0620 | S.D | 0.0652 | 0.2663 | 0.1066 | 0.0652 |
| Bias | 0.0017 | 0.0188 | -0.0009 | -0.0040 | Bias | 0.0007 | 0.0164 | 0.0042 | -0.0049 |
| RMSE | 0.0658 | 0.2612 | 0.1023 | 0.0620 | RMSE | 0.0652 | 0.2668 | 0.1067 | 0.0653 |

| Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|--------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 |
| Estimate | 0.8030 | 1.1055 | 0.5007 | 0.3621 | Estimate | 0.8029 | 1.1100 | 0.5035 | 0.3572 |
| S.D | 0.0658 | 0.2329 | 0.0958 | 0.0566 | S.D | 0.0659 | 0.2524 | 0.0964 | 0.0560 |
| Bias | 0.0030 | 0.0055 | -0.0007 | -0.0023 | Bias | 0.0029 | 0.0100 | 0.0035 | -0.0071 |
| RMSE | 0.0659 | 0.2330 | 0.0958 | 0.0567 | RMSE | 0.0660 | 0.2526 | 0.0965 | 0.0565 |

| Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|--------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3643 |
| Estimate | 0.8022 | 1.1192 | 0.4963 | 0.3644 | Estimate | 0.8025 | 1.1240 | 0.4986 | 0.3626 |
| S.D | 0.0668 | 0.2655 | 0.1057 | 0.0635 | S.D | 0.0665 | 0.2818 | 0.1086 | 0.0684 |
| Bias | 0.0022 | 0.0192 | -0.0037 | 0.0001 | Bias | 0.0025 | 0.0240 | -0.0014 | -0.0017 |
| RMSE | 0.0669 | 0.2662 | 0.1057 | 0.0635 | RMSE | 0.0665 | 0.2829 | 0.1086 | 0.0684 |
Table C.2: Misspecification of Marginals ($n = 1,000$) (True marginal: mixture of normals)

| Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|--------------------------------------|------------------------------------------|
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0911$                                 |
| Bias                                 | $-0.0059$                                |
| RMSE                                 | $0.0913$                                 |
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0911$                                 |
| Bias                                 | $-0.0059$                                |
| RMSE                                 | $0.0913$                                 |

| Parametric Estimation, Frank Copula   | Semiparametric Estimation, Frank Copula  |
|--------------------------------------|------------------------------------------|
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0899$                                 |
| Bias                                 | $0.0044$                                 |
| RMSE                                 | $0.0901$                                 |
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0899$                                 |
| Bias                                 | $0.0044$                                 |
| RMSE                                 | $0.0901$                                 |

| Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|--------------------------------------|------------------------------------------|
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0906$                                 |
| Bias                                 | $0.0065$                                 |
| RMSE                                 | $0.0908$                                 |
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0906$                                 |
| Bias                                 | $0.0065$                                 |
| RMSE                                 | $0.0908$                                 |

| Parametric Estimation, Gumbel Copula  | Semiparametric Estimation, Gumbel Copula |
|--------------------------------------|------------------------------------------|
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0893$                                 |
| Bias                                 | $-0.0151$                                |
| RMSE                                 | $0.0906$                                 |
| True Values                          | Estimate                                 |
| $\gamma$                             | $0.8000$                                 |
| $\delta_1$                           | $1.1000$                                 |
| $\rho_{sp}$                          | $0.5000$                                 |
| $ATE$                                | $0.1066$                                 |
| S.D                                  | $0.0893$                                 |
| Bias                                 | $-0.0151$                                |
| RMSE                                 | $0.0906$                                 |
Table C.3: Copula and Marginals Misspecification 1 \((n = 500)\) (True copula: Gaussian, true marginal: mixture of normals)

| Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula |
|-------------------------------------|----------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.1066 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.1066 |
| Estimate | 0.8140 | 1.3080 | 0.3775 | 0.2916 | Estimate | 0.8463 | 1.3514 | 0.4499 | 0.1351 |
| S.D | 0.1257 | 0.4899 | 0.1202 | 0.0862 | S.D | 0.1137 | 0.3502 | 0.0964 | 0.0686 |
| Bias | 0.0140 | 0.2080 | -0.1225 | 0.1849 | Bias | 0.0463 | 0.2514 | -0.0501 | 0.0285 |
| RMSE | 0.1265 | 0.5322 | 0.1716 | 0.2040 | RMSE | 0.1227 | 0.4311 | 0.1087 | 0.0743 |

| Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|--------------------------------------|-------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.1066 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.1066 |
| Estimate | 0.8244 | 1.5699 | 0.3691 | 0.3176 | Estimate | 0.8534 | 1.4386 | 0.4945 | 0.1586 |
| S.D | 0.1271 | 0.6609 | 0.1697 | 0.0999 | S.D | 0.1128 | 0.3398 | 0.1044 | 0.0734 |
| Bias | 0.0244 | 0.4699 | -0.1308 | 0.2110 | Bias | 0.0534 | 0.3386 | -0.0054 | 0.0520 |
| RMSE | 0.1294 | 0.8109 | 0.2143 | 0.2335 | RMSE | 0.1248 | 0.4797 | 0.1046 | 0.0899 |

| Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|--------------------------------------|-------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.1066 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.1066 |
| Estimate | 0.7981 | 1.0706 | 0.4232 | 0.2448 | Estimate | 0.8546 | 1.2025 | 0.4967 | 0.1137 |
| S.D | 0.1281 | 0.5795 | 0.1519 | 0.1077 | S.D | 0.1118 | 0.3611 | 0.1027 | 0.0600 |
| Bias | -0.0019 | -0.0294 | -0.0767 | 0.1382 | Bias | 0.0546 | 0.1025 | -0.0302 | 0.0070 |
| RMSE | 0.1281 | 0.5802 | 0.1702 | 0.1752 | RMSE | 0.1244 | 0.3754 | 0.1070 | 0.0604 |
Table C.4: Copula and Marginals Misspecification 2 ($n = 500$) (True copula: Frank, true marginal: mixture of normals)

|                              | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula | Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula | Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|------------------------------|----------------------------------------|--------------------------------------------|---------------------------------------|------------------------------------------|--------------------------------------|------------------------------------------|
|                              | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values                  | 0.8000   | 1.1000      | 0.5000      | 0.1066 | True Values | 0.8000   | 1.1000      | 0.5000      | 0.1066 |
| Estimate                     | 0.7992   | 1.1673      | 0.4517      | 0.2527 | Estimate    | 0.8500   | 1.1788      | 0.5173      | 0.1192 |
| S.D                          | 0.1342   | 0.6901      | 0.1680      | 0.1179 | S.D         | 0.1158   | 0.3602      | 0.1000      | 0.0652 |
| Bias                         | -0.0008  | 0.0673      | -0.0483     | 0.1461 | Bias        | 0.0500   | 0.0788      | 0.0173      | 0.0126 |
| RMSE                         | 0.1342   | 0.6934      | 0.1748      | 0.1877 | RMSE        | 0.1262   | 0.3687      | 0.1015      | 0.0664 |

|                              | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values                  | 0.8000   | 1.1000      | 0.5000      | 0.1066 | True Values | 0.8000   | 1.1000      | 0.5000      | 0.1066 |
| Estimate                     | 0.8235   | 1.6132      | 0.3870      | 0.3184 | Estimate    | 0.8484   | 1.3679      | 0.5212      | 0.1548 |
| S.D                          | 0.1329   | 0.7039      | 0.1670      | 0.1018 | S.D         | 0.1188   | 0.3416      | 0.1012      | 0.0755 |
| Bias                         | 0.0235   | 0.5132      | -0.1130     | 0.2118 | Bias        | 0.0484   | 0.2679      | 0.0212      | 0.0482 |
| RMSE                         | 0.1350   | 0.8711      | 0.2017      | 0.2350 | RMSE        | 0.1283   | 0.4341      | 0.1034      | 0.0896 |

|                              | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values                  | 0.8000   | 1.1000      | 0.5000      | 0.1066 | True Values | 0.8000   | 1.1000      | 0.5000      | 0.1066 |
| Estimate                     | 0.8001   | 1.1697      | 0.4202      | 0.2564 | Estimate    | 0.8485   | 1.1059      | 0.4997      | 0.1071 |
| S.D                          | 0.1347   | 0.6697      | 0.1608      | 0.1165 | S.D         | 0.1161   | 0.3548      | 0.0997      | 0.0601 |
| Bias                         | 0.0001   | 0.0697      | -0.0798     | 0.1498 | Bias        | 0.0485   | 0.0059      | -0.0003     | 0.0005 |
| RMSE                         | 0.1347   | 0.6733      | 0.1795      | 0.1897 | RMSE        | 0.1258   | 0.3548      | 0.0997      | 0.0601 |

Electronic copy available at: https://ssrn.com/abstract=3346834
Table C.5: Copula and Marginals Misspecification 3 ($n = 500$) (True copula: Clayton, true marginal: mixture of normals)

|                        | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula | Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula | Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|------------------------|----------------------------------------|--------------------------------------------|------------------------------------|----------------------------------------|--------------------------------------|----------------------------------------|
| True Values            | 0.8000 1.1000 0.5000 0.1066            | 0.8000 1.1000 0.5000 0.1066                | 0.8000 1.1000 0.5000 0.1066        | 0.8000 1.1000 0.5000 0.1066            | 0.8000 1.1000 0.5000 0.1066          | 0.8000 1.1000 0.5000 0.1066          |
| Estimate               | 0.7986 1.0471 0.4017 0.2392            | 0.8533 1.1780 0.4493 0.1076               | 0.8083 1.1559 0.3611 0.2712        | 0.8412 1.2404 0.4199 0.1160            | 0.8046 1.1937 0.3316 0.2680          | 0.8542 1.1610 0.4148 0.1046          |
| S.D                   | 0.1346 0.6366 0.1731 0.1181            | 0.1164 0.3438 0.1033 0.0569               | 0.1318 0.4453 0.1143 0.0856        | 0.1166 0.3408 0.0965 0.0611            | 0.1355 0.6663 0.1748 0.1220          | 0.1166 0.3283 0.1032 0.0557          |
| Bias                  | -0.0014 -0.0529 -0.0983 0.1325         | Bias 0.0533 0.0780 -0.0508 0.0009         | Bias 0.0559 -0.1389 0.1646         | Bias 0.0412 0.1404 -0.0802 0.0094     | Bias 0.0046 -0.1684 0.1613          | Bias 0.0542 0.0610 -0.0852 -0.0020   |
| RMSE                  | 0.1346 0.6388 0.1991 0.1775            | RMSE 0.1281 0.3525 0.1151 0.0569         | RMSE 0.1321 0.4488 0.1799 0.1855   | RMSE 0.1237 0.3686 0.1255 0.0619      | RMSE 0.1356 0.6728 0.2427 0.2022    | RMSE 0.1285 0.3339 0.1339 0.0557    |
Table C.6: Copula and Marginals Misspecification 4 ($n = 500$) (True copula: Gumbel, true marginal: mixture of normals)

|                        | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|------------------------|---------------------------------------|--------------------------------------------|
|                        | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values            | 0.8000   | 1.0000      | 0.5000       | 0.1066 | True Values | 0.8000   | 1.0000      | 0.5000       | 0.1066 |
| Estimate               | 0.7978   | 1.1488      | 0.4658       | 0.2523 | Estimate   | 0.8609   | 1.3801      | 0.4957       | 0.1460 |
| S.D                   | 0.1304   | 0.6489      | 0.1598       | 0.1117 | S.D        | 0.1132   | 0.3749      | 0.1052       | 0.0730 |
| Bias                  | -0.0022  | 0.0488      | -0.0342      | 0.1456 | Bias       | 0.0609   | 0.2801      | -0.0042      | 0.0393 |
| RMSE                  | 0.1304   | 0.6508      | 0.1634       | 0.1835 | RMSE       | 0.1286   | 0.4679      | 0.1053       | 0.0829 |

|                        | Parametric Estimation, Frank Copula   | Semiparametric Estimation, Frank Copula   |
|------------------------|---------------------------------------|--------------------------------------------|
|                        | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values            | 0.8000   | 1.0000      | 0.5000       | 0.1066 | True Values | 0.8000   | 1.0000      | 0.5000       | 0.1066 |
| Estimate               | 0.8140   | 1.4128      | 0.3834       | 0.3064 | Estimate   | 0.8532   | 1.4755      | 0.4543       | 0.1611 |
| S.D                   | 0.1290   | 0.5211      | 0.1184       | 0.0867 | S.D        | 0.1177   | 0.3466      | 0.0969       | 0.0752 |
| Bias                  | 0.0140   | 0.3128      | -0.1166      | 0.1998 | Bias       | 0.0532   | 0.3755      | -0.0457      | 0.0545 |
| RMSE                  | 0.1297   | 0.6078      | 0.1662       | 0.2178 | RMSE       | 0.1292   | 0.5110      | 0.1072       | 0.0929 |

|                        | Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|------------------------|---------------------------------------|--------------------------------------------|
|                        | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values            | 0.8000   | 1.1000      | 0.5000       | 0.1066 | True Values | 0.8000   | 1.0000      | 0.5000       | 0.1066 |
| Estimate               | 0.8276   | 1.8999      | 0.3208       | 0.3614 | Estimate   | 0.8603   | 1.6010      | 0.4823       | 0.1960 |
| S.D                   | 0.1321   | 0.7365      | 0.1753       | 0.0986 | S.D        | 0.1172   | 0.3103      | 0.1065       | 0.0799 |
| Bias                  | 0.0276   | 0.7999      | -0.1791      | 0.2548 | Bias       | 0.0603   | 0.5010      | -0.0177      | 0.0894 |
| RMSE                  | 0.1350   | 1.0873      | 0.2506       | 0.2732 | RMSE       | 0.1318   | 0.5893      | 0.1079       | 0.1199 |

Electronic copy available at: https://ssrn.com/abstract=3346834
Table C.7: Copula and Marginals Misspecification 1 \((n = 1,000)\) (True copula: Gaussian, true marginal: mixture of normals)

|                              | Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula |
|------------------------------|------------------------------------|----------------------------------------|
| True Values                  | \(0.8000\) 1.1000 0.5000 0.1066    | True Values \(0.8000\) 1.1000 0.5000 0.1066 |
| Estimate                     | \(0.8086\) 1.3159 0.3652 0.2975    | Estimate \(0.8549\) 1.3936 0.4376 0.1371 |
| S.D                          | \(0.0897\) 0.3636 0.0927 0.0650    | S.D \(0.0830\) 0.2548 0.0689 0.0506     |
| Bias                         | \(0.0086\) 0.2159 -0.1347 0.1909   | Bias \(0.0549\) 0.2936 -0.0623 0.0305   |
| RMSE                         | \(0.0901\) 0.4229 0.1636 0.2017    | RMSE \(0.0995\) 0.3887 0.0929 0.0591    |

|                              | Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|------------------------------|--------------------------------------|-------------------------------------------|
| True Values                  | \(0.8000\) 1.1000 0.5000 0.1066      | True Values \(0.8000\) 1.1000 0.5000 0.1066 |
| Estimate                     | \(0.8193\) 1.5478 0.3661 0.3205      | Estimate \(0.8613\) 1.4684 0.4886 0.1574  |
| S.D                          | \(0.0906\) 0.4574 0.1217 0.0705      | S.D \(0.0812\) 0.2351 0.0710 0.0514      |
| Bias                         | \(0.0193\) 0.4478 -0.1338 0.2139     | Bias \(0.0613\) 0.3684 -0.0113 0.0508    |
| RMSE                         | \(0.0927\) 0.6401 0.1809 0.2252      | RMSE \(0.1018\) 0.4370 0.0719 0.0722    |

|                              | Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|------------------------------|-------------------------------------|-----------------------------------------|
| True Values                  | \(0.8000\) 1.1000 0.5000 0.1066    | True Values \(0.8000\) 1.1000 0.5000 0.1066 |
| Estimate                     | \(0.7930\) 1.0391 0.4210 0.2453    | Estimate \(0.8620\) 1.2302 0.4574 0.1157  |
| S.D                          | \(0.0911\) 0.4010 0.1070 0.0771    | S.D \(0.0790\) 0.2554 0.0709 0.0439     |
| Bias                         | \(-0.0070\) -0.0609 -0.0789 0.1386 | Bias \(0.0620\) 0.1302 -0.0426 0.0090   |
| RMSE                         | \(0.0914\) 0.4056 0.1330 0.1586    | RMSE \(0.1004\) 0.2867 0.0827 0.0449    |
Table C.8: Copula and Marginals Misspecification 2 ($n = 1,000$) (True copula: Frank, true marginal: mixture of normals)

|                     | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|---------------------|----------------------------------------|--------------------------------------------|
|                     | $\hat{\gamma}$ | $\hat{\delta}_1$ | $\hat{\rho}_{sp}$ | $\hat{ATE}$ | $\hat{\gamma}$ | $\hat{\delta}_1$ | $\hat{\rho}_{sp}$ | $\hat{ATE}$ |
| **True Values**     | 0.8000 | 1.1000 | 0.5000 | 0.1066 | 0.8000 | 1.1000 | 0.5000 | 0.1066 |
| **Estimate**        | 0.7935 | 1.0825 | 0.4653 | 0.2465 | 0.8601 | 1.1832 | 0.5145 | 0.1196 |
| **S.D**             | 0.0926 | 0.4333 | 0.1152 | 0.0803 | 0.0768 | 0.2641 | 0.0723 | 0.0450 |
| **Bias**            | -0.0065 | -0.0175 | -0.0347 | 0.1399 | 0.0601 | 0.0832 | 0.0145 | 0.0130 |
| **RMSE**            | 0.0929 | 0.4336 | 0.1203 | 0.1613 | 0.0976 | 0.2769 | 0.0738 | 0.0468 |
|                     | Parametric Estimation, Clayton Copula  | Semiparametric Estimation, Clayton Copula  |
|                     | $\hat{\gamma}$ | $\hat{\delta}_1$ | $\hat{\rho}_{sp}$ | $\hat{ATE}$ | $\hat{\gamma}$ | $\hat{\delta}_1$ | $\hat{\rho}_{sp}$ | $\hat{ATE}$ |
| **True Values**     | 0.8000 | 1.1000 | 0.5000 | 0.1066 | 0.8000 | 1.1000 | 0.5000 | 0.1066 |
| **Estimate**        | 0.8188 | 1.5580 | 0.3941 | 0.3173 | 0.8583 | 1.3743 | 0.5200 | 0.1542 |
| **S.D**             | 0.0919 | 0.4621 | 0.1194 | 0.0708 | 0.0794 | 0.2439 | 0.0718 | 0.0526 |
| **Bias**            | 0.0188 | 0.4580 | -0.1059 | 0.2106 | 0.0583 | 0.2743 | 0.0200 | 0.0476 |
| **RMSE**            | 0.0938 | 0.6506 | 0.1595 | 0.2222 | 0.0985 | 0.3671 | 0.0746 | 0.0709 |
|                     | Parametric Estimation, Gumbel Copula  | Semiparametric Estimation, Gumbel Copula  |
|                     | $\hat{\gamma}$ | $\hat{\delta}_1$ | $\hat{\rho}_{sp}$ | $\hat{ATE}$ | $\hat{\gamma}$ | $\hat{\delta}_1$ | $\hat{\rho}_{sp}$ | $\hat{ATE}$ |
| **True Values**     | 0.8000 | 1.1000 | 0.5000 | 0.1066 | 0.8000 | 1.1000 | 0.5000 | 0.1066 |
| **Estimate**        | 0.7954 | 1.0843 | 0.4327 | 0.2496 | 0.8596 | 1.1105 | 0.4959 | 0.1082 |
| **S.D**             | 0.0927 | 0.4252 | 0.1119 | 0.0796 | 0.0765 | 0.2578 | 0.0708 | 0.0413 |
| **Bias**            | -0.0046 | -0.0157 | -0.0673 | 0.1429 | 0.0596 | 0.0105 | -0.0041 | 0.0016 |
| **RMSE**            | 0.0928 | 0.4255 | 0.1306 | 0.1636 | 0.0970 | 0.2580 | 0.0709 | 0.0413 |
Table C.9: Copula and Marginals Misspecification 3 \((n = 1,000)\) (True copula: Clayton, true marginal: mixture of normals)

|                           | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|---------------------------|---------------------------------------|------------------------------------------|
|                           | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values               | 0.8000     | 1.1000     | 0.5000     | 0.1066   | True Values | 0.8000     | 1.1000     | 0.5000     | 0.1066   |
| Estimate                  | 0.7928     | 0.9952     | 0.4102     | 0.2370   | Estimate    | 0.8618     | 1.2015     | 0.4441     | 0.1097   |
| S.D                       | 0.0929     | 0.4262     | 0.1233     | 0.0837   | S.D         | 0.0764     | 0.2527     | 0.0737     | 0.0411   |
| Bias                      | -0.0072    | -0.1048    | -0.0898    | 0.1303   | Bias        | 0.0618     | 0.1015     | -0.0559    | 0.0030   |
| RMSE                      | 0.0932     | 0.4389     | 0.1525     | 0.1549   | RMSE        | 0.0983     | 0.2723     | 0.0925     | 0.0412   |

|                           | Parametric Estimation, Frank Copula   | Semiparametric Estimation, Frank Copula  |
|---------------------------|---------------------------------------|-------------------------------------|
|                           | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values               | 0.8000     | 1.1000     | 0.5000     | 0.1066   | True Values | 0.8000     | 1.1000     | 0.5000     | 0.1066   |
| Estimate                  | 0.8048     | 1.1667     | 0.3480     | 0.2754   | Estimate    | 0.8510     | 1.2695     | 0.4101     | 0.1152   |
| S.D                       | 0.0910     | 0.3362     | 0.0918     | 0.0649   | S.D         | 0.0825     | 0.2578     | 0.0701     | 0.0453   |
| Bias                      | 0.0048     | 0.0667     | -0.1520    | 0.1688   | Bias        | 0.0510     | 0.1695     | -0.0899    | 0.0086   |
| RMSE                      | 0.0911     | 0.3428     | 0.1776     | 0.1808   | RMSE        | 0.0970     | 0.3085     | 0.1140     | 0.0461   |

|                           | Parametric Estimation, Gumbel Copula  | Semiparametric Estimation, Gumbel Copula |
|---------------------------|--------------------------------------|-----------------------------------------|
|                           | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values               | 0.8000     | 1.1000     | 0.5000     | 0.1066   | True Values | 0.8000     | 1.1000     | 0.5000     | 0.1066   |
| Estimate                  | 0.8046     | 1.1937     | 0.3316     | 0.2680   | Estimate    | 0.8594     | 1.1883     | 0.4090     | 0.1054   |
| S.D                       | 0.1355     | 0.6663     | 0.1748     | 0.1220   | S.D         | 0.0784     | 0.2373     | 0.0727     | 0.0412   |
| Bias                      | 0.0046     | 0.0937     | -0.1684    | 0.1613   | Bias        | 0.0594     | 0.0883     | -0.0911    | -0.0013  |
| RMSE                      | 0.1356     | 0.6728     | 0.2427     | 0.2022   | RMSE        | 0.0984     | 0.2532     | 0.1165     | 0.0412   |
Table C.10: Copula and Marginals Misspecification 4 ($n = 1,000$) (True copula: Gumbel, true DGP marginal: mixture of normals)

|                      | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|----------------------|----------------------------------------|--------------------------------------------|
|                      | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values          | 0.8000   | 1.1000     | 0.5000       | 0.1066 | True Values | 0.8000 | 1.1000   | 0.5000       | 0.1066 |
| Estimate             | 0.7905   | 1.1059     | 0.4669       | 0.2520 | Estimate   | 0.8660 | 1.4046   | 0.4893       | 0.1428 |
| S.D                  | 0.0896   | 0.4412     | 0.1167       | 0.0815 | S.D        | 0.0775 | 0.2644   | 0.0723       | 0.0508 |
| Bias                 | -0.0095  | 0.0059     | -0.0330      | 0.1454 | Bias       | 0.0660 | 0.3046   | -0.0107      | 0.0362 |
| RMSE                 | 0.0901   | 0.4412     | 0.1213       | 0.1667 | RMSE       | 0.1018 | 0.4034   | 0.0730       | 0.0624 |

|                      | Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula |
|----------------------|-------------------------------------|------------------------------------------|
|                      | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values          | 0.8000   | 1.0000     | 0.5000       | 0.1066 | True Values | 0.8000 | 1.0000   | 0.5000       | 0.1066 |
| Estimate             | 0.8123   | 1.4374     | 0.3701       | 0.3149 | Estimate   | 0.8628 | 1.5142   | 0.4473       | 0.1582 |
| S.D                  | 0.0901   | 0.3917     | 0.0930       | 0.0651 | S.D        | 0.0817 | 0.2377   | 0.0697       | 0.0545 |
| Bias                 | 0.0123   | 0.3374     | -0.1299      | 0.2083 | Bias       | 0.0628 | 0.4142   | -0.0526      | 0.0515 |
| RMSE                 | 0.0910   | 0.5169     | 0.1597       | 0.2182 | RMSE       | 0.1030 | 0.4776   | 0.0874       | 0.0750 |

|                      | Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|----------------------|--------------------------------------|--------------------------------------------|
|                      | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta_1$ | $\rho_{sp}$ | $ATE$ |
| True Values          | 0.8000   | 1.1000     | 0.5000       | 0.1066 | True Values | 0.8000 | 1.1000   | 0.5000       | 0.1066 |
| Estimate             | 0.8228   | 1.8913     | 0.3197       | 0.3656 | Estimate   | 0.8645 | 1.6249   | 0.4851       | 0.1894 |
| S.D                  | 0.0927   | 0.5234     | 0.1336       | 0.0714 | S.D        | 0.0808 | 0.2084   | 0.0742       | 0.0550 |
| Bias                 | 0.0228   | 0.7913     | -0.1803      | 0.2589 | Bias       | 0.0645 | 0.5249   | -0.0149      | 0.0828 |
| RMSE                 | 0.0955   | 0.9488     | 0.2244       | 0.2686 | RMSE       | 0.1034 | 0.5648   | 0.0757       | 0.0994 |
Table C.11: Misspecification of Marginals \((n = 500)\) (True Marginal: \(t(3)\))

| Parametric Estimation, Gaussian Copula | Semiparametric Estimation†, Gaussian Copula |
|--------------------------------------|---------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 |
| Estimate | 0.8060 | 1.1937 | 0.4938 | 0.3098 | Estimate | 0.7288 | 0.8480 | 0.5762 | 0.2499 |
| S.D | 0.1119 | 0.5749 | 0.1647 | 0.1068 | S.D | 0.1037 | 0.3832 | 0.1339 | 0.1085 |
| Bias | 0.0060 | 0.0937 | -0.0062 | -0.0143 | Bias | -0.0712 | -0.2520 | 0.0763 | -0.0742 |
| RMSE | 0.0125 | 0.3306 | 0.0271 | 0.0116 | RMSE | 0.0108 | 0.1468 | 0.0179 | 0.0173 |

| Parametric Estimation, Frank Copula | Semiparametric Estimation†, Frank Copula |
|-------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 |
| Estimate | 0.8080 | 1.0964 | 0.5164 | 0.2919 | Estimate | 0.7370 | 0.8618 | 0.5763 | 0.2459 |
| S.D | 0.1139 | 0.4602 | 0.1347 | 0.0922 | S.D | 0.1030 | 0.3226 | 0.1041 | 0.0884 |
| Bias | 0.0080 | -0.0036 | 0.0164 | -0.0323 | Bias | -0.0630 | -0.2382 | 0.0763 | -0.0783 |
| RMSE | 0.0130 | 0.2118 | 0.0181 | 0.0096 | RMSE | 0.0106 | 0.1041 | 0.0108 | 0.0139 |

| Parametric Estimation, Clayton Copula | Semiparametric Estimation†, Clayton Copula |
|--------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 |
| Estimate | 0.8000 | 1.0964 | 0.5164 | 0.2919 | Estimate | 0.7370 | 0.8618 | 0.5763 | 0.2459 |
| S.D | 0.1145 | 0.4385 | 0.1357 | 0.0946 | S.D | 0.1044 | 0.3479 | 0.1250 | 0.0989 |
| Bias | 0.0000 | -0.0798 | 0.0385 | -0.0456 | Bias | -0.0670 | -0.2554 | 0.0689 | -0.0695 |
| RMSE | 0.0131 | 0.1923 | 0.0184 | 0.0110 | RMSE | 0.0109 | 0.1210 | 0.0156 | 0.0146 |

| Parametric Estimation, Gumbel Copula | Semiparametric Estimation†, Gumbel Copula |
|--------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 | True Values | 0.8000 | 1.1000 | 0.5000 | 0.3242 |
| Estimate | 0.8098 | 1.2599 | 0.4767 | 0.3205 | Estimate | 0.7344 | 0.8905 | 0.5628 | 0.2559 |
| S.D | 0.1153 | 0.6137 | 0.1732 | 0.1140 | S.D | 0.1045 | 0.4106 | 0.1461 | 0.1144 |
| Bias | 0.0098 | 0.1599 | -0.0233 | -0.0037 | Bias | -0.0656 | -0.2095 | 0.0628 | -0.0682 |
| RMSE | 0.0133 | 0.3767 | 0.0300 | 0.0130 | RMSE | 0.0109 | 0.1686 | 0.0213 | 0.0177 |

†: The semiparametric models are specified with \(G = \Phi\), where \(\Phi(\cdot)\) is the standard normal distribution function.

Electronic copy available at: https://ssrn.com/abstract=3346834
Table C.12: Misspecification of Marginals \((n = 500)\) (True marginal: \(t(3)\))

|                     | Parametric Estimation, Gaussian Copula | Semiparametric Estimation†, Gaussian Copula |
|---------------------|----------------------------------------|---------------------------------------------|
|                     | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values         | 0.8000     | 1.1000       | 0.5000        | 0.3242   | True Values | 0.8000     | 1.1000       | 0.5000        | 0.3242   |
| Estimate            | 0.8124     | 1.1948       | 0.4913        | 0.3098   | Estimate    | 0.8098     | 1.1957       | 0.4930        | 0.3252   |
| S.D                 | 0.1149     | 0.5540       | 0.1626        | 0.1068   | S.D         | 0.1146     | 0.5905       | 0.1639        | 0.1107   |
| Bias                | 0.0124     | 0.0948       | -0.0086       | -0.0143  | Bias        | 0.0098     | 0.0957       | -0.0069       | 0.0010   |
| RMSE                | 0.0132     | 0.3069       | 0.0264        | 0.0116   | RMSE        | 0.0131     | 0.3487       | 0.0269        | 0.0123   |

|                     | Parametric Estimation, Frank Copula   | Semiparametric Estimation†, Frank Copula   |
|---------------------|---------------------------------------|--------------------------------------------|
|                     | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values         | 0.8000     | 1.1000       | 0.5000        | 0.3242   | True Values | 0.8000     | 1.1000       | 0.5000        | 0.3242   |
| Estimate            | 0.8063     | 1.0832       | 0.5246        | 0.2873   | Estimate    | 0.8087     | 1.1877       | 0.4930        | 0.3257   |
| S.D                 | 0.1152     | 0.4741       | 0.1354        | 0.0941   | S.D         | 0.1153     | 0.5132       | 0.1397        | 0.0971   |
| Bias                | 0.0063     | -0.0168      | 0.0246        | -0.0369  | Bias        | 0.0087     | 0.0877       | -0.0047       | 0.0015   |
| RMSE                | 0.0133     | 0.2247       | 0.0183        | 0.0102   | RMSE        | 0.0133     | 0.2633       | 0.0195        | 0.0094   |

|                     | Parametric Estimation, Clayton Copula | Semiparametric Estimation†, Clayton Copula |
|---------------------|--------------------------------------|------------------------------------------|
|                     | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values         | 0.8000     | 1.1000       | 0.5000        | 0.3242   | True Values | 0.8000     | 1.1000       | 0.5000        | 0.3242   |
| Estimate            | 0.8067     | 1.0312       | 0.5354        | 0.2797   | Estimate    | 0.8117     | 1.1871       | 0.4972        | 0.3172   |
| S.D                 | 0.1161     | 0.4525       | 0.1365        | 0.0950   | S.D         | 0.1163     | 0.5845       | 0.1500        | 0.0998   |
| Bias                | 0.0067     | -0.0688      | 0.0354        | -0.0445  | Bias        | 0.0117     | 0.0871       | -0.0028       | -0.0070  |
| RMSE                | 0.0135     | 0.2048       | 0.0186        | 0.0110   | RMSE        | 0.0135     | 0.3416       | 0.0225        | 0.0100   |

|                     | Parametric Estimation, Gumbel Copula | Semiparametric Estimation†, Gumbel Copula |
|---------------------|-------------------------------------|------------------------------------------|
|                     | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values         | 0.8000     | 1.1000       | 0.5000        | 0.3242   | True Values | 0.8000     | 1.1000       | 0.5000        | 0.3242   |
| Estimate            | 0.8101     | 1.2629       | 0.4780        | 0.3213   | Estimate    | 0.8062     | 1.1713       | 0.5024        | 0.3225   |
| S.D                 | 0.1153     | 0.5991       | 0.1711        | 0.1113   | S.D         | 0.1153     | 0.5477       | 0.1561        | 0.1103   |
| Bias                | 0.0101     | 0.1629       | -0.0220       | -0.0029  | Bias        | 0.0062     | 0.0713       | 0.0024        | -0.0017  |
| RMSE                | 0.0133     | 0.3589       | 0.0293        | 0.0124   | RMSE        | 0.0133     | 0.3000       | 0.0244        | 0.0122   |

†: The semiparametric models are specified with \(G = F_{t_3}\), where \(F_{t_3}\) is the distribution function of \(t(3)\).
Table C.13: Correct Specification ($n = 500$, $\rho_{sp} = 0.2$) (True marginal: normal)

|                         | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|-------------------------|---------------------------------------|--------------------------------------------|
| True Values             | 0.8000 1.1000 0.2000 0.3643            | True Values 0.8000 1.1000 0.2000 0.3643    |
| Estimate                | 0.8026 1.1342 0.2093 0.3643            | Estimate 0.8026 1.1342 0.2093 0.3526       |
| S.D                    | 0.0945 0.4199 0.1840 0.0963            | S.D 0.0945 0.4199 0.1840 0.0952           |
| Bias                   | 0.0026 0.0342 0.0093 0.0000            | Bias 0.0026 0.0342 0.0093 -0.0117         |
| RMSE                   | 0.0089 0.1763 0.0399 0.0093            | RMSE 0.0089 0.1763 0.0339 0.0092          |

|                         | Parametric Estimation, Frank Copula   | Semiparametric Estimation, Frank Copula   |
| True Values             | 0.8000 1.1000 0.2000 0.3643            | True Values 0.8000 1.1000 0.2000 0.3643    |
| Estimate                | 0.8037 1.0818 0.2216 0.3517            | Estimate 0.8051 1.0905 0.2278 0.3448       |
| S.D                    | 0.0974 0.3309 0.1468 0.0807            | S.D 0.0981 0.3591 0.1443 0.0856           |
| Bias                   | 0.0037 -0.0182 0.0215 -0.0126          | Bias 0.0051 -0.0095 0.0277 -0.0195        |
| RMSE                   | 0.0095 0.1095 0.0215 0.0067            | RMSE 0.0096 0.1290 0.0208 0.0077          |

|                         | Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
| True Values             | 0.8000 1.1000 0.1999 0.3643            | True Values 0.8000 1.1000 0.1999 0.3643    |
| Estimate                | 0.8036 1.0973 0.2138 0.3571            | Estimate 0.8046 1.1040 0.2216 0.3492       |
| S.D                    | 0.0934 0.3170 0.1498 0.0773            | S.D 0.0936 0.3593 0.1533 0.0818           |
| Bias                   | 0.0036 -0.0027 0.0139 -0.0072          | Bias 0.0046 0.0040 0.0217 -0.0151         |
| RMSE                   | 0.0087 0.1005 0.0224 0.0060            | RMSE 0.0088 0.1291 0.0235 0.0069          |

|                         | Parametric Estimation, Gumbel Copula  | Semiparametric Estimation, Gumbel Copula  |
| True Values             | 0.8000 1.1000 0.2000 0.3643            | True Values 0.8000 1.1000 0.2000 0.3643    |
| Estimate                | 0.8017 1.0886 0.2175 0.3524            | Estimate 0.8033 1.1245 0.2170 0.3495       |
| S.D                    | 0.0940 0.3640 0.1519 0.0867            | S.D 0.0954 0.4176 0.1578 0.0927           |
| Bias                   | 0.0017 -0.0114 0.0175 -0.0119          | Bias 0.0033 0.0245 0.0170 -0.0149         |
| RMSE                   | 0.0088 0.1325 0.0231 0.0077            | RMSE 0.0091 0.1744 0.0249 0.0088          |
Table C.14: Misspecification of Marginals \((n = 500, \rho_{sp} = 0.2)\) (True marginal: mixture of normals)

|                     | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula | Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula | Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula | Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|---------------------|----------------------------------------|--------------------------------------------|-------------------------------------|-----------------------------------------|--------------------------------------|----------------------------------------|-------------------------------------|----------------------------------------|
|                     | \(\gamma\)  | \(\delta_1\)  | \(\rho_{sp}\)  | ATE  | \(\gamma\)  | \(\delta_1\)  | \(\rho_{sp}\)  | ATE  | \(\gamma\)  | \(\delta_1\)  | \(\rho_{sp}\)  | ATE  | \(\gamma\)  | \(\delta_1\)  | \(\rho_{sp}\)  | ATE  | \(\gamma\)  | \(\delta_1\)  | \(\rho_{sp}\)  | ATE  |
| True Values         | 0.8000      | 1.1000      | 0.2000      | 0.1066 | True Values | 0.8000      | 1.1000      | 0.2000      | 0.1066 | True Values | 0.8000      | 1.1000      | 0.2000      | 0.1066 | True Values | 0.8000      | 1.1000      | 0.2000      | 0.1066 |
| Estimate            | 0.8038      | 0.9013      | 0.2088      | 0.2108 | Estimate   | 0.8544      | 1.2755      | 0.1821      | 0.1256 | Estimate   | 0.8391      | 1.2759      | 0.1854      | 0.1218 | Estimate   | 0.8459      | 1.2556      | 0.1878      | 0.1172 |
| S.D                 | 0.1308      | 0.5666      | 0.1823      | 0.1137 | S.D        | 0.1166      | 0.3865      | 0.1271      | 0.0638 | S.D        | 0.1185      | 0.3701      | 0.1214      | 0.0573 | S.D        | 0.0459      | 0.1556      | -0.0122     | 0.0105 |
| Bias                | 0.0038      | -0.1987     | 0.0088      | 0.1041 | Bias       | 0.0544      | 0.1755      | -0.0179     | 0.0190 | Bias       | 0.0495      | 0.1506      | -0.0146     | 0.0152 | Bias       | 0.0459      | 0.1556      | -0.0122     | 0.0105 |
| RMSE                | 0.0171      | 0.3210      | 0.0332      | 0.0238 | RMSE       | 0.0136      | 0.1494      | 0.0161      | 0.0044 | RMSE       | 0.0143      | 0.1288      | 0.0088      | 0.0035 | RMSE       | 0.0143      | 0.1370      | 0.0147      | 0.0034 |

Electronic copy available at: https://ssrn.com/abstract=3346834
|                                   | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|-----------------------------------|----------------------------------------|--------------------------------------------|
| True Values                       | γ          | δ₁       | ρ_sp | ATE     | γ          | δ₁       | ρ_sp | ATE     |
| Estimate                          | 0.8000    | 1.1000   | 0.7000 | 0.3643  | 0.8000    | 1.1000   | 0.7000 | 0.3643  |
| S.D                               | 0.0932    | 0.3663   | 0.1161 | 0.0860  | 0.0942    | 0.3909   | 0.1167 | 0.0928  |
| Bias                              | 0.0032    | 0.0403   | -0.0020 | 0.0016  | 0.0038    | 0.0475   | 0.0060 | -0.0028 |
| RMSE                              | 0.0087    | 0.1342   | 0.0135 | 0.0074  | 0.0089    | 0.1528   | 0.0136 | 0.0086  |

|                                   | Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula |
|-----------------------------------|-----------------------------------|----------------------------------------|
| True Values                       | γ          | δ₁       | ρ_sp | ATE     | γ          | δ₁       | ρ_sp | ATE     |
| Estimate                          | 0.8094    | 1.2165   | 0.6676 | 0.3858  | 0.8097    | 1.2244   | 0.6738 | 0.3783  |
| S.D                               | 0.0928    | 0.3185   | 0.0912 | 0.0675  | 0.0930    | 0.3138   | 0.0856 | 0.0748  |
| Bias                              | 0.0094    | 0.1165   | -0.0324 | 0.0214  | 0.0097    | 0.1244   | -0.0262 | 0.0140  |
| RMSE                              | 0.0090    | 0.1015   | 0.0083 | 0.0050  | 0.0086    | 0.0985   | 0.0073 | 0.0058  |

|                                   | Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|-----------------------------------|--------------------------------------|--------------------------------------------|
| True Values                       | γ          | δ₁       | ρ_sp | ATE     | γ          | δ₁       | ρ_sp | ATE     |
| Estimate                          | 0.8055    | 1.1382   | 0.6952 | 0.3666  | 0.8065    | 1.1581   | 0.7002 | 0.3598  |
| S.D                               | 0.0946    | 0.3188   | 0.0939 | 0.0709  | 0.0946    | 0.3441   | 0.0910 | 0.0750  |
| Bias                              | 0.0055    | 0.0382   | -0.0049 | 0.0023  | 0.0065    | 0.0581   | 0.0002 | -0.0045 |
| RMSE                              | 0.0090    | 0.1017   | 0.0088 | 0.0050  | 0.0090    | 0.1184   | 0.0083 | 0.0057  |

|                                   | Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|-----------------------------------|-------------------------------------|------------------------------------------|
| True Values                       | γ          | δ₁       | ρ_sp | ATE     | γ          | δ₁       | ρ_sp | ATE     |
| Estimate                          | 0.8036    | 1.1517   | 0.6945 | 0.3688  | 0.8055    | 1.1806   | 0.6979 | 0.3702  |
| S.D                               | 0.0937    | 0.3644   | 0.1185 | 0.0841  | 0.0942    | 0.3941   | 0.1197 | 0.0942  |
| Bias                              | 0.0036    | 0.0517   | -0.0055 | 0.0045  | 0.0055    | 0.0806   | -0.0021 | 0.0058  |
| RMSE                              | 0.0088    | 0.1328   | 0.0140 | 0.0071  | 0.0089    | 0.1553   | 0.0143 | 0.0089  |
Table C.16: Misspecification of Marginals \((n = 500, \rho_{sp} = 0.7)\) (True marginal: mixture of normals)

| Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|---------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 | True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 |
| Estimate | 0.7942 | 1.1740 | 0.6323 | 0.2582 | Estimate | 0.8565 | 1.2619 | 0.6932 | 0.1252 |
| S.D | 0.1276 | 0.6180 | 0.1331 | 0.1080 | S.D | 0.1084 | 0.3714 | 0.0835 | 0.0661 |
| Bias | -0.0058 | 0.0740 | -0.0676 | 0.1515 | Bias | 0.0565 | 0.1619 | -0.0068 | 0.0186 |
| RMSE | 0.0163 | 0.3820 | 0.0177 | 0.0346 | RMSE | 0.0118 | 0.1379 | 0.0070 | 0.0047 |

| Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula |
|-------------------------------------|-----------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 | True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 |
| Estimate | 0.8164 | 1.5022 | 0.5823 | 0.3157 | Estimate | 0.8566 | 1.3039 | 0.6787 | 0.1411 |
| S.D | 0.1237 | 0.5841 | 0.1151 | 0.0918 | S.D | 0.1094 | 0.3071 | 0.0619 | 0.0685 |
| Bias | 0.0164 | 0.4022 | -0.1177 | 0.2091 | Bias | 0.0566 | 0.2039 | -0.0212 | 0.0345 |
| RMSE | 0.0153 | 0.3412 | 0.0132 | 0.0521 | RMSE | 0.0120 | 0.0943 | 0.0038 | 0.0059 |

| Parametric Estimation, Clayton Copula | Semiparametric Estimation, Clayton Copula |
|--------------------------------------|-------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 | True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 |
| Estimate | 0.8219 | 1.3357 | 0.6006 | 0.2820 | Estimate | 0.8569 | 1.2553 | 0.6888 | 0.1272 |
| S.D | 0.1297 | 0.5681 | 0.1200 | 0.0902 | S.D | 0.1109 | 0.3197 | 0.0714 | 0.0628 |
| Bias | 0.0219 | 0.2357 | -0.0995 | 0.1754 | Bias | 0.0569 | 0.1553 | -0.0113 | 0.0206 |
| RMSE | 0.0168 | 0.3227 | 0.0144 | 0.0389 | RMSE | 0.0123 | 0.1022 | 0.0051 | 0.0044 |

| Parametric Estimation, Gumbel Copula | Semiparametric Estimation, Gumbel Copula |
|--------------------------------------|------------------------------------------|
| \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) | \(\gamma\) | \(\delta_1\) | \(\rho_{sp}\) | \(ATE\) |
| True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 | True Values | 0.8000 | 1.1000 | 0.7000 | 0.1066 |
| Estimate | 0.7874 | 0.4638 | 0.1139 | 0.0942 | Estimate | 0.8556 | 1.2614 | 0.6933 | 0.1251 |
| S.D | 0.1235 | 0.5168 | 0.1135 | 0.0942 | S.D | 0.1081 | 0.3526 | 0.0810 | 0.0661 |
| Bias | -0.0126 | 0.0463 | -0.0611 | 0.1501 | Bias | 0.0556 | 0.1614 | -0.0047 | 0.0184 |
| RMSE | 0.0152 | 0.2670 | 0.0129 | 0.0314 | RMSE | 0.0117 | 0.1243 | 0.0066 | 0.0047 |
Table C.17: Correct Specification \((n = 500, \rho_{sp} = -0.5)\) (True marginal: normal)

|                      | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|----------------------|---------------------------------------|---------------------------------------------|
| **True Values**      | \begin{array}{cccc}                  | \begin{array}{cccc}                        |
| \gamma & 0.8000 & 1.1000 & -0.5000 & 0.3643 \end{array} | \begin{array}{cccc}                        |
| \delta_1           | 1.1000 | -0.4912 & 0.3560  & \end{array} | \begin{array}{cccc}                        |
| \rho_{sp}           | 0.0970 | 0.4097  | 0.1727   & -0.0871 \end{array} | \begin{array}{cccc}                        |
| ATE                 | 0.0094 | 0.1678  | 0.0298   & 0.0077 \end{array} | \begin{array}{cccc}                        |
| **Bias**            | 0.0090 | 0.0134  | 0.0088   & -0.0084 \end{array} | \begin{array}{cccc}                        |
| **RMSE**            | 0.0970 | 0.4097  | 0.1727   & -0.0871 \end{array} | \begin{array}{cccc}                        |
| **Estimate**        | 0.8090 | 1.1134  | -0.4912  & 0.3560 \end{array} | \begin{array}{cccc}                        |
| **S.D**             | 0.0970 | 0.4097  | 0.1727   & -0.0871 \end{array} | \begin{array}{cccc}                        |
| **Bias**            | 0.0090 | 0.0134  | 0.0088   & -0.0084 \end{array} | \begin{array}{cccc}                        |
| **RMSE**            | 0.0970 | 0.4097  | 0.1727   & -0.0871 \end{array} | \begin{array}{cccc}                        |

Table C.18: Misspecification of Marginals \((n = 500, \rho_{sp} = -0.5)\) (True marginal: mixture of normals)

|                      | Parametric Estimation, Gaussian Copula | Semiparametric Estimation, Gaussian Copula |
|----------------------|---------------------------------------|---------------------------------------------|
| **True Values**      | \begin{array}{cccc}                  | \begin{array}{cccc}                        |
| \gamma & 0.8000 & 1.1000 & -0.5000 & 0.1066 \end{array} | \begin{array}{cccc}                        |
| \delta_1           | 1.1000 | -0.3171 & 0.1028  & \end{array} | \begin{array}{cccc}                        |
| \rho_{sp}           | 0.1301 | 0.4122  | 0.1665  & 0.1052 \end{array} | \begin{array}{cccc}                        |
| ATE                 | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                        |
| **Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                        |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                        |
| **Estimate**        | 0.8385 | 0.3931  | -0.3171 & 0.1028 \end{array} | \begin{array}{cccc}                        |
| **S.D**             | 0.1301 | 0.4122  | 0.1665  & 0.1052 \end{array} | \begin{array}{cccc}                        |
| **Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                        |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                        |

**Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                        |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                        |

**Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                        |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                        |

Table C.18: Misspecification of Marginals \((n = 500, \rho_{sp} = -0.5)\) (True marginal: mixture of normals)

|                      | Parametric Estimation, Frank Copula | Semiparametric Estimation, Frank Copula |
|----------------------|------------------------------------|-----------------------------------------|
| **True Values**      | \begin{array}{cccc}                | \begin{array}{cccc}                      |
| \gamma & 0.8000 & 1.1000 & -0.5000 & 0.1066 \end{array} | \begin{array}{cccc}                      |
| \delta_1           | 1.1000 | -0.3171 & 0.1028  & \end{array} | \begin{array}{cccc}                      |
| \rho_{sp}           | 0.1301 | 0.4122  | 0.1665  & 0.1052 \end{array} | \begin{array}{cccc}                      |
| ATE                 | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                      |
| **Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                      |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                      |
| **Estimate**        | 0.8385 | 0.3931  | -0.3171 & 0.1028 \end{array} | \begin{array}{cccc}                      |
| **S.D**             | 0.1301 | 0.4122  | 0.1665  & 0.1052 \end{array} | \begin{array}{cccc}                      |
| **Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                      |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                      |

**Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                      |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                      |

**Bias**            | 0.0385 | -0.7069 | 0.1829  & -0.0038 \end{array} | \begin{array}{cccc}                      |
| **RMSE**            | 0.0169 | 0.1699  | 0.0277  & 0.0111 \end{array} | \begin{array}{cccc}                      |

39
Table C.19: Coverage Probabilities of Bootstrap Confidence Intervals (Nominal Level = 0.95)

|         | Normal Approximation | Bootstrap Percentile |
|---------|----------------------|----------------------|
| $ATE$   | 0.9050               | 0.9300               |
| $\alpha$ | 0.9700             | 0.9500               |
| $\gamma$ | 0.9600             | 0.9250               |
| $\beta$  | 0.9400             | 0.9300               |
| $\delta$ | 0.8750             | 0.9200               |
| $\rho$   | 0.9000             | 0.9500               |

References

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics 6B*, 5549–5632. B.3, 3, B.4, B.4

Chen, X. and X. Shen (1998). Sieve extremum estimates for weakly dependent data. *Econometrica 66*(2), 289–314. B.4

Freyberger, J. and M. Masten (2015). Compactness of infinite dimensional parameter spaces. Technical report, cemmap working paper, Centre for Microdata Methods and Practice. B.3

Han, S. and S. Lee (2018). Estimation in a generalization of bivariate probit models with dummy endogenous regressors. *Available at SSRN 3233422* (document)

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis. A.2

Lorentz, G. (1966). *Approximation of functions*. Holt, Rinehart and Winston New York. B.4

Nelsen, R. B. (1999). *An introduction to copulas*. Springer Verlag. B.3

van de Geer, S. A. (2000). *Empirical Processes in M-estimation*, Volume 6. Cambridge university press. B.3, B.4

van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes*. Springer, New York. B.4