Appendix of "*Semiparametric Estimation and Variable Selection for Single-Index Copula Models*"

Bingduo Yang[*] Christian M. Hafner[†] Guannan Liu[‡] Wei Long[§]

This is the technical appendix of the paper entitled "Semiparametric Estimation and Variable Selection for Single-Index Copula Models".

Appendix

A: Asymptotic Properties for Unpenalized Estimators

To study the asymptotic properties of the unpenalized estimators, we introduce the following assumptions:

(A1) The copula is Lipschitz continuous in its dependence parameter θ. The function $\theta(\cdot)$ is continuous, bounded, not constant everywhere and has second order continuous derivatives on A_Λ with A_Λ being the domain of Λ.

(A2) There exists no perfect multicollinearity within the components of W, and none of the components of W is constant.

(A3) The first element of γ is positive and $\|\gamma\| = 1$, where $\|\cdot\|$ is the Euclidean norm (L_2 norm).

(A4) For any $\gamma \in A_\gamma$ and $\Lambda \in A_\Lambda$, the density function $f(\Lambda)$ is continuous and there exists $\epsilon > 0$ such that $f(\Lambda) \geq \epsilon$.

(A5) The copula likelihood function $\ell_{ct}(\psi, \gamma, \theta)$ has bounded third derivative with respect to θ and bounded second derivatives with respect to ψ. The marginal likelihood function $\ell_{mt}(\psi)$ has bounded second derivative with respect to ψ.

[*]Lingnan (University) College, Sun Yat-sen University, Guangzhou, Guangdong 510275, China.

[†]Louvain Institute for Data Analysis and Modelling in Economics and Statistics, Université catholique de Louvain, Voie du Roman Pays, 20, 1348 Louvain-la-Neuve, Belgium.

[‡]Corresponding author (email: gliuecon@gmail.com). MOE Key Laboratory of Econometrics, School of Economics and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China.

[§]Department of Economics, Tulane University, New Orleans, LA 70118, United States.

(A6) The kernel function $k(z)$ is twice continuously differentiable on its support, and its second derivative satisfies a Lipschitz condition. Define the kernel constants $\mu_2 = \int z^2 k(z)dz < \infty$ and $\nu_0 = \int k^2(z)dz < \infty$.

(A7) The bandwidth h satisfies $h \to 0$ and $Th \to \infty$, as $T \to \infty$.

(A8) Assume that $\{X_t, Z_t\}_{t=1}^T$ is a strictly stationary α-mixing sequence. There exists some constant $c > 2$ such that $E\|X_t\|^c < \infty$, $E\|Z_t\|^c < \infty$ and $E\|\ell'_{ct}(\theta_0(\Lambda))\|^c < \infty$, and the mixing coefficient $\alpha(\ell)$ satisfying $\sum_{l\geq 1} \ell^{c_1}\alpha(\ell)^{1-2/c} < \infty$ for some $c_1 > 1 - 2/c$. Further, assume that there exists a sequence of positive integers satisfying $\ell_T \to \infty$ and $\ell_T = o_p((Th)^{1/2})$ such that $(T/h)^{1/2}\alpha(\ell_T) \to 0$ with h being the bandwidth, as $T \to \infty$.

(A9) $\lim_{T\to\infty}\{\sqrt{T}p'_{\lambda_T}(|\gamma_k|)\} = 0$ and $p''_{\lambda_T}(|\gamma_k|) \to 0$ for $k = 1, \dots, d_1$, and $\lim_{T\to\infty}\{\sqrt{T}\inf_{|\gamma_k|\leq C/\sqrt{T}} p'_{\lambda_T}(|\gamma_k|)\}$ $\to \infty$ for $k = d_1 + 1, \dots, d$, and for any $C > 0$.

Many commonly used copulas, such as the Gaussian, Clayton, and Gumbel copula, satisfy Assumption (A1). Assumptions (A2)-(A3) are mild conditions for identification. It is obvious that γ cannot be identified if θ is a constant. The no perfect multicollinearity condition in Assumption (A2) is similar to that in classical linear models. A constant is excluded from W as it can be absorbed by the nonparametric function θ. As γ is identified up to sign and scale, Assumption (A3) imposes sign and scale restrictions for identification. Assumption (A4) is imposed so that nonparametric estimators are well defined. Assumption (A5) is for deriving the asymptotic distribution. Assumptions (A6) and (A7) are common in nonparametric estimation. In our simulation and empirical study, the commonly adopted Epanechnikov kernel function $k(u) = 3/4(1-u^2)I(|u| \leq 1)$ is used, where $I(|u| \leq 1)$ takes the value 1 if $|u| \leq 1$ and 0 otherwise. Assumption (A8) is for weakly dependent data, which can also be found in Section 6.6.2 in the book of Fan and Yao (2005). Many time series processess, such as ARMA and GARCH which are widely used in finance and econometrics, satisfy the α-mixing conditions under some mild conditions (e.g., Cai, 2002; Basrak, Davis and Mikosch, 2002). Assumption (A9) holds when the tuning parameter $\lambda_T \to 0$ and $\sqrt{T}\lambda_T \to \infty$ as $T \to \infty$, which are commonly employed in SCAD-based variable selection; see Fan and Li (2001) for details.

We define $\tilde{\xi} = (\hat{\psi}^\intercal, \tilde{\gamma}^\intercal)^\intercal$ with $\tilde{\gamma}$ being the unpenalized estimator of γ. The asymptotic properties for $\tilde{\xi}$ is summarized in the following theorem.

Theorem A.1 *Let $\{X_t, Z_t\}_{t=1}^T$ be a strictly stationary α-mixing sequence following the index* *copula model in (??). Under Assumptions (A1)-(A9), as* $T \to \infty$, $\|\tilde{\xi} - \xi_0\| = O_p(T^{-1/2})$ *and*

$$\sqrt{T}(\tilde{\xi} - \xi_0) \xrightarrow{d} N(0, V),$$

where $V = M^{-1}\Omega(M^{-1})^\intercal$ *with* $M = -\frac{1}{T}E\,\partial\Pi(\theta_0, m_w, \xi_0)/\partial\xi$ *and* $\Omega = \sum_{j=-\infty}^{\infty}\Gamma_j$ *with* $\Gamma_j = \mathrm{Cov}\,(\zeta_t, \zeta_{t-j})$ *and* $\zeta_t = (\ell'_{mt}(\psi_0)^\intercal, \pi_t(\theta_0, m_w, \xi_0)^\intercal)^\intercal$.

If the random vector sequence $\{\zeta_t\}_{t=1}^{\infty}$ is either i.i.d. or a martingale difference sequence, then the long-run variance Ω simplifies to $\Omega = \Gamma(0) = \mathrm{Var}(\zeta_t)$. Otherwise, the autocovariance function $\Gamma(j)$ may not be zero at least for some lag $j \neq 0$ due to the serial correlation of ζ_t.

Remark 1. *To find the asymptotic distribution of the unpenalized estimator* $\tilde{\gamma}$ *in the index copula model, we define* $\iota = (\boldsymbol{0}, I_d)$, *where* $\boldsymbol{0}$ *is a* $d \times d_m$ *matrix of zeros, and* I_d *is the d-dimensional identity matrix, where* d_m *and* d *are the dimensions of* ψ *and* γ, *respectively. Then, as* $T \to \infty$, $\sqrt{T}(\tilde{\gamma} - \gamma_0) \xrightarrow{d} N(0, V_\gamma)$ *where* $V_\gamma = \iota V \iota^\intercal$.

B: Mathematical Proofs

In this subsection, we prove the main results of theorems in Appendix A and Section 2.

Proof of Theorem A.1: The proof for Theorem A.1 is similar to the proof of Theorem 2 without the penalty term, so we omit it here.

□

Lemma 1. *For any given constant* C,

$$\underset{\|\xi-\xi_0\|\le CT^{-1/2}}{sup} \left\| T^{-1/2}\Pi(\hat{\theta}, \hat{m}_w, \xi) - T^{-1/2}\Pi(\theta_0, m_w, \xi_0) + T^{1/2}M(\xi - \xi_0) \right\| = o_p(1)$$

where $M = -\frac{1}{T}E\,\partial\Pi(\theta_0, m_w, \xi_0)/\partial\xi$.

Proof: We have

$$\begin{aligned}
& T^{-1/2}\Pi(\hat{\theta}, \hat{m}_w, \xi) - T^{-1/2}\Pi(\theta_0, m_w, \xi_0) \\
= \quad & T^{-1/2}\Pi(\hat{\theta}, \hat{m}_w, \xi) - T^{-1/2}\Pi(\hat{\theta}, m_w, \xi) + T^{-1/2}\Pi(\hat{\theta}, m_w, \xi) - T^{-1/2}\Pi(\theta_0, m_w, \xi) \\
& + T^{-1/2}\Pi(\theta_0, m_w, \xi) - T^{-1/2}\Pi(\theta_0, m_w, \xi_0) \\
\doteq \quad & A_1 + A_2 + A_3
\end{aligned}$$

where

$$
\begin{aligned}
A_1 = \quad & T^{-1/2}\Pi(\hat{\theta}, \hat{m}_w, \xi) - T^{-1/2}\Pi(\hat{\theta}, m_w, \xi) \\
= \quad & \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \ell'_{mt} \\ \ell'_{ct}(\psi, \gamma, \hat{\theta})\hat{\theta}'(\Lambda_t)(W_t - \hat{m}_w(\Lambda_t)) \end{pmatrix} - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \ell'_{mt} \\ \ell'_{ct}(\psi, \gamma, \hat{\theta})\hat{\theta}'(\Lambda_t)(W_t - m_w(\Lambda_t)) \end{pmatrix} \\
= \quad & \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} 0 \\ \ell'_{ct}(\psi, \gamma, \hat{\theta})\hat{\theta}'(\Lambda_t)(m_w(\Lambda_t) - \hat{m}_w(\Lambda_t)) \end{pmatrix}.
\end{aligned}
$$

Define $R_t = \ell'_{ct}(\psi, \gamma, \hat{\theta})\hat{\theta}'(\Lambda_t)(m_w(\Lambda_t) - \hat{m}_w(\Lambda_t))$. Note that

$$
\mathrm{Var}(R_1) = E\{\ell'^2_{c1}(\psi, \gamma, \hat{\theta})\hat{\theta}'^2(\Lambda_1)(m_w(\Lambda_1) - \hat{m}_w(\Lambda_1))^2\} = O_p(\frac{1}{Th})
$$

By stationarity, we have

$$
\mathrm{Var}(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} R_t) = \mathrm{Var}(R_1) + \sum_{s=1}^{T-1}(1 - s/T)cov(R_1, R_{s+1}).
$$

Following the same technology on pages 251-252 of Fan and Gijbels (1996), we can show $\sum_{s=1}^{T} |cov(R_1, R_{s+1})| = o_p(\frac{1}{Th})$. It follows that

$$
\mathrm{Var}(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} R_t) = o_p(1) \quad \text{and} \quad A_1 = o_p(1).
$$

Next,

$$
\begin{aligned}
A_2 = \quad & T^{-1/2}\Pi(\hat{\theta}, m_w, \xi) - T^{-1/2}\Pi(\theta_0, m_w, \xi) \\
= \quad & \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} 0 \\ \left[\ell'_{ct}(\psi, \gamma, \hat{\theta})\hat{\theta}'(\Lambda_t) - \ell'_{ct}(\psi, \gamma, \theta_0)\theta'_0(\Lambda_t)\right](W_t - m_w(\Lambda_t)) \end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
& T^{-1/2} \sum_{t=1}^{T} \left[\ell'_{ct}(\psi, \gamma, \hat{\theta})\hat{\theta}'(\Lambda_t) - \ell'_{ct}(\psi, \gamma, \theta_0)\theta'_0(\Lambda_t)\right](W_t - m_w(\Lambda_t)) \\
= \quad & T^{-1/2} \sum_{t=1}^{T} \left[\ell'_{ct}(\psi, \gamma, \hat{\theta})\hat{\theta}'(\Lambda_t) - \ell'_{ct}(\psi, \gamma, \hat{\theta})\theta'_0(\Lambda_t)\right](W_t - m_w(\Lambda_t)) \\
& + T^{-1/2} \sum_{t=1}^{T} \left[\ell'_{ct}(\psi, \gamma, \hat{\theta})\theta'_0(\Lambda_t) - \ell'_{ct}(\psi, \gamma, \theta_0)\theta'_0(\Lambda_t)\right](W_t - m_w(\Lambda_t)) \\
\doteq \quad & A_{21} + A_{22}
\end{aligned}
$$

By the same precedures for proving A_1, we can show

$$A_{21} = T^{-1/2}\sum_{t=1}^{T}\left[\ell'_{ct}(\psi,\gamma,\hat{\theta})\hat{\theta}'(\Lambda_t) - \ell'_{ct}(\psi,\gamma,\hat{\theta})\theta'_0(\Lambda_t)\right](W_t - m_w(\Lambda_t)) = o_p(1),$$

and

$$\begin{aligned} A_{22} = & \quad T^{-1/2}\sum_{t=1}^{T}\left[\ell'_{ct}(\psi,\gamma,\hat{\theta})\theta'_0(\Lambda_t) - \ell'_{ct}(\psi,\gamma,\theta_0)\theta'_0(\Lambda_t)\right](W_t - m_w(\Lambda_t)) \\ = & \quad T^{-1/2}\sum_{t=1}^{T}\left[\{\ell''_{ct}(\psi,\gamma,\theta_0)(\hat{\theta}-\theta_0)\}\{1+o_p(1)\}\right]\theta'_0(\Lambda_t)(W_t - m_w(\Lambda_t)) \\ = & \quad o_p(1). \end{aligned}$$

This implies that A_2 is of order $o_p(1)$. Finally,

$$\begin{aligned} A_3 = & \quad T^{-1/2}\Pi(\theta_0, m_w, \xi) - T^{-1/2}\Pi(\theta_0, m_w, \xi_0) \\ = & \quad T^{-1/2}\{(\partial\Pi(\theta_0, m_w, \xi_0)/\partial\xi)(\xi - \xi_0)\}\{1+o_p(1)\} \\ = & \quad \sqrt{T}(-M)(\xi-\xi_0) + o_p(1). \end{aligned}$$

Therefore, $T^{-1/2}\Pi(\hat{\theta}, \hat{m}_w, \xi) - T^{-1/2}\Pi(\theta_0, m_w, \xi_0) + T^{1/2}M(\xi-\xi_0) = A_1 + A_2 + A_3 + T^{1/2}M(\xi-\xi_0) = o_p(1)$, which implies the stated result.

□

Proof of Theorem 1:

Let $\xi = \xi_0 + T^{-1/2}\mathbf{v}$ with $\|\mathbf{v}\| = C$. For any small constant $\epsilon > 0$, if we can show there exists a large constant C such that

$$P\left\{\min_{\|\xi-\xi_0\|=T^{-1/2}C} \sqrt{T}(\xi-\xi_0)^\intercal(-M^\intercal)\frac{1}{\sqrt{T}}\Pi^P(\hat{\theta}, \hat{m}_w, \xi) > 0\right\} > 1-\epsilon,$$

where $\Pi^P(\hat{\theta}, \hat{m}_w, \xi) = \Pi(\hat{\theta}, \hat{m}_w, \xi) - T\mathbf{p}'_\lambda(|\xi|)\text{sgn}(\xi)$, then we can choose a $\sqrt{T}$-consistent estimator $\hat{\xi}$ satisfying both $||\hat{\xi}-\xi_0|| = O_p(1/\sqrt{T})$ and $\Pi^P(\hat{\theta}, \hat{m}_w, \hat{\xi}) = 0$.

Lemma 1 implies

$$\begin{aligned} & \sqrt{T}(\xi-\xi_0)^\intercal(-M^\intercal)\frac{1}{\sqrt{T}}\Pi^P(\hat{\theta}, \hat{m}_w, \xi) \\ = & \quad \sqrt{T}(\xi-\xi_0)^\intercal(-M^\intercal)\left[\frac{1}{\sqrt{T}}\Pi(\theta_0, m_w, \xi_0) - \sqrt{T}M(\xi-\xi_0) - \sqrt{T}\mathbf{p}'_\lambda(|\xi|)\text{sgn}(\xi) + o_p(1)\right] \\ \geq & \quad \sqrt{T}(\xi-\xi_0)^\intercal(-M^\intercal)\frac{1}{\sqrt{T}}\Pi(\theta_0, m_w, \xi_0) + \sqrt{T}(\xi-\xi_0)^\intercal M^\intercal M\sqrt{T}(\xi-\xi_0) \end{aligned}$$

$$
\begin{aligned}
& +\sqrt{T}(\xi-\xi_0)^\intercal M_2 \sqrt{T}\mathbf{p}'_\lambda(|\gamma_1|)\mathrm{sgn}(\gamma_1) \\
= \quad & \sqrt{T}(\xi-\xi_0)^\intercal(-M^\intercal)\frac{1}{\sqrt{T}}\Pi(\theta_0, m_w, \xi_0) + \sqrt{T}(\xi-\xi_0)^\intercal M^\intercal M\sqrt{T}(\xi-\xi_0) \\
& +\sqrt{T}(\xi-\xi_0)^\intercal M_2\{\sqrt{T}\mathbf{p}'_\lambda(|\gamma_{10}|)\mathrm{sgn}(\gamma_{10}) + \sqrt{T}\mathbf{p}''_\lambda(|\gamma_{10}|)(\gamma_1-\gamma_{10})\}\{1+o_p(1)\} \qquad (1)
\end{aligned}
$$

where M_2 is a submatrix of the partition $M^\intercal = (M_1, M_2, M_3)$ with M_1, M_2 and M_3 being $(d_m+d)\times d_m$, $(d_m+d)\times d_1$ and $(d_m+d)\times(d-d_1)$ matrices, respectively.

The first term on the right hand side of the last inequality in (1) is of order $C * O_p(1)$ and the second term is of order $C^2 * O_p(1)$. Using Assumption (A9), $\sqrt{T}\mathbf{p}'_\lambda(|\gamma_{10}|)\to 0$ and $\mathbf{p}''_\lambda(|\gamma_{10}|)\to 0$. By choosing the constant C sufficiently large, the second term will dominate the other two terms. This completes the proof.

□

Proof of Theorem 2: We first show the sparsity with $\hat\gamma_2 = 0$. Suppose a $\sqrt{T}$-consistent estimator $\hat\xi^* = (\hat\psi^\intercal, \hat\gamma_1^\intercal, \hat\gamma_2^\intercal)^\intercal$ with $\hat\gamma_2 \neq 0$ such that $\Pi^P(\hat\theta, \hat m_w, \hat\xi^*) = 0$ exists. By Lemma 1,

$$
\frac{1}{\sqrt{T}}\Pi(\theta_0, m_w, \xi_0) - \sqrt{T}M(\hat\xi^*-\xi_0) + o_p(1) = \sqrt{T}\mathbf{p}'(\hat\xi^*)\mathrm{sgn}(\hat\xi^*). \qquad (2)
$$

The first two components on the left hand side of (2) are of order $O_p(1)$. However, the last $d-d_1$ elements of $\sqrt{T}\mathbf{p}'(\hat\xi^*)$ on the right hand side diverge to infinity by Assumption (A9). Therefore, by contradiction, we conclude that $\hat\gamma_2 = 0$ must hold.

Second, we show asymptotic normality. By Lemma 1, we have

$$
\frac{1}{\sqrt{T}}\Pi(\theta_0, m_{w_1}, \xi_{10}) - \sqrt{T}M_1(\hat\xi_1-\xi_{10}) + o_p(1) = \sqrt{T}\mathbf{p}'(\hat\xi_1)\mathrm{sgn}(\hat\xi_1)
$$

where $\hat\xi_1 = (\hat\psi^\intercal, \hat\gamma_1^\intercal)^\intercal$. The term $\sqrt{T}\mathbf{p}'(\hat\xi_1) = 0$ as $T\to\infty$ according to the conditions in (A9). It follows that

$$
\sqrt{T}M_1(\hat\xi_1-\xi_{10}) = \frac{1}{\sqrt{T}}\Pi(\theta_0, m_{w_1}, \xi_{10}) + o_p(1).
$$

Further, the condition (i) of Theorem 2.21 in Fan and Yao (2005) holds by Assumption (A8), and the asymptotic normality can be obtained by Theorem 2.21 in Fan and Yao (2005).

□

For the proof of Theorem 3 we need the following lemma.

Lemma 2. *Assume that the parametric estimators $\hat\psi$ and $\hat\gamma$ and the local constant estimator $\hat\theta$ are obtained from the three-step procedure of Section 2.1 and satisfy* $\|\hat\psi-\psi_0\| = O_p(1/\sqrt{T})$,

$\|\hat{\gamma}-\gamma_0\| = O_p(1/\sqrt{T})$ *and* $\|\hat{\theta}-\theta_0\| = O_p(1/\sqrt{Th})$. *Define the local log-likelihood function as*

$$L_h(\psi,\gamma,\theta) = \sum_{t=1}^{T} \ell_{ct}(\psi,\gamma,\theta) k_h(\gamma^{\intercal} W_t - \Lambda),$$

where $\ell_{ct}(\psi,\gamma,\theta) = \log c(u_t;\theta(\gamma^{\intercal} W_t))$. *Under Assumptions (A1)-(A9), we have*

$$L_h(\hat{\psi},\hat{\gamma},\hat{\theta}) - L_h(\psi_0,\gamma_0,\theta_0) = L_h(\psi_0,\gamma_0,\hat{\theta}) - L_h(\psi_0,\gamma_0,\theta_0) + o_p(1/h).$$

Proof: Let

$$\begin{aligned}
& L_h(\hat{\psi},\hat{\gamma},\hat{\theta}) - L_h(\psi_0,\gamma_0,\theta_0) \\
= \quad & \underbrace{L_h(\hat{\psi},\hat{\gamma},\hat{\theta}) - L_h(\hat{\psi},\gamma_0,\hat{\theta})}_{I_1} + \underbrace{L_h(\hat{\psi},\gamma_0,\hat{\theta}) - L_h(\psi_0,\gamma_0,\hat{\theta})}_{I_2} + \underbrace{L_h(\psi_0,\gamma_0,\hat{\theta}) - L_h(\psi_0,\gamma_0,\theta_0)}_{I_3}.
\end{aligned}$$

By Taylor expansion and the conditions $\|\hat{\psi}-\psi_0\| = O_p(1/\sqrt{T})$, $\|\hat{\gamma}-\gamma_0\| = O_p(1/\sqrt{T})$ and $\|\hat{\theta}-\theta_0\| = O_p(1/\sqrt{Th})$, the first term I_1 is given by

$$\begin{aligned}
I_1 = \quad & L_h(\hat{\psi},\hat{\gamma},\hat{\theta}(\hat{\psi},\hat{\gamma})) - L_h(\hat{\psi},\gamma_0,\hat{\theta}(\hat{\psi},\gamma_0)) \\
= \quad & \left[\frac{1}{\sqrt{T}} \frac{\partial L_h(\hat{\psi},\gamma_0,\hat{\theta}(\hat{\psi},\gamma_0))}{\partial \gamma}\right] \sqrt{T}(\hat{\gamma}-\gamma_0)\{1+o_p(1)\} \\
= \quad & \left[\frac{1}{\sqrt{T}} \frac{\partial L_h(\psi_0,\gamma_0,\hat{\theta}(\psi_0,\gamma_0))}{\partial \gamma}\{1+o_p(1)\}\right] \sqrt{T}(\hat{\gamma}-\gamma_0)\{1+o_p(1)\} \\
= \quad & \left[\frac{1}{\sqrt{T}} \frac{\partial L_h(\psi_0,\gamma_0,\theta_0(\psi_0,\gamma_0))}{\partial \gamma}\{1+o_p(1)\}\{1+o_p(1)\}\right] \sqrt{T}(\hat{\gamma}-\gamma_0)\{1+o_p(1)\}
\end{aligned}$$

is of order $O_p(1)$. In the same vein, we can show that the second term,

$$\begin{aligned}
I_2 = \quad & L_h(\hat{\psi},\gamma_0,\hat{\theta}(\hat{\psi},\gamma_0)) - L_h(\psi_0,\gamma_0,\hat{\theta}(\psi_0,\gamma_0)) \\
= \quad & \left[\frac{1}{\sqrt{T}} \frac{\partial L_h(\psi_0,\gamma_0,\hat{\theta}(\psi_0,\gamma_0))}{\partial \psi}\right] \sqrt{T}(\hat{\psi}-\psi_0)\{1+o_p(1)\} \\
= \quad & \left[\frac{1}{\sqrt{T}} \frac{\partial L_h(\psi_0,\gamma_0,\theta_0(\psi_0,\gamma_0))}{\partial \psi}\{1+o_p(1)\}\right] \sqrt{T}(\hat{\psi}-\psi_0)\{1+o_p(1)\}
\end{aligned}$$

is of order $O_p(1)$. The term on the right hand side $T^{-1/2}\partial L_h(\psi_0,\gamma_0,\theta(\psi_0,\gamma_0))/\partial\psi$ is of order $O_p(1)$ since the first order derivative of the marginal likelihood $T^{-1/2}\partial L_m(\psi_0)/\partial\psi$ and the first order derivative of the full likelihood $T^{-1/2}\partial L_m(\psi_0)/\partial\psi + T^{-1/2}\partial L_h(\psi_0,\gamma_0,\theta(\psi_0,\gamma_0))/\partial\psi$ are of order $O_p(1)$. This implies that $L_h(\hat{\psi},\gamma_0,\hat{\theta}) - L_h(\psi_0,\gamma_0,\hat{\theta})$ is of order $O_p(1)$.

Furthermore, by Taylor expansion and the condition $\|\hat{\theta}-\theta_0\| = O_p(1/\sqrt{Th})$, the last term

$$I_3 = \frac{1}{h}\left[\sqrt{\frac{h}{T}}\frac{\partial L_h(\psi_0,\gamma_0,\theta_0)}{\partial\theta}\right]\sqrt{Th}(\hat{\theta}-\theta_0)\{1+o_p(1)\}$$

is of order $O_p(1/h)$ and dominates the other two terms. This completes the proof. □

Lemma 2 suggests that we can derive the asymptotic distribution of $\hat{\theta}$ without considering the errors from the parametric estimation. The estimators of $(\hat{\psi},\hat{\gamma})$ have little effect on the estimation of $\hat{\theta}$ if the sample size T is large. This result is in line with the fact that the convergence rate of the parametric part of the model is faster than that of the nonparametric component.

Proof of Theorem 3: Using Lemma 2 we can assume that ψ_0 and γ_0 are known for simplicity. Define the kernel constants $\mu_2 = \int z^2k(z)dz$, $\nu_0 = \int k^2(z)dz$ and $\nu_2 = \int z^2k^2(z)dz$. Let $\Lambda_t = \gamma_0^\intercal W_t$, $\ell_{ct}(\theta(\Lambda)) = \ell_{ct}(\psi_0,\gamma_0,\theta)$, $L(\theta(\Lambda)) = \frac{1}{T}\sum_{t=1}^T \ell_{ct}(\theta(\Lambda))k_h(\Lambda_t-\Lambda)$, $L'(\theta(\Lambda)) = \frac{1}{T}\sum_{t=1}^T \ell'_{ct}(\theta(\Lambda))k_h(\Lambda_t-\Lambda)$ and $L''(\theta(\Lambda)) = \frac{1}{T}\sum_{t=1}^T \ell''_{ct}(\theta(\Lambda))k_h(\Lambda_t-\Lambda)$. For a fixed point Λ lying in the interior of the support A_Λ, the normal equation for the local likelihood-based estimator is given by $L'(\hat{\theta}(\Lambda)) = 0$. By a Taylor expansion, it can be written as

$$L'(\theta_0(\Lambda)) + L''(\theta_0(\Lambda))(\hat{\theta}(\Lambda)-\theta_0(\Lambda)) + o_p(1/\sqrt{Th}) = 0,$$

which leads to

$$\hat{\theta}(\Lambda)-\theta_0(\Lambda) = -[L''(\theta_0(\Lambda))]^{-1}L'(\theta_0(\Lambda)) + o_p(1/\sqrt{Th}).$$

By the moment condition, we have

$$\begin{aligned} 0 &= E\{\ell'_{ct}(\theta_0(\Lambda_t))|\Lambda_t=\Lambda\} \\ &= E\{\ell'_{ct}(\theta_0(\Lambda)+r_t)|\Lambda_t=\Lambda\} \\ &\approx E\{\ell'_{ct}(\theta_0(\Lambda))|\Lambda_t=\Lambda\} + r_tE\{\ell''_{ct}(\theta_0(\Lambda))|\Lambda_t=\Lambda\} + o_p(r_t), \end{aligned}$$

where $r_t = \theta_0'(\Lambda)(\Lambda_t-\Lambda)+\frac{1}{2}\theta_0''(\Lambda)(\Lambda_t-\Lambda)^2+o_p(\Lambda_t-\Lambda)^2$. By construction, we have $E\{\ell'_{ct}(\theta_0(\Lambda))|\Lambda_t = \Lambda\} \approx -r_tE\{\ell''_{ct}(\theta_0(\Lambda))|\Lambda_t=\Lambda\} + o_p(r_t)$. Thus,

$$\begin{aligned} E\{L'(\theta_0(\Lambda))|\Lambda_t=\Lambda\} &= -\frac{1}{T}\sum_{t=1}^T r_tE\{\ell''_{ct}(\theta_0(\Lambda))|\Lambda_t=\Lambda\}k_h(\Lambda_t-\Lambda) \\ &= \frac{1}{T}\Sigma(\Lambda)\sum_{t=1}^T r_tk_h(\Lambda_t-\Lambda) \end{aligned}$$

where $\Sigma(\Lambda) = -E\{\ell''_{ct}(\theta_0(\Lambda))|\Lambda_t = \Lambda\}$. Note that

$$\begin{aligned} E\{L''(\theta_0(\Lambda))|\Lambda_t = \Lambda\} \quad &= \frac{1}{T}\sum_{t=1}^{T} E\{\ell''_{ct}(\theta_0(\Lambda))|\Lambda_t = \Lambda\}k_h(\Lambda_t - \Lambda) \\ &= -f(\Lambda)\Sigma(\Lambda) + o_p(1). \end{aligned}$$

It follows by a Taylor expansion and the Riemann sum approximation of an integral that the bias term of $\hat{\theta}(\Lambda)$ can be expressed as

$$\begin{aligned} &\mathrm{E}\{\hat{\theta}(\Lambda)|\Lambda_t = \Lambda\} - \theta_0(\Lambda) \\ &= -[\mathrm{E}\{L''(\theta_0(\Lambda))|\Lambda_t = \Lambda\}]^{-1}\mathrm{E}\{L'(\theta_0(\Lambda))|\Lambda_t = \Lambda\} \\ &\approx \frac{1}{f(\Lambda)}\frac{1}{T}\sum_{t=1}^{T}\left[\theta'_0(\Lambda)(\Lambda_t - \Lambda) + \frac{1}{2}\theta''_0(\Lambda)(\Lambda_t - \Lambda)^2\right]k_h(\Lambda_t - \Lambda) \\ &\approx \frac{1}{f(\Lambda)}\int \theta'_0(\Lambda)(\Lambda_t - \Lambda)f(\Lambda_t)k_h(\Lambda_t - \Lambda)d\Lambda_t + \frac{1}{2f(\Lambda)}\int \theta''_0(\Lambda)(\Lambda_t - \Lambda)^2 f(\Lambda_t)k_h(\Lambda_t - \Lambda)d\Lambda_t \\ &= \frac{h}{f(\Lambda)}\int \theta'_0(\Lambda)uf(\Lambda + uh)k(u)du + \frac{h^2}{2f(\Lambda)}\int \theta''_0(\Lambda)u^2 f(\Lambda + uh)k(u)du \\ &= \frac{h^2}{f(\Lambda)}\theta'_0(\Lambda)f'(\Lambda)\mu_2 + \frac{h^2}{2f(\Lambda)}\theta''_0(\Lambda)\mu_2 + o_p(h^2) \\ &= h^2 B(\Lambda) + o_p(h^2), \end{aligned}$$

where $B(\Lambda) = \frac{1}{f(\Lambda)}(\theta'_0(\Lambda)f'(\Lambda)\mu_2 + \frac{1}{2}\theta''_0(\Lambda)\mu_2)$.

To find the expression for $\mathrm{Var}\{L'(\theta_0(\Lambda))|\Lambda_t = \Lambda\}$, let $Q = \frac{1}{T}\sum_{t=1}^{T} Q_t$, where $Q_t = \ell'_{ct}(\theta_0(\Lambda))k_h(\Lambda_t - \Lambda)$. Note that $\mathrm{Var}(Q_1) = \frac{\nu_0 f(\Lambda)}{h}\Phi(\Lambda) + o_p\left(\frac{1}{h}\right)$ with $\Phi(\Lambda) = E\{\ell'_{ct}(\theta_0(\Lambda))\ell'_{ct}(\theta_0(\Lambda))^\intercal|\Lambda_t = \Lambda\}$.

By stationarity, we have

$$\mathrm{Var}(Q) = \frac{1}{T}\mathrm{Var}(Q_1) + \frac{1}{T}\sum_{s=1}^{T-1}(1 - s/T)cov(Q_1, Q_{s+1}).$$

Define

$$J_1 = \sum_{s=1}^{d_T - 1}|cov(Q_1, Q_{s+1})| \quad \text{and} \quad J_2 = \sum_{s=d_T}^{T-1}|cov(Q_1, Q_{s+1})|,$$

where d_T satisfies $d_T \to \infty$ and $d_T h \to 0$. Following the same technology on pages 251-252 of Fan and Gijbels (1996), we can show $J_1 = o_p(1/h)$ and $J_2 = o_p(1/h)$. It follows that

$$\sum_{s=1}^{T-1}|cov(Q_1, Q_{s+1})| = o_p(1/h) \quad \text{and} \quad \mathrm{Var}(Q) = \frac{\nu_0 f(\Lambda)}{Th}\Phi(\Lambda).$$

Therefore, the variance term is given by

$$
\begin{aligned}
& \operatorname{Var}\{\hat{\theta}(\Lambda)|\Lambda_t = \Lambda\} \\
= \, & \mathrm{E}\{L''(\theta_0(\Lambda))|\Lambda_t = \Lambda\}^{-1}\operatorname{Var}\{L'(\theta_0(\Lambda))|\Lambda_t = \Lambda\}\mathrm{E}\{L''(\theta_0(\Lambda))|\Lambda_t = \Lambda\}^{-1} \\
= \, & \frac{1}{Thf(\Lambda)}\nu_0\Sigma(\Lambda)^{-1}\Phi(\Lambda)\Sigma(\Lambda)^{-1}.
\end{aligned}
$$

By using Assumption (A8), and the score function with $Q_t = \ell'_{ct}(\theta_0(\Lambda))k_h(\Lambda_t - \Lambda)$, we establish the asymptotic normality for $\hat{\theta}(\Lambda)$ by Doob's small-block and large-block technique, which is similar to the proofs on pages 252-255 of Fan and Gijbels (1996). Details are omitted here. This completes the proof.

$\square$