Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score: Auxiliary Appendix[1]

Joshua D. Angrist
MIT

Guido M. Kuersteiner
Georgetown University

This Revision: November 2009

[1]This paper is a revised version of NBER working paper 10975, December 2004. We thank Christina Romer for sharing the Romers' data, Alberto Abadie, Xiaohong Chen, Jordi Gali, Simon Gilchrist, Yanqin Fan, Stefan Hoderlein, Bernard Salanié and James Stock for helpful discussions and seminar and conference participants at ASSA 2006, Auckland, Berkeley, BU, Columbia, Harvard-MIT, Georgetown, Indiana, Illinois, Konstanz, LAMES 2006, 2004 NSF/NBER Time Series Conference, CIREQ Monteral, Rochester, Rutgers, Stanford, St. Gallen, the Triangle Econometrics Workshop, UCLA, Wisconsin, Zurich and the 2^{nd} conference on evaluation research in Mannheim for helpful comments. Thanks also go to the editor and referees for helpful suggestions and comments. Kuersteiner gratefully acknowledges support from NSF grants SES-0095132 and SES-0523186.

Abstract

This Auxiliary Appendix contains detailed proofs for the results presented in the main part of the paper.

A Introduction

We repeat a number of definitions given in the main paper for ease of reference. We also attach the appendix on implementation which is also part of the main paper for the same reason.

The two identifying restrictions are recalled first:

Condition 1 *The sharp null hypothesis of no causal effects means that* $Y_{t,j}^{\psi\prime}(d') = Y_{t,j}^{\psi}(d)$, $j > 0$ *for all* d, d' *and for all possible policy functions* $\psi, \psi' \in \Psi_t$. *In addition, under the no-effects null hypothesis,* $Y_{t,j}^{\psi}(d) = Y_{t+j}$ *for all* d, ψ, t, j.

Condition 2 *Selection on observables:*

$$Y_{t,1}^{\psi}(d), Y_{t,2}^{\psi}(d), ... \perp D_t | z_t, \text{ for all } d \text{ and } \psi \in \Psi_t.$$

Combining Conditions 1 and 2 produces the key testable conditional independence assumption, written in terms of observable distributions as:

$$Y_{t+1}, ..., Y_{t+j}, ... \perp D_t | z_t. \tag{1}$$

Let $U_t = (y_t, z_t)$ and choose a suitable test function $\phi(.,.) : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{H}$ where $\mathbb{H}$ is either $\mathbb{H} = \mathbb{R}^{\mathcal{M}} \times \mathbb{R}^{\mathcal{M}}$ or $\mathbb{H} = \mathbb{R}$. The corresponding unconditional moment restriction is $E\left[\phi(U_t, v)(\mathbf{1}(D_t = i) - p_i(z_t))\right] = 0$ for $i = 1, ...\mathcal{M}$. To move from population moment conditions to the sample, we start by defining the empirical process

$$V_n(v) = n^{-1/2} \sum_{t=1}^{n} m(y_t, D_t, z_t, \theta_0; v)$$

with

$$m(y_t, D_t, z_t, \theta; v) = \phi(U_t, v)\left[\mathcal{D}_t - p(z_t, \theta)\right].$$

Let

$$H(v) = \int_{-\infty}^{v} \left(\text{diag}\left(p(u_2)\right) - p(u_2)p\left(u_2\right)'\right) dF_u(u) \tag{2}$$

with $p(z_t) = (p_1(z_t), ..., p_{\mathcal{M}}(z_t))'$ and $p_i(z_t) = \Pr(D_t = i | z_t)$. Also, $F_u(u)$ is the cumulative marginal distribution function of U_t. Define the covariance kernel of the limiting process $V(v)$ of $V_n(v)$ as

$$\Gamma(v, \tau) = \lim_{n\to\infty} E\left[V_n(v)V_n(\tau)'\right] = \int \phi(u, v) dH(u)\, \phi(u, \tau)'. \tag{3}$$

The statistic $V_n(v)$ can be used to test the null hypothesis of conditional independence by comparing the value of KS$= \sup_v \|V_n(v)\|$ or

$$\text{VM} = \int \|V_n(v)\|^2 dF_u(v) \tag{4}$$

Let $\hat{V}_n(v)$ denote the empirical process of interest where $p(z_t, \theta)$ is replaced by $p(z_t, \hat{\theta})$ and the estimator $\hat{\theta}$ is assumed to satisfy the following asymptotic linearity property:

$$n^{1/2}\left(\hat{\theta}-\theta_0\right) = n^{-1/2}\sum_{t=1}^{n} l\left(D_t, z_t, \theta_0\right) + o_p(1). \tag{5}$$

Define the function $\bar{m}(v,\theta) = E\left[m(y_t, D_t, z_t, \theta; v)\right]$ and let

$$\dot{m}(v,\theta) = -\frac{\partial \bar{m}(v,\theta)}{\partial \theta'}.$$

The empirical process $\hat{V}_n(v)$ converges to a limiting process $\hat{V}\left(v\right)$ with covariance function

$$\hat{\Gamma}(v,\tau) = \Gamma\left(v,\tau\right) - \dot{m}(v,\theta_0)L(\theta_0)\dot{m}(\tau,\theta_0)', \tag{6}$$

with $L\left(\theta_0\right) = E\left[l\left(D_t, z_t, \theta_0\right) l\left(D_t, z_t, \theta_0\right)'\right]$.

Let $\{A_\lambda\}$ be a family of measurable subsets of $[-\infty,\infty]^k$, indexed by $\lambda \in [-\infty,\infty]$ such that $A_{-\infty} = \varnothing$, $A_\infty = [-\infty,\infty]^k$, $\lambda \leq \lambda' \implies A_\lambda \subset A_{\lambda'}$ and $A_{\lambda'} \backslash A_\lambda \to \varnothing$ as $\lambda' \downarrow \lambda$. Define the projection $\pi_\lambda f(v) = \mathbf{1}\left(v \in A_\lambda\right) f(v)$ and $\pi_\lambda^\perp = 1 - \pi_\lambda$ such that $\pi_\lambda^\perp f(v) = \mathbf{1}\left(v \notin A_\lambda\right) f(v)$. We then define the inner product $\langle f(.), g\left(.\right)\rangle \equiv \int f(u)' dH\left(u\right) g(u)$ and, for

$$\bar{l}(v,\theta) = \left(\operatorname{diag}\left(p(v_2)\right) - p(v_2)p\left(v_2\right)'\right)^{-1}\frac{\partial p(v_2,\theta)}{\partial \theta'},$$

define the matrix

$$C_\lambda = \left\langle \pi_\lambda^\perp \bar{l}(.,\theta), \pi_\lambda^\perp \bar{l}(.,\theta)\right\rangle = \int \pi_\lambda^\perp \bar{l}(u,\theta)' dH(u) \pi_\lambda^\perp \bar{l}(u,\theta).$$

Next, note that $V(v)$ can be represented in terms of a vector of Gaussian processes $b(v)$ with covariance function $H(v \wedge \tau)$ as $V(\phi(.,v)) = V(v) = \int \phi(u,v) db(u)$, and similarly $V(l\left(.,\theta_0\right)) = \int l(u,\theta_0) db(u)$ and $\hat{V}\left(f\right) = V(f(.)) - \left\langle f(.), \bar{l}(.,\theta_0)\right\rangle \Sigma_\theta^{-1} V(\bar{l}(.,\theta_0)')$. Define the transformation T as

$$W(v) \equiv T\hat{V}\left(v\right) = \hat{V}\left(v\right) - \int \left\langle \phi\left(.,v\right)', d\left(\pi_\lambda \bar{l}(.,\theta)\right)\right\rangle C_\lambda^{-1} \hat{V}(\pi_\lambda^\perp \bar{l}(.,\theta)') \tag{7}$$

and

$$\hat{W}_n(v) \equiv T_n V_n\left(v\right) = \hat{V}_n\left(v\right) - \int \left(\int \phi(u,v) d\hat{H}_n(u) d\left(\pi_\lambda \bar{l}(u,\theta)\right)\right) \hat{C}_\lambda^{-1} \hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \tag{8}$$

with $\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') = n^{-1/2}\sum_{s=1}^{n} \pi_\lambda^\perp \bar{l}(U_s,\hat{\theta})' \left(\mathcal{D}_s - p(z_s,\hat{\theta})\right)$ and the empirical distribution $\hat{H}_n(v)$ is defined in Appendix C. Now specializing A_λ to

$$A_\lambda = [-\infty,\lambda] \times [-\infty,\infty]^{k-1}, \tag{9}$$

leads to test statistics with simple closed form expressions. Denote the first element of y_t by y_{1t}. Then (8) can be expressed more explicitly as

$$\hat{W}_n(v) = \hat{V}_n(v) - n^{-1/2}\sum_{t=1}^{n}\left[\phi\left(U_t,v\right)\frac{\partial p(z_t,\hat{\theta})}{\partial \theta'}\hat{C}_{y_{1t}}^{-1} n^{-1}\sum_{s=1}^{n}\mathbf{1}\left\{y_{1s} > y_{1t}\right\}\bar{l}(U_s,\hat{\theta})'\left(\mathcal{D}_s - p(z_s,\hat{\theta})\right)\right] \tag{10}$$

To describe the Rosenblatt transform let $U_t = [U_{t1}, ..., U_{tk}]$ and define the transformation $w = T_R(v)$ component wise by $w_1 = F_1(v_1) = \Pr(U_{t1} \le v_1)$, $w_2 = F_2(v_2|v_1) = \Pr(U_{t2} \le v_2|U_{1t} = v_1), ..., w_k = F_k(v_k|v_{k-1}, ..., v_1)$ where $F_k(v_k|v_{k-1}, ..., v_1) = \Pr(U_{tk} \le v_k|U_{tk-1} = v_{k-1}, ..., U_{t1} = v_1)$. The inverse $v = T_R^{-1}(w)$ of this transformation is obtained recursively as $v_1 = F_1^{-1}(u_1)$,

$$v_2 = F_2^{-1}\left(w_2|F_1^{-1}(w_1)\right), ...$$

Using the Rosenblatt transformation we define

$$m_w(w_t, D_t, \theta|v) = \phi(w_t, w)\left[\mathcal{D}_t - p\left(\left[T_R^{-1}(w_t)\right]_z, \theta\right)\right]$$

where $w = T_R(v)$ and $z_t = \left[T_R^{-1}(w_t)\right]_z$ denotes the components of T_R^{-1} corresponding to z_t. The test statistic $V_n(v)$ becomes the marked process

$$V_{w,n}(w) = n^{-1/2}\textstyle\sum_{t=1}^n m_w(w_t, D_t, \theta|w).$$

We denote by $V_w(w)$ the limit of $V_{w,n}(w)$ and by $\hat{V}_w(w)$ the limit of $\hat{V}_{w,n}(w)$ which is the process obtained by replacing θ with $\hat{\theta}$ in $V_{w,n}(w)$. Define the transform $T_w\hat{V}_w(w)$ as before by

$$W_w(w) \equiv T_w\hat{V}_w(w) = \hat{V}_w(w) - \int \left\langle \phi(., w)', d\pi_\lambda \bar{l}_w(., \theta)\right\rangle C_\lambda^{-1}\hat{V}_w(\pi_\lambda^\perp \bar{l}_w(., \theta)'). \tag{11}$$

Finally, to convert $W_w(w)$ to a process which is asymptotically distribution free let

$$h_w(.) = \left(\text{diag}\left(p(\left[T_R^{-1}(.)\right]_z)\right) - p(\left[T_R^{-1}(.)\right]_z)p\left(\left[T_R^{-1}(.)\right]_z\right)'\right)$$

and

$$B_w(w) = W_w\left(\phi(., w)(h_w(.))^{-1/2}\right)$$

where $B_w(w)$ is a Gaussian process with covariance function $\int_0^1 \cdots \int_0^1 \phi(u, w)\phi(u, w')'du$.

For $\omega \ge 2$ let $K_k(x) = (2\pi)^{-k/2}\sum_{j=1}^{\omega}\theta_j|\sigma_j|^{-k}\exp\left(-1/2x'x/\sigma_j^2\right)$ with $\sum_{j=1}^{\omega}\theta_j = 1$ and $\sum_{j=1}^{\omega}\theta_j|\sigma_j|^{2\ell} = 0$ and all $\ell = 1, 2, ..., \omega - 1$. Let $m_n = O(n^{-(1-\kappa)/2k})$ for some κ with $0 < \kappa < 1$ be a bandwidth sequence and define

$$\begin{aligned}
\hat{F}_1(x_1) &= n^{-1}\sum_{t=1}^{n}\mathbf{1}\{U_{t1} \le x_1\} \\
&\vdots \\
\hat{F}_k(x_k|x_{k-1}, ..., x_1) &= \frac{n^{-1}\sum_{t=1}^n \mathbf{1}\{U_{tk} \le x_k\} K_{k-1}((x_{k-} - U_{tk-})/m_n)}{n^{-1}\sum_{t=1}^n K_{k-1}((x_{k-} - U_{tk-})/m_n)}
\end{aligned}$$

where $x_{k-} = (x_{k-1}, ..., x_1)'$ and $U_{tk-} = (U_{tk-1}, ..., U_{t1})'$. An estimate $\hat{w}_t$ of w_t is then obtained from the recursions

$$\begin{aligned} \hat{w}_{t1} &= \hat{F}_1(U_{t1}) \\ &\vdots \\ \hat{w}_{tk} &= \hat{F}_k(U_{tk}|U_{tk-1}, ..., U_{t1}). \end{aligned}$$

We define $\hat{W}_{w,n}(w) = T_{w,n}\hat{V}_{w,n}(w)$ where $T_{w,n}$ is the empirical version of the Khmaladze transform applied to the vector w_t. Let $\hat{W}_{\hat{w},n}(w)$ denote the process $\hat{W}_{w,n}(w)$ where w_t has been replaced with $\hat{w}_t$. For a detailed formulation of this statistic see Appendix C. An estimate of $h_w(w)$ is defined as

$$\hat{h}_w(.) = \left(\operatorname{diag}\left(p(., \hat{\theta})\right) - p(., \hat{\theta})p\left(., \hat{\theta}\right)'\right).$$

The empirical version of the transformed statistic is

$$\begin{aligned} \hat{B}_{\hat{w},n}(w) &= \hat{W}_{\hat{w},n}\left(\phi(., w)\hat{h}_w(.)^{-1/2}\right) \\ &= n^{-1/2}\sum_{t=1}^{n}\phi\left(\hat{w}_t, w\right)\hat{h}(z_t)^{-1/2}\left[D_t - p(z_t, \hat{\theta}) - \hat{A}_{n,t}\right] \end{aligned} \quad (12)$$

where $\hat{A}_{n,s} = n^{-1}\sum_{t=1}^{n}\mathbf{1}\left\{\hat{w}_{t1} > \hat{w}_{s1}\right\}\frac{\partial p(z_s, \hat{\theta})}{\partial \theta'}\hat{C}_{\hat{w}_{1s}}^{-1}\bar{l}(z_t, \hat{\theta})'\left(D_t - p(z_t, \hat{\theta})\right)$.

Bootstrap based critical values are obtained as follows: the wild bootstrap error distribution is constructed by sampling $\varepsilon_{t,s}^{*}$ for $s = 1, ..., S$ bootstrap replications according to

$$\varepsilon_{t,s}^{*} = \varepsilon_{t,s}^{**}/\sqrt{2} + \left(\left(\varepsilon_{t,s}^{**}\right)^2 - 1\right)/2 \quad (13)$$

where $\varepsilon_{t,s}^{**} \sim N(0,1)$ is independent of the sample. Let the moment condition underlying the transformed test statistic (12) be denoted by

$$m_{T,t}\left(v, \hat{\theta}\right) = \phi\left(\hat{w}_t, w\right)\hat{h}(z_t)^{-1/2}\left[D_t - p(z_t, \hat{\theta}) - \hat{A}_{n,t}\right]$$

and write

$$\hat{B}_{\hat{w},n;s}^{*}(w) = n^{-1/2}\sum_{t=1}^{n}\varepsilon_{t,s}^{*}\left(m_{T,t}\left(v, \hat{\theta}\right) - \bar{m}_{n;T}\left(v, \hat{\theta}\right)\right) \quad (14)$$

to denote the test statistic in a bootstrap replication, with $\bar{m}_{n;T}\left(v, \hat{\theta}\right) = n^{-1}\sum_{t=1}^{n} m_{T,t}\left(v, \hat{\theta}\right)$.

B Asymptotic Critical Values

This Section provides formal results on the distribution of the test statistics described above and forms the basis for the construction of asymptotic critical values. The theorems and proofs use the additional notation outlined below.

B.1 Additional Notation and Assumptions

We focus initially on the process $V_n(v)$ and the associated transformation T. Results for $V_{w,n}(w)$ and the transformed process $T_w V_{w,n}(w)$ then follow as a special case.

Let $\chi_t = [y_t', z_t', D_t]'$ be the vector of observations. Assume that $\{\chi_t\}_{t=1}^{\infty}$ is strictly stationary with values in the measurable space $(\mathbb{R}^{k+1}, \mathcal{B}^{k+1})$ where $\mathcal{B}^{k+1}$ is the Borel σ-field on $\mathbb{R}^{k+1}$ and k is fixed with $2 \leq k < \infty$. Let $\mathcal{A}_1^l = \sigma(\chi_1, ..., \chi_l)$ be the sigma field generated by $\chi_1, ..., \chi_l$. The sequence χ_t is β-mixing or absolutely regular if

$$\beta_m = \sup_{l \geq 1} E\left[\sup_{A \in \mathcal{A}_{l+m}^{\infty}} \left|\Pr\left(A|\mathcal{A}_1^l\right) - \Pr(A)\right|\right] \to 0 \text{ as } m \to \infty.$$

Condition 3 *Let χ_t be a stationary, absolutely regular process such that for some $2 < p < \infty$ and some $\delta > 0$ the β-mixing coefficient of χ_t satisfies $m^{(p+\delta)/(p-2)} (\log m)^{2(p-1)/(p-2)} \beta_m \to 0$.*

Condition 4 *Let $F_u(u)$ be the marginal distribution of U_t. Assume that $F_u(.)$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^k$ and has a density $f_u(u)$ with $f_u(u) > 0$ for all $u \in \mathbb{R}^k$.*

Condition 5 *The matrix of functions $\phi(.,.)$ belongs to a VC subgraph class of functions with envelope $M(\chi_t)$ such that $E\|M(\chi_t)\|^{p+\delta} < \infty$ for the same p and δ as in Condition 3.*

We note that $|m(y_t, D_t, z_t, \theta_0|v)| \leq 2$ for $\phi(., v) = \mathbf{1}\{. \leq v\}$ such that by Pollard (1984) Theorem II.25,

$$m_v(W_t) = m(y_t, D_t, z_t, \theta_0|v) \tag{15}$$

is a VC subgraph class of functions indexed by v with envelope 2.

Condition 6 *Let $H(v)$ be as defined in (2). Assume that $H(v)$ is absolutely continuous in v with respect to Lebesgue measure and for all v, τ such that $v \leq \tau$ with $v_i < \tau_i$ for at least one element v_i of v it follows that $H(v) < H(\tau)$. Let the $\mathcal{M} \times \mathcal{M}$ matrix of derivatives $h(v) = \partial^k H(v)/\partial v_1 ... \partial v_k$ and assume that* $\det(h(v)) > 0$ *for all $v \in \mathbb{R}^k$.*

Remark 1 *A sufficient condition for Condition 6 is that $0 < p_i(z_t, \theta_0) < 1$ almost surely for all $i = 0, 1, ..., \mathcal{M}$, together with Condition 4.*

B.2 Limiting Distributions

Let $\mathfrak{D}[-\infty, \infty]^k$ be the space of functions that are continuous from the right with left limits (Cadlag) mapping $[-\infty, \infty]^k \to \mathbb{R}$. We consider weak convergence on $\mathfrak{D}[-\infty, \infty]^k$ equipped with the sup norm. Here $[-\infty, \infty]^k$ denotes the k-fold product space of the extended real line equipped with the metric $q(v, \tau) =$

$\left(\sum_{i=1}^{k} |\Phi(v_i) - \Phi(\tau_i)|^2\right)^{1/2}$ where Φ is a fixed, bounded and strictly increasing function. It follows that $[-\infty, \infty]^k$ is totally bounded. The function space $\mathcal{F} = \left\{m(.,v)|v \in [-\infty,\infty]^k\right\}$ of functions m indexed by v then is a subset of the space of all bounded functions on $[-\infty,\infty]^k$ denoted by $l^\infty([-\infty,\infty]^k)$.

Proposition 1 *Assume that Conditions 1-6 are satisfied. Let $v_i \in [-\infty,\infty]^k$ for $i = 1, ..., s$ be a finite collection of points. Then, for all finite s, $V_n(v_1), ..., V_n(v_s)$ converges in distribution to a Gaussian limit with mean zero and covariance function $\Gamma(v_i, v_j)$, defined in (3). Moreover, $V_n(v)$ converges in $\mathfrak{D}[-\infty,\infty]^k$ to a Gaussian process $V(v)$ with covariance kernel $\Gamma(v,\tau)$ with $v, \tau \in [-\infty,\infty]^k$ and $V(-\infty) = 0$, $H(v)$ is positive definite with $H(v)$ increasing in v.*

Proof of Proposition 1. As noted before, under H_0, $m_v(\chi_t)$, defined in (15) is a martingale difference sequence such that $E(m_v(\chi_t)|z_t) = 0$. Let $\lambda = (\lambda_1, ..., \lambda_s)'$ with $\|\lambda\| = 1$ and $\lambda_i \in \mathbb{R}^{\mathcal{M}}$. For finite dimensional convergence we apply Corollary 3.1 of Hall and Heyde (1980) to $Y_{t,\lambda} = \lambda_1' m_{v_1}(\chi_t) + \lambda_2' m_{v_2}(\chi_t) + ... + \lambda_s' m_{v_s}(\chi_t)$. Then, clearly $Y_{t,\lambda}$ is also a martingale difference sequence. Consider $Y_{nt} = Y_{t,\lambda}/\sqrt{n}$. Then, for all $\varepsilon > 0$,

$$\sum_t E\left(Y_{nt}^2 \mathbf{1}\left\{|Y_{nt}| \geq \varepsilon\right\}|\mathcal{A}_1^{t-1}\right) \leq \sum_t E\left(Y_{nt}^2 \mathbf{1}\left\{\|M(\chi_t)\| \textstyle\sum_i \|\lambda_i\| \geq \sqrt{n}\varepsilon\right\}|\mathcal{A}_1^{t-1}\right) \to 0 \text{ a.s.}$$

because $E\|M(\chi_t)\|^{2+\delta}$ is bounded for some $\delta > 0$. Also,

$$\begin{aligned}
\sum_t E\left[Y_{nt}^2|\mathcal{A}_1^{t-1}\right] &= n^{-1}\sum_{t=1}^{n} E\left[Y_t^2|\mathcal{A}_1^{t-1}\right] \\
&= n^{-1}\sum_{t=1}^{n}\sum_{i,j=1}^{s} E\left[\lambda_i' \boldsymbol{\phi}(u_t, v_i)\left(\operatorname{diag}(p(z_t)) - p(z_t)p(z_t)'\right)\boldsymbol{\phi}(u_t, v_j)'\lambda_j|\mathcal{A}_1^{t-1}\right] \\
&\xrightarrow{p} \sum_{i,j=1}^{s} \lambda_i' \Gamma(v_i, v_j)\lambda_j
\end{aligned}$$

where the last line is a consequence of Theorem 2.1 in Arcones and Yu (1994). By the Cramer-Wold theorem this establishes finite dimensional convergence. The functional central limit theorem again follows from Theorem 2.1 in Arcones and Yu (1994). ■

The next proposition establishes a linear approximation to the process $\hat{V}_n(v)$ evaluated at the estimated parameter value $\hat{\theta}$. The fact that $l(D_t, z_t, \theta_0)$ is a martingale difference sequence is critical to the development of a distribution free test statistic. The next condition states that the propensity score $p(z_t, \theta)$ is the correct parametric model for the conditional expectation of D_t and lists a number of additional regularity conditions.

Condition 7 *Let $\theta_0 \in \Theta$ where $\Theta \subset \mathbb{R}^d$ is a compact set and $d < \infty$. Assume that $E[D_t|z_t] = p(z_t|\theta_0)$ and for all $\theta \neq \theta_0$ it follows $E[D_t|z_t] \neq p(z_t|\theta)$. Assume that $p(z_t|\theta)$ is differentiable a.s. for $\theta \in \{\theta \in \Theta | \|\theta - \theta_0\| \leq \delta\} \equiv N_\delta(\theta_0)$ for some $\delta > 0$. Let $N(\theta_0)$ be a compact subset of the union of all neighborhoods $N_\delta(\theta_0)$ where $\partial p(z_t|\theta)/\partial\theta$, $\partial^2 p(z_t|\theta)/\partial\theta_i\partial\theta_j$ exists and assume that $N(\theta_0)$ is not empty. Let $\partial p_i(z_t|\theta)/\partial\theta_j$ be the i,j-th element of the matrix of partial derivatives $\partial p(z_t|\theta)/\partial\theta'$ and let $\bar{l}_{i,j}(z_t,\theta)$ be the i,j-th element of $\bar{l}(z_t,\theta)$. Assume that there exists a function $B(x)$ and a constant $\alpha > 0$ such that*

$$\left|\partial p_i(x|\theta)/\partial\theta_j - \partial p_i(x|\theta')/\partial\theta_j\right| \leq B(x)\left\|\theta - \theta'\right\|^\alpha,$$

$$\left|\partial^2 p_k(x|\theta)/\partial\theta_i\partial\theta_j - \partial^2 p_k(x|\theta)/\partial\theta_i\partial\theta_j\right| \leq B(x)\left\|\theta - \theta'\right\|^\alpha \text{ and}$$

$$\left|\partial\bar{l}_{i,j}(x|\theta)/\partial\theta_k - \partial\bar{l}_{i,j}(x|\theta')/\partial\theta_k\right| \leq B(x)\left\|\theta - \theta'\right\|^\alpha$$

for all i,j,k and $\theta,\theta' \in \operatorname{int} N(\theta_0)$, $E|B(z_t)|^{2p+\delta} < \infty$, $E|\partial p_i(z_t|\theta_0)/\partial\theta_j|^{2p+\delta} < \infty$,

$$E\left[p_i(z_t,\theta_0)^{-(2p+\delta)}\right] < \infty$$

and

$$E\left[\left|\partial p_i(z_t|\theta_0)/\partial\theta_j\right|^{\frac{2p+\delta}{2}}\right] < \infty$$

for all $i = 0,...,\mathcal{M}$, and j and some $\delta > 0$.

Remark 2 *By Pakes and Pollard (1989, Lemma 2.13) the uniform Lipschitz condition for the derivatives $\partial p_i(z_t|\theta)/\partial\theta_j$ guarantees that the functions $\partial p(z_t|\theta)/\partial\theta'$ indexed by θ form a Euclidean class for the envelope $B(z_t)\left(2\sqrt{d}\sup_{N(\theta_0)}\left\|\theta - \theta'\right\|\right)^\alpha + |\partial p_i(z_t|\theta_0)/\partial\theta_j|$.*

Remark 3 *In Condition 7, $p_0(z_t|\theta)$ is defined as $p_0(z_t|\theta) = 1 - \sum_{i=1}^{\mathcal{M}} p_i(z_t,\theta)$.*

Condition 8 *Let*

$$l(D_t,z_t,\theta) = \Sigma_\theta^{-1}\frac{\partial p'(z_t,\theta)}{\partial\theta}h(z_t,\theta)^{-1}(\mathcal{D}_t - p(z_t,\theta)) \tag{16}$$

where

$$h(z_t,\theta) = \left(\operatorname{diag}(p(z_t,\theta)) - p(z_t,\theta)p(z_t,\theta)'\right)$$

and

$$\Sigma_\theta = E\left[\frac{\partial p'(D_t|z_t,\theta)}{\partial\theta}h(z_t,\theta)^{-1}\frac{\partial p(D_t|z_t,\theta)}{\partial\theta'}\right]. \tag{17}$$

Assume that Σ_θ is positive definite for all θ in some neighborhood $N \subset \Theta$ such that $\theta_0 \in \operatorname{int} N$ and $0 < \|\Sigma_\theta\| < \infty$ for all $\theta \in N$. Let $l_i(D_t,z_t,\theta)$ be the i-th element of $l(D_t,z_t,\theta)$. Assume that there exists a function $B(x_1,x_2)$ and a constant $\alpha > 0$ such that $\left\|\partial l_i(x_1,x_2,\theta)/\partial\theta_j - \partial l_i(x_1,x_2,\theta')/\partial\theta_j\right\| \leq B(x_1,x_2)\left\|\theta - \theta'\right\|^\alpha$ for all i,j and $\theta,\theta' \in \operatorname{int} N$, $E[B(D_t,z_t)] < \infty$ and $E[|l_i(D_t,z_t,\theta)|] < \infty$ for all i.

Remark 4 *Note that for* $P(z_t,\theta) = \mathrm{diag}\,(p(z_t,\theta))$ *it follows that*

$$h(z_t,\theta)^{-1} = P(z_t,\theta)^{-1} + \frac{P(z_t,\theta)^{-1} p(z_t,\theta) p(z_t,\theta)' P(z_t,\theta)^{-1}}{\left(1 - p(z_t,\theta)' P(z_t,\theta)^{-1} p(z_t,\theta)\right)} = P(z_t,\theta)^{-1} + \frac{\mathbf{1}\mathbf{1}'}{1-\sum_{i=1}^{\mathcal{M}} p_i(z_t,\theta)}. \quad (18)$$

Simple algebra then shows that

$$(\mathcal{D}_t - p(z_t,\theta))' h(z_t,\theta)^{-1} \partial p(D_t|z_t,\theta)/\partial\theta' = \partial\ell(\mathcal{D}_t, z_t,\theta)/\partial\theta'$$

where $\ell(\mathcal{D}_t, z_t,\theta) = \sum_{j=0}^{\mathcal{M}} D_{j,t} \log p_i(z_t,\theta)$ *is the log likelihood of the multinomial distribution and* $D_{j,t} = \mathbf{1}\{D_t = j\}$.

Proposition 2 *Assume that Conditions 1-8 are satisfied. Then*

$$\sup_{v\in[-\infty,\infty]^k} \left\| \hat{V}_n(v) - V_n(v) + \dot{m}(v,\theta_0) n^{-1/2} \sum_{t=1}^{n} l(D_t, z_t,\theta_0) \right\| = o_p(1) \quad (19)$$

and if $l(D_t, z_t, \theta_0)$ *is as defined in 16 and 17 then* $\hat{V}_n(v)$ *converges weakly in* $\mathfrak{D}[-\infty,\infty]^k$ *equipped with the sup norm to a limiting Gaussian process with mean zero and covariance function* $\hat{\Gamma}(v,\tau) = \Gamma(v,\tau) - \dot{m}(v,\theta_0)L(\theta_0)\dot{m}(\tau,\theta_0)'$ *where* $L(\theta_0) = \Sigma_{\theta_0}^{-1}$ *is defined in 17.*

Proof of Proposition 2. Note that $\hat{V}_n(v) - V_n(v) = n^{-1/2}\sum_t^n \phi(U_t, v)\left[p(z_t,\theta_0) - p(z_t,\hat{\theta})\right]$such that we can approximate

$$\begin{aligned} \hat{V}_n(v) - V_n(v) &= \frac{1}{n}\sum_t^n \left(\phi(U_t,v)\left[\frac{\partial p(z_t,\theta_n)}{\partial\theta'} - \frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right]\right)\left(n^{1/2}\left(\hat{\theta}-\theta_0\right)\right) \\ &\quad + \frac{1}{n}\sum_t^n \left(\phi(U_t,v)\frac{\partial p(z_t,\theta_0)}{\partial\theta}\right)\left(n^{1/2}\left(\hat{\theta}-\theta_0\right)\right) \end{aligned}$$

where $\|\theta_n - \theta_0\| \leq \left\|\hat{\theta}-\theta_0\right\|$ by the mean value theorem. Let $\dot{m}(\theta,v) = -E\left[\phi(U_t,v)\frac{\partial p(z_t,\theta)}{\partial\theta'}\right]$ and $\dot{m}(U_t,\theta,v) = \phi(U_t,v)\frac{\partial p(z_t,\theta)}{\partial\theta'} - \dot{m}(\theta,v)$. From Pakes and Pollard (1989, Lemmas 2.13 and 2.14) and Condition 7 it follows that $\dot{m}(.,\theta,v)$ is a matrix of functions in a Euclidean class indexed on $N(\theta_0)\times[-\infty,\infty]^k$ with envelope $\mathcal{M}\left(B(z_t)\left(2\sqrt{d}\sup_{N(\theta_0)}\|\theta-\theta'\|\right)^{\alpha} + \sum_{i=1}^{\mathcal{M}}|\partial p_i(z_t|\theta_0)/\partial\theta_j|\right)M(\chi_t)$ for all the elements in the j-th column of $\dot{m}(U_t,\theta,v)$. Note that the factor $\mathcal{M}$ can be replaced with the constant 1 if $\phi(U_t,v)$ is scalar valued. Then

$$\begin{aligned} &\left\|\frac{1}{n}\sum_t^n \phi(U_t,v)\left[\frac{\partial p(z_t,\theta_n)}{\partial\theta} - \frac{\partial p(z_t,\theta_0)}{\partial\theta}\right]\right\| \\ \leq &\sup_{\|\theta-\theta_0\|\leq\delta}\sup_v \left\|\frac{1}{n}\sum_t^n \left[\dot{m}(U_t,\theta,v) - \dot{m}(U_t,\theta_0,v)\right]\right\| + \sup_{\|\theta-\theta_0\|\leq\delta}\|\dot{m}(\theta,v) - \dot{m}(\theta_0,v)\| + o_p(1) = o_p(1) \end{aligned}$$

since $\sup_{\|\theta-\theta_0\|\leq\delta}\sup_v \left\|\frac{1}{n}\sum_t^n \left[\dot{m}(U_t\theta, v) - \dot{m}(U_t, \theta_0, v)\right]\right\| = o_p(1)$ by applying Lemma 2.1 of Arcones and Yu (1994) to each element $\sup_{\|\theta-\theta_0\|\leq\delta}\sup_v \left|\frac{1}{n}\sum_t^n \left[\dot{m}_{i,j}(U_t\theta, v) - \dot{m}_{i,j}(U_t, \theta_0, v)\right]\right|$. This completes the proof of 19 because $\frac{1}{n}\sum_t^n \left(\phi(U_t, v)\frac{\partial p(z_t,\theta_0)}{\partial\theta}\right) \to_p -\dot{m}\left(\theta_0, v\right)$ and $n^{1/2}\left(\hat{\theta}-\theta_0\right) = n^{-1/2}\sum_{t=1}^n l\left(D_t, z_t, \theta_0\right) + o_p\left(1\right)$ by (5).

The second part of the result follows from the fact that the class of functions

$$\mathcal{F} = \left[m_v(.)\right]_i + \left[\dot{m}\left(\theta, v\right) l(.,.,\theta_0)\right]_i$$

where $[.]_i$ denotes the i-th element of a vector, is a Euclidean class by Lemma 2.14 of Pakes and Pollard (1989). Since $m_v(X_t) + \dot{m}\left(\theta, v\right) l(D_t, z_t, \theta_0)$ is a martingale difference sequence with respect to the filtration $\mathcal{A}_1^{t-1}$, finite dimensional convergence to a Gaussian random vector with zero mean and covariance function $\hat{\Gamma}(v,\tau)$ follows from the martingale CLT (Hall and Heyde, Corollary 3.1) and the fact that $0 < \|\Sigma_{\theta_0}\| < \infty$ by Condition 8. Convergence to a weak limit in $\mathfrak{D}\left[-\infty,\infty\right]^k$ then follows again by Theorem 2.1 of Arcones and Yu (1994) as well as van der Vaart and Wellner (1996, Corollary 1.4.5) together with Pakes and Pollard (1989, Lemmas 2.13 and 2.15) to handle the vector case. ■

We now establish that the process $T\hat{V}(v)$, defined in (7) is zero mean Gaussian with covariance function $\Gamma(v,\tau)$. This establishes that the process $W(v) \equiv T\hat{V}(v)$ can be transformed to a distribution free process via Lemma 3.5 and Theorem 3.9 of Khmaladze (1993).

In order to define the transform T we choose a grid $-\infty = \lambda_0 < \lambda_1 < ... < \lambda_N = \infty$ on $[-\infty,\infty]$, let $\Delta\pi_{\lambda_i} = \pi_{\lambda_{i+1}} - \pi_{\lambda_i}$ and set

$$c_N(V) = \sum_{i=1}^N \left\langle \phi\left(., v\right)', \Delta\pi_{\lambda_i}\bar{l}(.,\theta)\right\rangle C_{\lambda_i}^{-1} V(\pi_{\lambda_i}^{\perp}\bar{l}(\vartheta,\theta)). \tag{20}$$

This construction is the same as in Khmaladze (1993) except that we work on $[-\infty,\infty]$ rather than $[0,1]$. In Proposition (3) we show that $c_N(V)$ converges as $N\to\infty$ and $\max_i\left(\Phi(\lambda_{i+1}) - \Phi\left(\lambda_i\right)\right) \to 0$. Let the limit of $c_N(V)$ be denoted as $c(V) = \int \left\langle \phi\left(., v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} V\left(\pi_\lambda^{\perp}\bar{l}(.,\theta)\right)$

Condition 9 *Let $\{A_\lambda\}$ be a family of measurable subsets of $[-\infty,\infty]^k$, indexed by $\lambda\in[-\infty,\infty]$ such that $A_{-\infty} = \varnothing$, $A_\infty = [-\infty,\infty]^k$, $\lambda\leq\lambda' \Longrightarrow A_\lambda \subset A_{\lambda'}$ and $A_{\lambda'}\backslash A_\lambda \to \varnothing$ as $\lambda'\downarrow\lambda$. Assume that the sets $\{A_\lambda\}$ form a V-C class (polynomial class) of sets as defined in Pollard (1984, p.17). Define the projection $\pi_\lambda f(v) = \mathbf{1}\left(v\in A_\lambda\right) f(v)$ and $\pi_\lambda^{\perp} = 1 - \pi_\lambda$ such that $\pi_\lambda^{\perp} f(v) = \mathbf{1}\left(v\notin A_\lambda\right) f(v)$. We then define the inner product $\langle f(.), g(.)\rangle \equiv \int f(u)' dH(u) g(u)$ and the matrix*

$$C_\lambda = \left\langle \pi_\lambda^{\perp}\bar{l}(.,\theta), \pi_\lambda^{\perp}\bar{l}(.,\theta)\right\rangle = \int \pi_\lambda^{\perp}\bar{l}(u,\theta)' dH(u)\pi_\lambda^{\perp}\bar{l}(u,\theta).$$

Assume that $\langle f(v), \pi_\lambda g(v)\rangle$ is absolutely continuous in λ and C_λ is invertible for $\lambda\in[-\infty,\infty)$.

Proposition 3 *Assume Conditions 1-8 hold. Define* $\Upsilon_x = \left\{ v \in [-\infty, \infty]^k \,| v = \pi_x v \right\}$ *for some* $x < \infty$. *Let* $c_N(v)$ *be defined as in 20. Then* $c_N(v)$ *converges with probability* 1 *to* $c(v)$ *for all* $v \in \Upsilon_x$. *Let* $T\hat{V}(v)$ *be as defined in 7. Then* $T\hat{V}(v)$ *is a Gaussian process with zero mean and covariance function* $\Gamma(v, \tau)$ *for all* $v, \tau \in \Upsilon_x$.

Proof of Proposition 3. The proof of this result follows closely Khmaladze (1993) with the necessary adjustments pointed out. First, let $V(v)$ be a Gaussian process on $[-\infty, \infty]^k$ and taking values in $\mathbb{R}^{\mathcal{M}}$ with zero mean and covariance function $\Gamma(v, \tau)$ and $V(-\infty) = 0$. See Kallenberg (1997, p. 201) for the construction of such a process. Then, $V(\pi_\lambda^\perp \bar{l}(., \theta))$ is a process with trajectories that are continuous in λ by essentially the same argument as in Lemma 3.2 of Khmaladze. To see this fix $\alpha \in \mathbb{R}^{\mathcal{M}}$ such that $\alpha' V(\pi_\lambda^\perp \bar{l}(., \theta))$ is a Wiener process on $[-\infty, \infty]$ with mean zero, $\alpha' V(\pi_\infty^\perp \bar{l}(., \theta)) = 0$ and variance $\alpha' C_\lambda \alpha$ with almost all trajectories continuous in λ on $[-\infty, \infty]$. To show that $c_N(v) \to c(v)$ almost surely we adapt the proof of Lemma 3.3 of Khmaladze (1993). As there, define $\rho_1(\xi) = |\xi_1| + ... + |\xi_k|$ for any vector $\xi = (\xi_1, ..., \xi_k) \in \mathbb{R}^k$ and $\rho_\infty(\xi) = \max_i |\xi_i|$. Set $\xi = \left\langle \phi, \Delta\pi_\mu \bar{l}(., \theta) \right\rangle$ and $\eta(\mu, \lambda) = C_\mu^{-1} V(\pi_\mu^\perp \bar{l}(., \theta)) - C_\lambda^{-1} V(\pi_\lambda^\perp \bar{l}(., \theta))$. By Condition 9 the matrix C_λ is invertible on $[-\infty, \infty)$ and C_λ^{-1} is continuous in λ. Then, since $V(\pi_\lambda^\perp \bar{l}(., \theta))$ is continuous in λ almost surely, we have

$$\sup_{\substack{|\Phi(\lambda) - \Phi(\mu)| < \delta \\ \lambda, \mu \in [-\infty, x]}} \rho_\infty(\eta(\mu, \lambda)) \to 0$$

with probability 1 for any fixed $x < \infty$. The remainder of the proof in Khmaladze (1993) then goes through without change.

We first represent $\hat{V}(v)$ in terms of $V(v)$. Let $V(l(., \theta_0)) = \int l(u, \theta_0) db(u)$ as before for any function $l(u, \theta)$ with $\langle l(., \theta), l(., \theta) \rangle < \infty$ and $b(u)$ a zero mean vector Gaussian process with covariance function $H(v \wedge \tau)$ and note that $\hat{V}(v) = V(\phi(., v)) - \dot{m}(v, \theta)\Sigma_\theta^{-1} V(\bar{l}(., \theta_0)')$. In order to establish a corresponding result to Lemma 3.4 of Khmaladze (1993) we first show that $\hat{V}(v) = V(\phi(., v)) - \dot{m}(v, \theta)\Sigma_\theta^{-1} V(\bar{l}(., \theta_0)')$ is a valid representation of the limiting distribution of $\hat{V}_n(v)$ which was derived in Proposition 2. Clearly, $\hat{V}(v)$ is zero mean Gaussian and the covariance function is

$$\begin{aligned} &E\left[V(v)V(\tau)'\right] - \dot{m}(v, \theta_0)\Sigma_\theta^{-1} \int \phi(u, \tau) H(du)\bar{l}(u, \theta_0) - \left(\int \phi(u, v) H(du)\bar{l}(u, \theta_0) \right) \Sigma_\theta^{-1} \dot{m}(\tau, \theta_0)' \\ &+ \dot{m}(v, \theta_0)'\Sigma_\theta^{-1} \left(\int \bar{l}(u, \theta_0)' H(du)\bar{l}(u, \theta_0) \right) \Sigma_\theta^{-1} \dot{m}(\tau, \theta_0). \end{aligned}$$

Note that $dH(u) = \left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right) dF_u(u)$ such that

$$\begin{aligned} \int \phi(u, \tau) dH(u)\bar{l}(u, \theta_0) &= \int \phi(u, \tau) dH(u) \left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right)^{-1} \frac{\partial p(u_2, \theta)}{\partial \theta'} \\ &= \int \phi(u, \tau) \frac{\partial p(u_2, \theta_0)}{\partial \theta} dF_u(u) = \dot{m}(\tau, \theta_0) \end{aligned}$$

and

$$\begin{aligned}
&\int \bar{l}(u,\theta_0)' dH(u) \bar{l}(u,\theta_0) \\
= &\int \frac{\partial p'(u_2,\theta)}{\partial \theta} \left(\operatorname{diag}\left(p(u_2)\right) - p(u_2) p\left(u_2\right)'\right)^{-1} \frac{\partial p(u_2,\theta)}{\partial \theta'} dF_u(u) = \Sigma_\theta
\end{aligned}$$

such that $E\left[\hat{V}(v)\hat{V}(\tau)'\right] = H(v \wedge \tau) - \dot{m}(v,\theta_0)'\Sigma_\theta^{-1}\dot{m}(\tau,\theta_0)$ as required.

We now verify that the transformation T has the required properties. Note that

$$\begin{aligned}
\left\langle \phi(.,v)', \bar{l}(.,\theta)\right\rangle &= \int \phi\left(u,v\right) dH(u) \left(\operatorname{diag}\left(p(u_2)\right) - p(u_2)p\left(u_2\right)'\right)^{-1} \frac{\partial p(u_2,\theta)}{\partial \theta'} \\
&= \dot{m}(v,\theta_0)
\end{aligned}$$

such that $\hat{V}\left(v\right) = V(\phi\left(.,v\right)) - \left\langle \phi\left(.,\tau\right)', \bar{l}(.,\theta)\right\rangle C_{-\infty}^{-1} V(\bar{l}(v,\theta))$.

In order to establish that $T\hat{V}(v) = \hat{V}\left(v\right) - \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \hat{V}(\pi_\lambda^{\perp}\bar{l}(.,\theta))$ has covariance function $\Gamma(v,\tau)$ we first consider $E\left[TV(v)TV\left(v\right)'\right]$ where

$$\begin{aligned}
&E\left[TV(v)TV\left(v\right)'\right] \\
= \; &\Gamma(v,v) - \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^{\perp}\bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
&- \int \left\langle \phi\left(.,v\right)', \pi_\lambda^{\perp}\bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle d\pi_\lambda \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
&+ \int\int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left(\int \pi_\lambda^{\perp}\bar{l}(u,\theta)' dH(u) \pi_\mu^{\perp}\bar{l}(u,\theta)\right) C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
= \; &\Gamma(v,v) - \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^{\perp}\bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
&- \int \left\langle \phi\left(.,v\right)', \pi_\lambda^{\perp}\bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle d\pi_\lambda \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
&+ \int\int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} C_{\lambda\vee\mu} C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle.
\end{aligned}$$

Note that $\left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} C_{\lambda\vee\mu} C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle$ is symmetric in λ and μ such that

$$\begin{aligned}
&\int\int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} C_{\lambda\vee\mu} C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
= &\int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \int_\lambda^\infty \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
&+ \int\int_\mu^\infty \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
= &\int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^{\perp}\bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\
&+ \int \left\langle \phi\left(.,v\right)', \pi_\lambda^{\perp}\bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle d\pi_\lambda \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle
\end{aligned}$$

such that $E\left[TV(v)TV(v)'\right] = \Gamma(v,v)$. By the same arguments it follows that $E[TV(v)TV(\tau)'] = \Gamma(v,\tau)$.

That the result then also holds for $T\hat{V}(v)$ follows from Khmaladze (1993, Theorem 3.9). ■

Khmaladze (1993, Lemmas 3.2-3.4) shows that the argument need not be limited to all v such that $v \in \Upsilon_x$. As noted by Koul and Stute, however, once T is replaced by T_n convergence can only be shown on the subset $\pi_x v$ of $[-\infty,\infty]^k$ for some finite x due to the instability of the estimated matrix C_λ as $\lambda \to \infty$.

The next step is to analyze the transform T when applied to the empirical processes $V_n(v)$ and $\hat{V}_n(v)$ and in particular to show convergence to the limiting counterpart, $T\hat{V}(v)$.

Proposition 4 *Assume Conditions 1-9 are satisfied. Fix $x < \infty$ arbitrary and define*

$$\Upsilon_x = \left\{ v \in [-\infty,\infty]^k \, | v = \pi_x v \right\}.$$

Then,

$$\sup_{v\in\Upsilon_x} \left| T\hat{V}_n(v) - TV_n(v) \right| = o_p(1)$$

and $TV_n(v) \Rightarrow TV(v)$ in $\mathfrak{D}[\Upsilon_x]$ where $\Rightarrow$ denotes weak convergence.

Proof of Proposition 4. By Proposition 2 we have uniformly on $[-\infty,\infty]^k$ that $\hat{V}_n(v) - V_n(v) = \dot{m}(v,\theta_0) n^{-1/2} \sum_{t=1}^n l(D_t, z_t, \theta_0) + o_p(1)$. Thus consider the difference

$$\begin{aligned} & T\hat{V}_n - TV_n \\ = & -\dot{m}(v,\theta_0) n^{-1/2} \sum_{t=1}^n l(D_t, z_t, \theta_0) \\ & - \int \left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0) \right\rangle C_\lambda^{-1} \left(\hat{V}_n \left(\pi_\lambda^\perp \bar{l}(.,\theta_0)' \right) - V_n \left(\pi_\lambda^\perp \bar{l}(.,\theta_0)' \right) \right) + o_p(1). \end{aligned} \tag{21}$$

For $\|\theta_n - \theta_0\| \leq \left\| \hat{\theta} - \theta_0 \right\|$ it follows by the mean value theorem that

$$\begin{aligned} & \hat{V}_n \left(\pi_\lambda^\perp \bar{l}(.,\theta_0)' \right) - V_n \left(\pi_\lambda^\perp \bar{l}(.,\theta_0)' \right) \\ = & n^{-1/2} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t,\theta_0)' \left(p(z_t,\theta_0) - p(z_t,\hat{\theta}) \right) \\ = & n^{-1/2} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t,\theta_0)' \left(\frac{\partial p(z_t,\theta_n)}{\partial\theta'} - \frac{\partial p(z_t,\theta_0)}{\partial\theta'} \right) \left(\hat{\theta} - \theta_0 \right) \\ & + n^{-1/2} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t,\theta_0)' \frac{\partial p(z_t,\theta_0)}{\partial\theta'} \left(\hat{\theta} - \theta_0 \right) \\ \equiv & R_1(\lambda) + R_2(\lambda). \end{aligned}$$

Let $\dot{m}(\theta) = E\left[\frac{\partial p(z_t,\theta)}{\partial \theta'}\right]$ and $\dot{m}(z_t,\theta) = \frac{\partial p(z_t,\theta)}{\partial \theta} - \dot{m}(\theta)$. First consider

$$\begin{aligned}
\sup_{\lambda} \|R_1(\lambda)\| &\leq n^{1/2}\left\|\hat{\theta}-\theta_0\right\| n^{-1}\sum_{t=1}^{n}\left\|\bar{l}(z_t,\theta_0)\right\|\left\|\frac{\partial p(z_t,\theta_n)}{\partial \theta'} - \frac{\partial p(z_t,\theta_0)}{\partial \theta'}\right\| \\
&\leq n^{1/2}\left\|\hat{\theta}-\theta_0\right\| n^{-1}\sum_{t=1}^{n}\left\|\bar{l}(z_t,\theta_0)\right\|\left\|\dot{m}(z_t,\theta_n)-\dot{m}(z_t,\theta_0)\right\| \\
&\quad + n^{1/2}\left\|\hat{\theta}-\theta_0\right\| n^{-1}\sum_{t=1}^{n}\left\|\bar{l}(z_t,\theta_0)\right\|\left\|\dot{m}(\theta_n)-\dot{m}(\theta_0)\right\| \\
&\leq n^{1/2}\left\|\hat{\theta}-\theta_0\right\|\left(n^{-1}\sum_{t=1}^{n}\left\|\bar{l}(z_t,\theta_0)\right\|^2\right)^{1/2}\left(n^{-1}\sum_{t=1}^{n}\left\|\dot{m}(z_t,\theta_n)-\dot{m}(z_t,\theta_0)\right\|^2\right)^{1/2} + o_p(1)
\end{aligned}$$

where the third inequality follows from Hölder's inequality and the fact that $\|\theta_n - \theta_0\| = o_p(1)$ implies by the continuous mapping theorem that $\|\dot{m}(\theta_n) - \dot{m}(\theta_0)\| = o_p(1)$. Together with $E\left\|\bar{l}(z_t,\theta_0)\right\| < \infty$ and Lemma 2.1 of Arcones and Yu (1994) this implies that

$$n^{1/2}\left\|\hat{\theta}-\theta_0\right\| n^{-1}\sum_{t=1}^{n}\left\|\bar{l}(z_t,\theta_0)\right\|\left\|\dot{m}(\theta_n)-\dot{m}(\theta_0)\right\| = o_p(1).$$

By Condition 7 it follows that

$$\left\|\dot{m}(z_t,\theta_n)-\dot{m}(z_t,\theta_0)\right\|^2 \leq k\,|B(z_t)|^2\,\|\theta_n-\theta_0\|^{2\alpha}$$

for some $\alpha > 0$ such that

$$n^{-1}\sum_{t=1}^{n}\left\|\dot{m}(z_t,\theta_n)-\dot{m}(z_t,\theta_0)\right\|^2 \leq k\,\|\theta_n-\theta_0\|^{2\alpha}\,n^{-1}\sum_{t=1}^{n}|B(z_t)|^2 = o_p(1).$$

This establishes $\sup_\lambda \|R_1(\lambda)\| = o_p(1)$ such that uniformly on Υ_x

$$\left\|\int \left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} R_1(\lambda)\right\| \leq \sup_\lambda \|R_1(\lambda)\| \sup_{\lambda:\pi_\lambda\in\Upsilon_x}\|C_\lambda\|^{-1}\int \left\|\left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle\right\| = o_p(1).$$

Next consider $R_2(\lambda) - \left(\int \pi_\lambda^{\perp}\bar{l}(u_2,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u(u)\right) n^{1/2}\left(\hat{\theta}-\theta_0\right)$. Note that

$$E\left[\mathbf{1}\{U_t \notin A_\lambda\}\,\bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right] = \int \pi_\lambda^{\perp}\bar{l}(u_2,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u(u)$$

and

$$\begin{aligned}
& \sup_{\lambda}\left\|\mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)'\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\| \\
\leq & \left\|\bar{l}(z_t,\theta_0)'\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\| \\
= & \left\|\frac{\partial p'(z_t,\theta_0)}{\partial\theta}\left(\operatorname{diag}\left(p(u_2)\right)-p(u_2)p\left(u_2\right)'\right)^{-1}\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\| \\
\leq & \left\|\frac{\partial p'(z_t,\theta_0)}{\partial\theta}\right\|^2\left\|\left(\operatorname{diag}\left(p(u_2)\right)-p(u_2)p\left(u_2\right)'\right)^{-1/2}\right\|^2 \\
\leq & \left(\sup_{u_2}\left(\mathbf{1}_{\mathcal{M}}'\left(\operatorname{diag}\left(p(u_2)\right)-p(u_2)p\left(u_2\right)'\right)^{-1}\mathbf{1}_{\mathcal{M}}\right)\right)\sum_{i=1}^{\mathcal{M}}\sum_{j=1}^{d}\left(\frac{\partial p_i(z_t,\theta_0)}{\partial\theta_j}\right)^2
\end{aligned}$$

where $\left(\sup_{u_2}\left(\mathbf{1}_{\mathcal{M}}'\left(\operatorname{diag}\left(p(u_2)\right)-p(u_2)p\left(u_2\right)'\right)^{-1}\mathbf{1}_{\mathcal{M}}\right)\right)$ is bounded by Condition 6 and $E\left[\left(\frac{\partial p_i(z_t,\theta_0)}{\partial\theta_j}\right)^2\right]$ is bounded by Condition 7. This shows that $\left(1-\mathbf{1}\left\{(y_t,z_t)\in A_\lambda\right\}\right)\bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}$ is a Euclidean class with integrable envelope $\left\|\bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\|$ such that by Lemma 2.1 of Arcones and Yu it follows that

$$\sup_{\lambda}\left\|R_2\left(\lambda\right)-\left(\int\pi_\lambda^{\perp}\bar{l}(u_2,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u\left(u\right)\right)n^{1/2}\left(\hat{\theta}-\theta_0\right)\right\|=o_p(1).$$

It then follows that uniformly on Υ_x

$$\int\left\langle\phi\left(.,v\right)',d\pi_\lambda\bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}\left[R_2\left(\lambda\right)-\int\pi_\lambda^{\perp}\bar{l}(u_2,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u\left(u\right)n^{1/2}\left(\hat{\theta}-\theta_0\right)\right]=o_p\left(1\right).$$

Now note that $\int\pi_\lambda^{\perp}\bar{l}(u_2,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u\left(u\right)=C_\lambda$ such that

$$\begin{aligned}
& \int\left\langle\phi\left(.,v\right)',d\pi_\lambda\bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}\int\pi_\lambda^{\perp}\bar{l}(\vartheta,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u\left(u\right)n^{1/2}\left(\hat{\theta}-\theta_0\right) \\
= & \int\left\langle\phi\left(.,v\right)',d\pi_\lambda\bar{l}(.,\theta_0)\right\rangle n^{1/2}\left(\hat{\theta}-\theta_0\right) \\
= & -\dot{m}(v,\theta_0)n^{1/2}\left(\hat{\theta}-\theta_0\right)=-\dot{m}(v,\theta_0)n^{-1/2}\sum_{t=1}^{n}l\left(D_t,z_t,\theta_0\right)+o_p\left(1\right).
\end{aligned}$$

Substituting back in (21) then shows that $\sup_{v\in\Upsilon_x}\left|T\hat{V}_n\left(v\right)-TV_n\left(v\right)\right|=o_p(1)$.

For the second part of the proposition consider

$$TV_n\left(v\right)=V_n\left(v\right)-\int\left\langle\phi\left(.,v\right)',d\pi_\lambda\bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}n^{-1/2}\sum_{t=1}^{n}\mathbf{1}\left\{U_t\notin A_\lambda\right\}\bar{l}(z_t,\theta_0)'\left(\mathcal{D}_t-p(z_t,\theta_0)\right).$$

Under H_0 it follows that

$$\begin{aligned}
& E\left[\mathbf{1}\left\{U_t\notin A_\lambda\right\}\bar{l}(z_t,\theta_0)\left(\mathcal{D}_t-p(z_t,\theta_0)\right)|z_t\right] \\
= & E\left[\mathbf{1}\left\{U_t\notin A_\lambda\right\}|z_t\right]\bar{l}(z_t,\theta_0)E\left[\left(\mathcal{D}_t-p(z_t,\theta_0)\right)|z_t\right]=0
\end{aligned}$$

such that $TV_n(v)$ is a martingale. The finite dimensional distributions can therefore be obtained from a martingale difference CLT. Let

$$g(y_t, z_t, v) = \int \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta_0) \right\rangle C_\lambda^{-1} \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \theta_0)'$$

such that $TV_n(v) = n^{-1/2} \sum_{t=1}^n (\phi(U_t, v) - g(y_t, z_t, v))(D_t - p(z_t, \theta_0))$. Then let

$$\begin{aligned} Y_{1t}(v) &= \phi(U_t, v)(D_t - p(z_t, \theta_0)), \\ Y_{2t}(v) &= g(y_t, z_t, v)(D_t - p(z_t, \theta_0)), \end{aligned}$$

$Y_t(v) = Y_{1t}(v) - Y_{2t}(v)$ and $Y_{nt}(v) = n^{-1/2} Y_t(v)$. It follows that

$$E\left[Y_{1t}(v) Y_{1t}(v)'\right] = \Gamma(v, v),$$

$$\begin{aligned} & E\left[Y_{2t}(v) Y_{2t}(v)'\right] \\ = & \int\int \left\{ \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta_0) \right\rangle C_\lambda^{-1} \right. \\ & \times E\left[E\left[\mathbf{1}\{U_t \notin A_\lambda\} \mathbf{1}\{U_t \notin A_\mu\} | z_t\right] \frac{\partial p_t(z_t, \theta_0)}{\partial \theta'} \left(\operatorname{diag}(p(z_t)) - p(z_t) p(z_t)'\right)^{-1} \frac{\partial p_t(z_t, \theta_0)}{\partial \theta} \right] \\ & \left. \times\, C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(., \theta_0), \phi(., v)' \right\rangle \right\} \\ = & \int\int \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta_0) \right\rangle C_\lambda^{-1} C_{\mu \vee \lambda} C_\mu^{-1} \left\langle \phi(., v), d\pi_\mu \bar{l}(., \theta_0)' \right\rangle \\ = & \int \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta) \right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^\perp \bar{l}(., \theta), \phi(., v)' \right\rangle \\ & + \int \left\langle \phi(., v)', \pi_\lambda^\perp \bar{l}(., \theta) \right\rangle C_\lambda^{-1} \left\langle d\pi_\lambda \bar{l}(., \theta), \phi(., v)' \right\rangle \end{aligned}$$

and

$$\begin{aligned} E\left[Y_{1t}(v) Y_{2t}(v)'\right] &= \int \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta_0) \right\rangle C_\lambda^{-1} E\left[E\left[\mathbf{1}\{U_t \notin A_\lambda\} \phi(U_t, v) | z_t\right] \frac{\partial p_t(z_t, \theta_0)}{\partial \theta'} \right] \\ &= \int \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta) \right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^\perp \bar{l}(., \theta), \phi(U_t, v)' \right\rangle \end{aligned}$$

which shows that $E\left[Y_t(v) Y_t(v)'\right] = \Gamma(v, v)$. Also, $E\left[Y_{1t}(v) Y_{1t}(\tau)'\right] = \Gamma(v, \tau)$,

$$\begin{aligned} E\left[Y_{2t}(v) Y_{2t}(\tau)'\right] &= \int \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta) \right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^\perp \bar{l}(., \theta), \phi(., \tau)' \right\rangle \\ &\quad + \int \left\langle \phi(., v)', \pi_\lambda^\perp \bar{l}(., \theta) \right\rangle C_\lambda^{-1} \left\langle d\pi_\lambda \bar{l}(., \theta), \phi(., \tau)' \right\rangle \end{aligned}$$

and

$$E\left[Y_{1t}(v) Y_{2t}(\tau)'\right] = \int \left\langle \phi(., v)', \pi_\lambda^\perp \bar{l}(., \theta) \right\rangle C_\lambda^{-1} \left\langle d\pi_\lambda \bar{l}(., \theta), \phi(., \tau)' \right\rangle$$

such that $E\left[Y_t(v)\,Y_t(\tau)'\right] = \Gamma(v,\tau)$. It also follows that $E\left\|Y_t\right\|^{2+\delta} < \infty$ such that the conditional Lindeberg condition of the CLT is satisfied. We conclude that the finite dimensional distributions of $TV_n(v)$ converge to a Gaussian limit with mean zero and covariance function $\Gamma(v,\tau)$. For weak convergence in the function space note that

$$\begin{aligned}\|g(y_t,z_t,v)\| &\leq \int \left\|\left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} \bar{l}(z_t,\theta_0)'\right\| \\ &\leq \int \left\|\left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}\right\| \left\|\bar{l}(z_t,\theta_0)\right\|\end{aligned}$$

where $\int \left\|\left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}\right\|$ is uniformly bounded on Υ_x and $\left\|\bar{l}(z_t,\theta_0)\right\|^2 = \sum_{j=1}^{\mathcal{M}} \sum_{i=1}^{d} \left|\bar{l}_{i,j}(z_t,\theta_0)\right|^2$ such that by the Hölder inequality

$$\left|\bar{l}_{i,j}(z_t,\theta_0)\right|^{2+\delta} \leq (\mathcal{M}+1)^{1+\delta/2}\left(\left|\frac{\partial p_i(z_t,\theta_0)/\partial\theta_j}{p_i(z_t,\theta_0)}\right|^{2+\delta} + \frac{\sum_{i=1}^{\mathcal{M}} \left|\partial p_i(z_t,\theta_0)/\partial\theta_j\right|^{2+\delta}}{\left|1-\sum_{j=1}^{\mathcal{M}} p_j(z_t,\theta_0)\right|^{2+\delta}}\right).$$

By the Cauchy Schwartz inequality it then follows that

$$\begin{aligned}& E\left|\bar{l}_{i,j}(z_t,\theta_0)\right|^{2+\delta} \\ \leq\; & (\mathcal{M}+1)^{1+\delta/2}\left(E\left[\left|\partial p_i(z_t,\theta_0)/\partial\theta_j\right|^{4+2\delta}\right]\right)^{1/2}\left(E\left[p_i(z_t,\theta_0)^{-(4+2\delta)}\right]\right)^{1/2} \\ & + (\mathcal{M}+1)^{1+\delta/2}\sum_{j=1}^{\mathcal{M}}\left(E\left[\left|\partial p_j(z_t,\theta_0)/\partial\theta_j\right|^{4+2\delta}\right]\right)^{1/2}\left(E\left[\left|1-\sum_{j=1}^{\mathcal{M}} p_j(z_t,\theta_0)\right|^{-(4+2\delta)}\right]\right)^{1/2} \\ <\; & \infty\end{aligned}$$

which is bounded for some δ by Condition 7. This shows that $g(y_t,z_t,v)$ is a Euclidean class of functions and by Lemma 2.14 of Pakes and Pollard it follows that $Y_t(v)$ is a Euclidean class of functions. Lemma 2.1 of Arcones and Yu then can be used to establish weak convergence on $\mathfrak{D}\left[\Upsilon_x\right]$. ■

Our main formal result is established next.

Theorem 5 *Assume Conditions 1-9 are satisfied. Fix $x < \infty$ arbitrary and define*

$$\Upsilon_x = \left\{v \in [-\infty,\infty]^k \,|\, v = \pi_x v\right\}.$$

Then, for T_n defined in (8),

$$\sup_{v\in\Upsilon_x}\left|T_n\hat{V}_n(v) - TV_n(v)\right| = o_p(1).$$

Proof of Theorem 5. We start by considering $\hat{C}_\lambda - C_\lambda$. Let

$$C_\lambda(\theta) = E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\}\bar{l}(z_t,\theta)'\frac{\partial p(z_t,\theta)}{\partial\theta'}\right]$$

such that $C_\lambda = C_\lambda(\theta_0)$ and

$$\begin{aligned}\hat{C}_\lambda - C_\lambda &= n^{-1}\sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\}\bar{l}(z_t,\hat{\theta})'\frac{\partial p(z_t,\hat{\theta})}{\partial\theta'} - C_\lambda(\theta_0) \\ &= n^{-1}\sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\}\bar{l}(z_t,\hat{\theta})'\frac{\partial p(z_t,\hat{\theta})}{\partial\theta'} - C_\lambda\left(\hat{\theta}\right) + C_\lambda\left(\hat{\theta}\right) - C_\lambda(\theta_0).\end{aligned}$$

Note that $C_\lambda(\theta) = \int (1 - \mathbf{1}(u \in A_\lambda))\bar{l}(u,\theta)'H(du)\bar{l}(u,\theta)$ such that for any λ,θ it follows that

$$\begin{aligned}\left\|C_{\lambda'}(\theta') - C_\lambda(\theta)\right\| &\leq \left\|\int (\mathbf{1}(u \in A_{\lambda'}) - \mathbf{1}(u \in A_\lambda))\bar{l}(u,\theta')'dH(u)\bar{l}(u,\theta')\right\| \\ &\quad + \left\|\int \mathbf{1}(u \in A_\lambda)\left(\bar{l}(u,\theta')'dH(u)\bar{l}(u,\theta') - \bar{l}(u,\theta)'dH(u)\bar{l}(u,\theta)\right)\right\|\end{aligned}$$

where $|\mathbf{1}(u \in A_{\lambda'}) - \mathbf{1}(u \in A_\lambda)| \leq \mathbf{1}\left(u \in A_{\max(\lambda,\lambda')}\backslash A_{\min(\lambda,\lambda')}\right) \to 0$ as $\lambda' \to \lambda$ by Condition 9. Continuity of $\bar{l}(u,\theta)'\bar{l}(u,\theta)$ and integrability of the envelope function $\left\|\bar{l}(u,\theta_0)\right\|^2$ then establish uniform continuity of $C_\lambda(\theta)$ on $\Upsilon_x \times N(\theta_0)$ by use of the dominated convergence theorem. By continuity of $C_\lambda(\theta)$ and the continuous mapping theorem it now follows that $\left\|C_\lambda\left(\hat{\theta}\right) - C_\lambda(\theta_0)\right\| = o_p(1)$ uniformly on $\Upsilon_x \times N(\theta_0)$. Let $v_n(\theta,\lambda) = n^{-1}\sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\}\bar{l}(z_t,\theta)'\frac{\partial p(z_t,\theta)}{\partial\theta'} - C_\lambda(\theta)$. We note that

$$\left\|\mathbf{1}\{U_t \notin A_\lambda\}\bar{l}(z_t,\theta)'\frac{\partial p(z_t,\theta)}{\partial\theta'}\right\| \leq 2\left\|\bar{l}(z_t,\theta)\right\|^2\left\|\operatorname{diag}(p(z_t)) - p(z_t)p(z_t)'\right\| \leq 2\mathcal{M}\left\|\bar{l}(z_t,\theta)\right\|^2$$

where $\bar{l}_{i,j}(z_t,\theta)$ has the integrable Envelope $B(z_t)\left(2\sqrt{d}\sup_{N(\theta_0)}\left\|\theta - \theta'\right\|\right)^\alpha + \left|\bar{l}_{i,j}(z_t,\theta_0)\right|$ on $N(\theta_0)$ by Condition 7. By Condition 9 the functions $\mathbf{1}\{(y_t,z_t) \in A_\lambda\}$ form a Euclidean class. It now follows from Lemma 2.1 of Arcones and Yu (1994) that, because $n^{1/2}v_n(\theta,\lambda)$ converges weakly to a Gaussian limit, a tightness condition must hold, i.e. for any $\varepsilon,\eta > 0$, $\exists\delta > 0$ such that

$$\limsup_n \Pr\left(\sup_{\lambda,\theta\in\Upsilon_x\times N(\theta_0)}\sup_{\lambda',\theta':d((\lambda,\theta),(\lambda',\theta'))<\delta}\left\|v_n(\theta',\lambda') - v_n(\theta,\lambda)\right\| > \varepsilon\right) < \eta. \tag{22}$$

Property (22) together with the boundedness of the space $\Upsilon_x \times N(\theta_0)$ now implies by a conventional approximation argument, that

$$\sup_{\lambda,\theta\in\Upsilon_x\times N(\theta_0)}\left\|v_n(\theta,\lambda)\right\| = o_p(1).$$

It now follows that

$$\Pr\left(\left\|\hat{C}_\lambda - C_\lambda\left(\hat{\theta}\right)\right\| > \varepsilon\right) \leq \Pr\left(\sup_{\lambda,\theta\in\Upsilon_x\times N(\theta_0)}\left\|v_n(\theta,\lambda)\right\| > \varepsilon\right) + \Pr\left(\hat{\theta} \notin N(\theta_0)\right) \xrightarrow{p} 0 \tag{23}$$

such that $\sup_{\lambda\in\Upsilon_x}\left\|\hat{C}_\lambda - C_\lambda\right\| = o_p(1)$.

Then

$$
\begin{aligned}
T_n\hat{V}_n(v) - TV_n(v) &= -\dot{m}(v,\theta_0)n^{-1/2}\sum_{t=1}^{n} l\left(D_t, z_t, \theta_0\right) + o_p(1) \\
&\quad - \int d\left(\int \phi\left(u,v\right) d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right) \hat{C}_\lambda^{-1}\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \\
&\quad + \int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(.,\theta_0)')
\end{aligned}
$$

where the first line follows from Proposition 2. From before we have

$$
\begin{aligned}
&\int d\left(\int \phi\left(u,v\right) d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right) \hat{C}_\lambda^{-1}\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \\
= &\int d\left(\int \phi\left(u,v\right) d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right) \left(\hat{C}_\lambda^{-1} - C_\lambda^{-1}\right) \hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \\
&+ \int d\left(\int \phi\left(u,v\right) d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right) C_\lambda^{-1}\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})')
\end{aligned}
$$

where

$$
\begin{aligned}
&\left\| \int d\left(\int \phi\left(u,v\right) d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right) \left(\hat{C}_\lambda^{-1} - C_\lambda^{-1}\right) \hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta}))\right\| \\
\leq &\sup_{\lambda\in[-\infty,x]} \left\|\hat{C}_\lambda^{-1} - C_\lambda^{-1}\right\| \int \left\| d\left(\int \phi\left(u,v\right) d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right)\right\| \left\|\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta}))\right\| = o_p(1)
\end{aligned}
$$

by (23). Next we consider

$$
\begin{aligned}
\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') &= n^{-1/2}\sum_{t=1}^{n} \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(U_t,\hat{\theta})' \left(\mathcal{D}_t - p(z_t,\hat{\theta})\right) \\
&= n^{-1/2}\sum_{t=1}^{n} \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(U_t,\theta_0)' \left(\mathcal{D}_t - p(z_t,\theta_0)\right) \\
&\quad + \left[n^{-1/2}\sum_{t=1}^{n} \mathbf{1}\left\{U_t \notin A_\lambda\right\} \left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_n)'}{\partial \theta'}\right] \left(\hat{\theta} - \theta_0\right) \\
&\quad - \left[n^{-1/2}\sum_{t=1}^{n} \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}((y_t,z_t),\theta_0)' \frac{\partial p(z_t,\theta_n)}{\partial \theta'}\right] \left(\hat{\theta} - \theta_0\right) \\
&\quad - \left(\left(\hat{\theta} - \theta_0\right)' \otimes I_{\mathcal{M}}\right) \left[n^{-1/2}\sum_{t=1}^{n} \mathbf{1}\left\{U_t \notin A_\lambda\right\} \left(\frac{\partial p'(z_t,\theta_n)}{\partial \theta} \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_n)}{\partial \theta'}\right] \left(\hat{\theta} - \theta_0\right) \\
&\equiv R_1\left(\lambda\right) + R_2\left(\lambda\right)\left(\hat{\theta} - \theta_0\right) + R_3\left(\lambda\right) n^{1/2}\left(\hat{\theta} - \theta_0\right) + n^{1/2}\left(\hat{\theta} - \theta_0\right)' R_4\left(\lambda\right)\left(\hat{\theta} - \theta_0\right)
\end{aligned}
$$

where $\|\theta_n - \theta_0\| \leq \left\|\hat{\theta} - \theta\right\|$ and we have used the mean value theorem. Note that $R_1 = \int \pi_\lambda^\perp \bar{l}(\vartheta, \theta_0) dV_n(u)$,

$$\begin{aligned}
R_2(\lambda) &= n^{-1/2} \sum_{t=1}^{n} \mathbf{1}\{U_t \notin A_\lambda\} \left((\mathcal{D}_t - p(z_t, \theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t, \theta_0)'}{\partial \theta'} \\
&\quad + n^{-1/2} \sum_{t=1}^{n} \mathbf{1}\{U_t \notin A_\lambda\} \left((\mathcal{D}_t - p(z_t, \theta_0))' \otimes I_{\mathcal{M}}\right) \left(\frac{\partial \operatorname{vec} \bar{l}(U_t, \theta_n)'}{\partial \theta'} - \frac{\partial \operatorname{vec} \bar{l}(U_t, \theta_0)'}{\partial \theta'}\right) \\
&\equiv R_{21}(\lambda) + R_{22}(\lambda, \theta_n)
\end{aligned}$$

satisfies $ER_{21}(\lambda) = 0$ because

$$\begin{aligned}
&E\left[\mathbf{1}\{U_t \notin A_\lambda\} \left((\mathcal{D}_t - p(z_t, \theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t, \theta_0)'}{\partial \theta'} | z_t\right] \\
&= E\left[\left((\mathcal{D}_t - p(z_t, \theta_0))' \otimes I_{\mathcal{M}}\right) | z_t\right] E\left[\mathbf{1}\{U_t \notin A_\lambda\} \frac{\partial \operatorname{vec} \bar{l}(U_t, \theta_0)'}{\partial \theta'} | z_t\right] = 0
\end{aligned}$$

under H_0 such that finite dimensional convergence follows by the martingale difference CLT and uniform convergence follows from the fact that $\mathbf{1}\{U_t \notin A_\lambda\} \left((\mathcal{D}_t - p(z_t, \theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t, \theta_0)'}{\partial \theta'}$ is a Euclidean class of functions by Condition 9. It thus follows that $\sup_\lambda R_{21}(\lambda) = O_p(1)$ and $R_{21}(\lambda)\left(\hat{\theta} - \theta_0\right) = o_p(1)$ uniformly in λ. For the term $R_{22}(\lambda, \theta_n)$ we note that

$$E\left[\mathbf{1}\{U_t \notin A_\lambda\} \left((\mathcal{D}_t - p(z_t, \theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t, \theta)'}{\partial \theta'} | z_t\right] = 0$$

for any θ. By Lemma 2.1 of Arcones and Yu it thus follows that $R_{22}(\lambda, \theta)$ converges to a Gaussian limit process uniformly in λ and θ. Consequently, a tightness condition implied by this result can be used to show that $\limsup \Pr\left[\sup_{\theta: d(\theta, \theta_0) \leq \delta} \|R_{22}(\lambda, \theta)\| > \varepsilon\right] < \eta$ for all $\varepsilon, \eta > 0$ and some $\delta > 0$. Use root-n convergence of θ_n to conclude from this that $R_{22}(\lambda, \theta_n) = o_p(1)$. The terms involving θ_n in the remainder terms R_3 and R_4 containing θ_n can be handled in similar form and we therefore only consider the leading terms where θ_n is replaced by θ_0. For $R_4(\lambda)$ where

$$R_4(\lambda) = n^{-1} \sum_{t=1}^{n} \mathbf{1}\{U_t \notin A_\lambda\} \left(\frac{\partial p'(z_t, \theta_n)}{\partial \theta} \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t, \theta_n)}{\partial \theta'}$$

we note that $n^{1/2}(R_4(\lambda) - ER_4(\lambda))$ satisfies the conditions of Lemma 2.1 of Arcones and Yu (1994) such that it follows by similar arguments as before that $\sup_\lambda R_4(\lambda) = O_p(1)$. Then conclude that $n^{1/2}\left(\hat{\theta} - \theta_0\right)' R_4(\lambda)\left(\hat{\theta} - \theta_0\right) = o_p(1)$ uniformly in λ.

For $R_3(\lambda)$ note that

$$R_3(\lambda) = n^{-1} \sum_{t=1}^{n} \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(U_t, \theta_0)' \frac{\partial p(z_t, \theta_0)}{\partial \theta'}$$

uniformly converges to

$$E[R_3(\lambda)] = E\left[\mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(U_t, \theta_0) \frac{\partial p(z_t, \theta_0)}{\partial \theta'}\right] = C_\lambda.$$

We have thus established that

$$\sup_{\lambda}\left\|\hat{V}_n(\pi_\lambda^\perp\bar{l}(.,\hat{\theta}))-V_n(\pi_\lambda^\perp\bar{l}(.,\theta_0))-C_\lambda n^{1/2}\left(\hat{\theta}-\theta_0\right)\right\|=o_p\left(1\right).$$

Using this result we obtain

$$\begin{aligned}&\int d\left(\int\phi\left(u,v\right)d\hat{H}_n(u)\pi_\lambda\bar{l}(u,\hat{\theta})\right)C_\lambda^{-1}\left(\hat{V}_n(\pi_\lambda^\perp\bar{l}(.,\hat{\theta}))-V_n(\pi_\lambda^\perp\bar{l}(.,\theta_0))\right)\\&\qquad=\int d\left(\int\phi\left(u,v\right)d\hat{H}_n(u)\pi_\lambda\bar{l}(u,\hat{\theta})\right)n^{1/2}\left(\hat{\theta}-\theta_0\right)+o_p(1).\end{aligned}$$

The leading term is then

$$\begin{aligned}\operatorname{vec}\int d\left(\int\phi\left(u,v\right)d\hat{H}_n(u)\pi_\lambda\bar{l}(u,\hat{\theta})\right)&=\operatorname{vec}\int d\left(\int\phi\left(u,v\right)dH_n(u)\pi_\lambda\bar{l}(u,\theta_0)\right)\\&\quad+\int d\left(\int\phi\left(u,v\right)\pi_\lambda\frac{\partial\operatorname{vec}\partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'}d\hat{F}_u(u)\right)\left(\hat{\theta}-\theta_0\right)\end{aligned}\tag{24}$$

where $\hat{F}_u(u)$ is defined in (27) in Appendix C.1 and

$$\begin{aligned}&\left\|\int d\int\phi\left(u,v\right)\pi_\lambda\frac{\partial\operatorname{vec}\partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'}d\hat{F}_u(u)\right\|\\\leq\;&n^{-1}\sum_{t=1}^n\left\|\phi\left(U_t,v\right)\frac{\partial\operatorname{vec}\partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'}\right\|\\\leq\;&n^{-1}\sum_{t=1}^n\left\|M\left(\chi_t\right)\right\|\left\|\frac{\partial\operatorname{vec}\partial p(u_2,\theta_0)/\partial\theta'}{\partial\theta'}\right\|\\&+n^{-1}\sum_{t=1}^n\left\|\frac{\partial\operatorname{vec}\partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'}-\frac{\partial\operatorname{vec}\partial p(u_2,\theta_0)/\partial\theta'}{\partial\theta'}\right\|\\\leq\;&n^{-1}\sum_{t=1}^n\left\|M\left(\chi_t\right)\right\|\left\|\frac{\partial\operatorname{vec}\partial p(u_2,\theta_0)/\partial\theta'}{\partial\theta'}\right\|+C\left\|\theta_n-\theta_0\right\|^\alpha n^{-1}\sum_{t=1}^n B(z_t)\\=\;&O_p(1)\end{aligned}$$

where C is a finite constant, the third inequality uses Condition 7 and the last equality follows from a standard law of large numbers for strong mixing sequences. The first term in 24 then is

$$\int d\left(\int\phi\left(u,v\right)dH_n(u)\pi_\lambda\bar{l}(u,\theta_0)\right)=n^{-1}\sum_{t=1}^n\phi\left(U_t,v\right)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}$$

where $E\left[\phi\left(U_t,v\right)\frac{\partial p(z_t,\theta_0)}{\partial\theta}\right]=-\dot{m}(v,\theta_0)$ for $v\in\Upsilon_x$. It thus follows again by a law or large numbers that $\int d\left(\int\phi\left(u,v\right)dH_n(u)\pi_\lambda\bar{l}(u,\theta_0)\right)=-\dot{m}(v,\theta_0)+o_p\left(1\right)$ uniformly on Υ_x.

Finally we need to show that

$$\int\left(d\left(\int\phi\left(u,v\right)dH_n(u)\pi_\lambda\bar{l}(u,\theta_0)\right)-\left\langle\phi\left(.,v\right)',d\pi_\lambda\bar{l}(.,\theta_0)\right\rangle\right)C_\lambda^{-1}V_n(\pi_\lambda^\perp\bar{l}(u,\theta_0))=o_p(1).\tag{25}$$

Let $g(z_t, \lambda, v) = \phi(U_t, v) \mathbf{1}\{U_t \in A_\lambda\} \frac{\partial p(z_t,\theta_0)}{\partial\theta}$. We first note that uniformly in λ on $[-\infty, x]$ and $v \in \Upsilon_x$,

$$\int \phi(., v) \pi_\lambda dH_n(v)\bar{l}(., \theta_0) - \left\langle \phi(., v)', \pi_\lambda \bar{l}(., \theta_0) \right\rangle = n^{-1} \sum_{t=1}^{n} g(z_t, \lambda, v) - E(g(z_t, \lambda, v)) \to 0 \text{ a.s.}$$

Weak convergence of $C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(u, \theta_0))$ uniformly in λ on $[-\infty, x]$ can be established by the same methods as for $TV_n(v) \Rightarrow TV(v)$ in the second part of the proof of Proposition 4. We can thus proceed in the same way as Koul and Stute (1999, Lemma 4.2). Let $G_n(\lambda, v) = n^{-1} \sum_{t=1}^{n} g(z_t, \lambda, v)$, $G(\lambda, v) = E(g(z_t, \lambda, v))$ and let $\zeta_n(\lambda) = C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(u, \theta_0)')$. Then each component $\zeta_{ni}(\lambda)$ of the $d \times 1$ vector $\zeta_n(\lambda)$ is asymptotically tight by Prohorov's Theorem. In other words there exists a compact set $\mathbb{K} \subset \mathfrak{D}[-\infty, x]$ such that $\zeta_{ni}(\lambda) \in \mathbb{K}$ with probability no less than $1 - \eta$ for any $\eta > 0$. Following the proof of Lemma 3.1 of Chang (1990) we choose step functions $a_1(\lambda), a_2(\lambda), ..., a_k(\lambda) \in \mathfrak{D}[-\infty, x]$ such that for any $\zeta \in \mathbb{K}$, $d_0(a_i, \zeta) < \varepsilon$ for all $i, 1 \leq i \leq d$ and $d_0(., .)$ is the Skorohod metric. The right hand side of 25 can now be written as $\int_{-\infty}^{x} \zeta_n(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v))$ such that for any $\delta > 0$

$$\begin{aligned} \Pr\left(\left\| \int_{-\infty}^{x} \zeta_n(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v)) \right\| > \eta\right) \quad &\leq \quad \Pr\left(\sup_{\zeta \in \mathbb{K}, v \in \Upsilon_x} \left\| \int_{-\infty}^{x} \zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v)) \right\| > \delta\right) \\ &\quad + \Pr(\zeta_n \notin \mathbb{K}). \end{aligned}$$

Since $\zeta \in \mathbb{K}$ it follows that

$$\sup_{\zeta \in \mathbb{K}, v \in \Upsilon_x} \left| \int_{-\infty}^{x} \zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v)) \right| \leq \sup_{\zeta \in \mathbb{K}} \|\zeta(\lambda)\| \left(\sup_{v \in \Upsilon_x} \int_{-\infty}^{x} \|G(d\lambda, v)\| + \sup_{v \in \Upsilon_x} \int_{-\infty}^{x} \|G_n(d\lambda, v)\| \right)$$

where $\int_{-\infty}^{x} \|G(d\lambda, v)\| = \|G(x, v)\|$ and $\int_{-\infty}^{x} \|G_n(d\lambda, v)\| = \|G_n(x, v)\|$. Since $G(x, v) \to 0$ uniformly in v as $x \to -\infty$ and $G_n(\lambda, v)$ converges uniformly to $G(x, v)$ we can focus on a subset $[x_u, x] \subset [-\infty, x]$ where x_u is such that

$$\sup_{\zeta \in \mathbb{K}, v \in \Upsilon_x} \left\| \int_{-\infty}^{x_u} \zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v)) \right\| < \delta$$

with probability tending to one. Now, for any component i, there exists a strictly increasing, continuous mapping κ of $[-\infty, x]$ onto itself, depending on ζ_i such that $\sup_{-\infty \leq \lambda \leq x} |\kappa(\lambda) - \lambda| < \varepsilon$ and $\sup_{-\infty \leq \lambda \leq x} |\zeta_i(\lambda) - a_i(\kappa(\lambda))| < \varepsilon$. Then for any component i, j of $\zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v))$

$$\begin{aligned} \left| \int_{x_u}^{x} \zeta_i(\lambda) (G_{ni,j}(d\lambda, v) - G_i(d\lambda, v)) \right| \quad &\leq \quad \left| \int_{x_u}^{x} (\zeta_i(\lambda) - a_i(\kappa(\lambda))) (G_{ni,j}(d\lambda, v) - G_{i,j}(d\lambda, v)) \right| \\ &\quad + \left| \int_{x_u}^{x} a_i(\kappa(\lambda)) (G_{ni,j}(d\lambda, v) - G_{i,j}(d\lambda, v)) \right| \end{aligned}$$

which implies that for some N_0 and all $n > N_0$, $\left| \int_{-\infty}^{x} \zeta_i(\lambda) (G_{ni,j}(d\lambda, v) - G_{i,j}(d\lambda, v)) \right| < 3\varepsilon$ uniformly on $\mathbb{K} \times \Upsilon_x$ by the arguments of Chang (1994, p.396) which establishes 25. This now implies that $T_n \hat{V}_n(v) - TV_n(v) = o_p(1)$. ■

Theorem 5 together with Propositions 4 and 3 implies that $\hat{W}_n(v) - V_n(v) = o_p(1)$ uniformly in $v \in \Upsilon_x$. This in turn means that the limiting distribution of $\hat{W}_n(v)$ is a zero mean Gaussian process with covariance function $H(v, \tau)$. This distribution is not nuisance parameter free but can be computed conditional on the sample relatively easily as pointed out in Section 4.

Section 4.2 introduced the distribution free statistic $\hat{B}_{w,n}(w)$, defined as $\hat{B}_{w,n}(w) = \hat{W}_{w,n}\left(\phi(.,w)h_w(.)^{-1/2}\right)$. By the arguments preceding Theorem 5, it follows that $\hat{B}_{w,n}(w) \Longrightarrow B_w(w)$ on $\Upsilon_{[0,1]}$. The only adjustments necessary are a restriction of $[-\infty,\infty]^k$ to $[0,1]^k$. What remains to be shown is that

$$\sup_{w \in \Upsilon_{[0,1]}} \left| \hat{B}_{\hat{w},n}(w) - \hat{B}_{w,n}(w) \right| = o_p(1). \tag{26}$$

This is done in the next Theorem. We impose the following assumptions on the kernel function and density.

Condition 10 *The density $f_u(u)$ is continuously differentiable to some integral order $\omega \geq \max(2,k)$ on $\mathbb{R}^k$ with $\sup_{x \in \mathbb{R}^k} |D^\mu f(x)| < \infty$ for all $|\mu| \leq \omega$ where $\mu = (\mu_1, ..., \mu_k)$ is a vector of non-negative integers, $|\mu| = \sum_{j=1}^k \mu_j$, and $D^\mu f(x) = \partial^{|\mu|} f(x)/\partial x_1^{\mu_1} ....\partial x_k^{\mu_k}$ is the mixed partial derivative of order $|\mu|$. The kernel $K(.)$ satisfies i) $\int K(x)dx = 1, \int x^\mu K(x)dx = 0$ for all $1 \leq |\mu| \leq \omega - 1$, $\int |x^\mu K(x)| dx < \infty$ for all μ with $|\mu| \leq \omega$, $K(x) \to 0$ as $\|x\| \to \infty$ and $\sup_{x \in \mathbb{R}^k} \max(1, \|x\|) |D^{e_i} K(x)| < \infty$ for all $i \leq k$ and e_i is the i-th elementary vector in $\mathbb{R}^k$. ii) $K(x)$ is absolutely integrable and has Fourier transform $\mathfrak{K}(r) = (2\pi)^k \int \exp(ir'x)K(x)dx$ that satisfies $\int (1 + \|r\|) \sup_{b \geq 1} |\mathfrak{K}(br)| dr < \infty$ where $i = \sqrt{-1}$.*

Theorem 6 *Assume Conditions 1-10 are satisfied. Fix $x < 1$ arbitrary and define*

$$\Upsilon_{[0,1]} = \{w \in \Upsilon_\varepsilon | w = \pi_x w\}$$

where Υ_ε is a compact subset of the interior of $[0,1]^k$ with volume $1 - \varepsilon$ for some $\varepsilon > 0$. Then,

$$\sup_{w \in \Upsilon_{[0,1]}} \left| \hat{B}_{\hat{w},n}(w) - \hat{B}_{w,n}(w) \right| = o_p(1).$$

Proof of Theorem 6:. Let $\hat{f}_{u,k-1}(x_{k-}) = n^{-1} \sum_{t=1}^n K_{k-1}((x_{k-} - U_{tk-})/m_n)$ and choose a sequence of positive constants $d_n = O\left(n^{-\kappa/3}\right)$. By Theorem 1(b) of Andrews (1995) it follows that

$$\sup_{x: \hat{f}_{k-1}(x_{k-}) > d_n} \left| \hat{F}_k(x_k | x_{k-1}, ..., x_1) - F_k(x_k | x_{k-1}, ..., x_1) \right| = O_p(T^{-1/2} m_n^{-k} d_n^{-2}) + O_p(m_n^\omega) = o_p(1).$$

Theorem 1(a) of Andrews (1995) implies that $\sup_{x \in \mathbb{R}^{k-1}} \left| \hat{f}_{u,k-1}(x) - f_{u,k-1}(x) \right| = o_p(1)$ where $f_{u,k-1}(x)$ is the marginal density of $f_u(u)$ associated with the frist $k-1$ dimensions. Then for any $\varepsilon, \delta > 0$ there is an n large enough such that

$$\begin{aligned}
&\Pr\left(\sup_{x:T_R(x)\in\Upsilon_{[0,1]}}\left|\hat{F}_k(x_k|x_{k-1},...,x_1)-F_k(x_k|x_{k-1},...,x_1)\right|>\varepsilon\right)\\
\leq\;&\Pr\left(\sup_{x:\hat{f}_{u,k-1}(x_{k-})>d_n}\left|\hat{F}_k(x_k|x_{k-1},...,x_1)-F_k(x_k|x_{k-1},...,x_1)\right|>\varepsilon\right)+\Pr\left(\sup_{x:T_R(x)\in\Upsilon_{[0,1]}}\hat{f}_{u,k-1}\left(x_{k-}\right)\leq d_n\right)\\
\leq\;&\delta/2+1-\Pr\left(\sup_{x:T_R(x)\in\Upsilon_{[0,1]}}\left|\hat{f}_{u,k-1}\left(x_{k-}\right)-f_{u,k-1}\left(x_{k-}\right)\right|+\inf_{x:T_R(x)\in\Upsilon_{[0,1]}}f_{u,k-1}\left(x_{k-}\right)>d_n\right)\leq\delta
\end{aligned}$$

where the last inequality follows from the fact that $\inf_{x:T_R(x)\in\Upsilon_{[0,1]}} f_{u,k-1}\left(x_{k-}\right)>0$ by Condition 4.

By Pakes and Pollard (1989, Lemma 2.15) it follows that the composition of a function from a Euclidean class with envelope M and a measurable map with envelope M_1 forms another Euclidean class with envelope $M\circ M_1$. Since $F_k(x_k|x_{k-1},...,x_1)$ takes values in $[0,1]$ it clearly has an envelope M_1. It follows that $\hat{W}_{w,n}$ is a sample average over functions that belong to a Euclidean class plus remainder terms that vanish by similar arguments as before. It thus follows by the same arguments as before that for all $\varepsilon,\delta>0$ there exists an $\eta>0$ such that

$$\limsup_n \Pr\left(\sup_{\substack{w,w',w_1,w_1'\in\Upsilon_{[0,1]}\\ \|w-w'\|<\eta,\left\|w_1-w_1'\right\|<\eta}}\left|\hat{B}_{w_1,n}(w)-\hat{B}_{w_1',n}(w')\right|>\varepsilon\right)<\delta.$$

It then follows that $\hat{B}_n(s)\Rightarrow B(s)$. ■

This result allows us to conduct inference using critical values that do not depend on nuisance parameters. Although these critical values must be calculated numerically, they are invariant to the sample distribution for a given design.

The next result establishes the validity of the bootstrap procedure proposed in Section 4.3.

Theorem 7 *Assume Conditions 1-10 are satisfied. Fix $x<1$ arbitrary and define*

$$\Upsilon_{[0,1]}=\left\{w\in\Upsilon_\varepsilon|w=\pi_x w\right\}.$$

where Υ_ε is a compact subset of the interior of $[0,1]^k$ with volume $1-\varepsilon$ for some $\varepsilon>0$. For $\hat{B}^*_{\hat{w},n}(w)$ defined in (14) it follows that $\hat{B}^*_{\hat{w},n}(w)$ converges on $\Upsilon_{[0,1]}$ to a Gaussian process $B_w(w)$.

Proof. Following Chen and Fan (1999) we note that conditional on the data, $\hat{B}^*_{\hat{w},n}(w)$ is a Gaussian process with covariance function given by

$$\hat{\Gamma}_w(v,\tau)=n^{-1}\sum_{t=1}^n\left(m_{T,t}\left(v,\hat{\theta}\right)-\bar{m}_{n;T}\left(v,\hat{\theta}\right)\right)\left(m_{T,t}\left(\tau,\hat{\theta}\right)-\bar{m}_{n;T}\left(\tau,\hat{\theta}\right)\right)'.$$

By (26) and similar arguments as in the proof of Proposition 3 and Theorems 4 and 5 it follows that $\hat{\Gamma}_w(v,\tau)$ converges uniformly on $\Upsilon_{[0,1]}$ to the covariance function of $B_w(w)$, $\int \phi(u,v)\phi(u,\tau)du$. The result then follows in the same way as Theorem 5.2 of Chen and Fan (1999). ■

C Implementation Details

C.1 Details for the Khmaladze Transform

To construct the test statistic proposed in the theoretical discussion we must deal with the fact that the transformation T is unknown and needs to be replaced by an estimator. In this section, we discuss the details that lead to the formulation in (10). We also present results for general sets A_λ. We start by defining the empirical distribution

$$\hat{F}_u(v) = n^{-1}\sum_{t=1}^{n}\{U_t \leq v\}, \tag{27}$$

and let

$$\begin{aligned} H_n(v) &= \int_{-\infty}^{v}\left(\operatorname{diag}\left(p(u_2,\theta_0)\right) - p(u_2,\theta_0)p\left(u_2,\theta_0\right)'\right)d\hat{F}_u\left(u\right) \\ &= n^{-1}\sum_{t=1}^{n}\left(\operatorname{diag}\left(p(z_t,\theta_0)\right) - p(z_t,\theta_0)p\left(z_t,\theta_0\right)'\right)\mathbf{1}\left\{U_t \leq v\right\} \end{aligned}$$

as well as

$$\begin{aligned} \hat{H}_n(v) &= \int_{-\infty}^{v}\left(\operatorname{diag}\left(p(z_t,\hat{\theta})\right) - p(z_t,\hat{\theta})p\left(z_t,\hat{\theta}\right)'\right)d\hat{F}_u\left(u\right) \\ &= n^{-1}\sum_{t=1}^{n}\left(\operatorname{diag}\left(p(z_t,\hat{\theta})\right) - p(z_t,\hat{\theta})p\left(z_t,\hat{\theta}\right)'\right)\mathbf{1}\left\{U_t \leq v\right\}. \end{aligned}$$

We now use the sets A_λ and projections π_λ as defined in Section 4.1. Let

$$\begin{aligned} \hat{C}_\lambda &= \int \pi_\lambda^{\perp}\bar{l}(v,\hat{\theta})' d\hat{H}_n(v)\pi_\lambda^{\perp}\bar{l}(v,\hat{\theta}) \\ &= n^{-1}\sum_{t=1}^{n}\left(1-\mathbf{1}\left\{U_t\in A_\lambda\right\}\right)\bar{l}(U_t,\hat{\theta})'\left(\operatorname{diag}\left(p(z_t,\hat{\theta})\right) - p(z_t,\hat{\theta})p\left(z_t,\hat{\theta}\right)'\right)\bar{l}(U_t,\hat{\theta}) \end{aligned}$$

such that

$$T_n\hat{V}_n(v) = \hat{V}_n(v) - \int d\left(\int \phi(u,v)d\hat{H}_n(u)\pi_\lambda\bar{l}(u,\theta)\right)\hat{C}_\lambda^{-1}\hat{V}_n(\pi_\lambda^{\perp}\bar{l}(u,\hat{\theta}))$$

where

$$\int \phi(u,v)\,d\hat{H}_n(u)\pi_\lambda\bar{l}(.,\hat{\theta}) = n^{-1}\sum_{t=1}^{n}\mathbf{1}\left\{U_t\in A_\lambda\right\}\phi(U_t,v)\frac{\partial p(z_t,\hat{\theta})}{\partial\theta'}.$$

Finally, write

$$\hat{V}_n(\pi_\lambda^\perp \bar{l}(u,\hat{\theta})) = n^{-1/2}\sum_{t=1}^{n}\left(1-\mathbf{1}\left\{U_t \in A_\lambda\right\}\right)\bar{l}(U_t,\hat{\theta})'\left(\mathcal{D}_t - p(z_t,\hat{\theta})\right).$$

We now specialize the choice of sets A_λ to $A_\lambda = [-\infty,\lambda]\times[-\infty,\infty]^{k-1}$. Denote the first element of y_t by y_{1t}. Then

$$\hat{C}_\lambda = n^{-1}\sum_{t=1}^{n}\mathbf{1}\left\{y_{1t} > \lambda\right\}\bar{l}(z_t,\hat{\theta})\left(\operatorname{diag}\left(p(z_t,\hat{\theta})\right) - p(z_t,\hat{\theta})p\left(z_t,\hat{\theta}\right)'\right)\bar{l}(z_t,\hat{\theta})', \tag{28}$$

$$\hat{V}_n(\pi_\lambda^\perp \bar{l}(u,\hat{\theta})) = n^{-1/2}\sum_{t=1}^{n}\mathbf{1}\left\{y_{1t} > \lambda\right\}\bar{l}(U_t,\hat{\theta})'\left(\mathcal{D}_t - p(z_t,\hat{\theta})\right) \tag{29}$$

and

$$\int \phi(u,v)d\hat{H}_n(u)\pi_\lambda \bar{l}(u,\hat{\theta}) = n^{-1}\sum_{t=1}^{n}\mathbf{1}\left\{y_{1t} \le \lambda\right\}\phi\left\{U_t,v\right\}\frac{\partial p(z_t,\hat{\theta})}{\partial\theta'} \tag{30}$$

Combining 28, 29 and 30 then leads to the formulation 10.

C.2 Details for the Rosenblatt Transform

As before implementation requires replacement of θ with an estimate. We therefore work with the process $\hat{V}_{w,n}(v) = n^{-1/2}\sum_{t=1}^{n} m_w(w_t,\mathcal{D}_t,\hat{\theta};w)$. Define

$$E\left[m_w(w_t,D_t,\theta);w)\right] = \int_0^1\cdots\int_0^1 \phi(u,w)\left(p\left(\left[T_R^{-1}(u)\right]_z,\theta_0\right) - p(\left[T_R^{-1}(u)\right]_z,\theta)\right)du$$

such that $\dot{m}(w,\theta)$ evaluated at the true parameter value θ_0 is

$$\begin{aligned}\dot{m}_w(w,\theta_0) &= E\left[\phi(U_t,w)\partial p(z_t,\theta_0)/\partial\theta'\right]\\ &= \int_{[0,1]^k}\phi(u,w)\frac{\partial p(\left[T_R^{-1}(u)\right]_z,\theta_0)}{\partial\theta'}du\end{aligned}$$

It therefore follows that $\hat{V}_{w,n}(v)$ can be approximated by $V_{w,n}(v) - \dot{m}_w(w,\theta_0)'n^{-1/2}\sum_{t=1}^{n} l(\mathcal{D}_t,z_t,\theta_0)$. This approximation converges to a limiting process $\hat{V}_w(v)$ with covariance function

$$\hat{\Gamma}_w(w,\tau) = \Gamma_w(w,\tau) - \dot{m}_w(w,\theta_0)'L(\theta_0)\dot{m}_w(\tau,\theta_0)$$

where

$$\Gamma_w(w,\tau) = \int_{[0,1]^k}\phi(u,w)h_w(u)\phi(u,\tau)'du.$$

where $h_w(.,\theta) = \left(\operatorname{diag}\left(p(\left[T_R^{-1}(.),\theta\right]_z)\right) - p(\left[T_R^{-1}(.),\theta\right]_z)p(\left[T_R^{-1}(.)\right]_z,\theta)'\right)$ and $h_w(.) \equiv h_w(.,\theta_0)$.

We represent $\hat{V}_w$ in terms of V_w. Let $V_w(l_w(.,\theta_0)) = \int l_w(w,\theta_0)b_w(dv)$ where $b_w(v)$ is a Gaussian process on $[0,1]^k$ with covariance function $\Gamma_w(v,\tau)$ as before, for any function $l_w(w,\theta)$. Also, define

$$\bar{l}_w(w,\theta) = h_w(w,\theta)^{-1} \frac{\partial p(\left[T_R^{-1}(w)\right]_z,\theta)}{\partial \theta'}$$

such that $\hat{V}_w(w) = V_w(w) - \dot{m}_w(w,\theta_0)V_w\left(\bar{l}_w(w,\theta)\right)$ as before.

Let $\{A_{w,\lambda}\}$ be a family of measurable subsets of $[0,1]^k$, indexed by $\lambda \in [0,1]$ such that $A_{w,0} = \varnothing$, $A_{w,1} = [0,1]^k$, $\lambda \leq \lambda' \Longrightarrow A_{w,\lambda} \subset A_{w,\lambda'}$ and $A_{w,\lambda'} \backslash A_{w,\lambda} \to \varnothing$ as $\lambda' \downarrow \lambda$. We then define the inner product $\langle f(.), g(.)\rangle_w \equiv \int_{[0,1]^k} f(w)' dH_w(w) g(w)$ where

$$H_w(w) = \int_{u \leq w} h_w(u)\, du$$

and the matrix

$$C_{w,\lambda} = \left\langle \pi_\lambda^\perp \bar{l}_w(.,\theta), \pi_\lambda^\perp \bar{l}_w(.,\theta)\right\rangle_w = \int \pi_\lambda^\perp \bar{l}_w(w,\theta)' dH_w(w) \pi_\lambda^\perp \bar{l}_w(w,\theta).$$

and define the transform $T_w V_w(w)$ as before by

$$W_w(w) \equiv T_w \hat{V}_w(w) = \hat{V}_w(w) - \int \left\langle \phi(.,w)', d\pi_\lambda \bar{l}_w(.,\theta)\right\rangle C_\lambda^{-1} \hat{V}_w(\pi_\lambda^\perp \bar{l}_w(.,\theta)').$$

Finally, to convert $W_w(w)$ to a process which is asymptotically distribution free we apply a modified version of the final transformation proposed by Khmaladze (1988, p. 1512) to the process $W(v)$. In particular, using the notation $W_w(\phi(.,w)) = W_w(w)$ to emphasize the dependence of W on ϕ, it follows from the previous discussion that

$$B_w(w) = W_w\left(\phi(.,w)(h_w(.))^{-1/2}\right)$$

where $B_w(w)$ is a Gaussian process on $[0,1]^k$ with covariance function $\int_0^1 \cdots \int_0^1 \phi(u,w)\phi(u,w')du$.

The empirical version of $W_w(w)$, denoted by $\hat{W}_{w,n}(w) = \hat{T}_w \hat{V}_{w,n}(w)$, is obtained as before from

$$\hat{W}_{w,n}(w) = n^{-1/2} \sum_{t=1}^n \left[m_w(w_t, D_t, \hat{\theta}|w) - \phi(w_t,w) \frac{\partial p(z_t,\hat{\theta})}{\partial \theta'} \hat{C}_{w_{t1}}^{-1} n^{-1} \sum_{s=1}^n \mathbf{1}\{w_{s1} > w_{t1}\} \bar{l}(z_s,\hat{\theta})' \left(\mathcal{D}_s - p(z_s,\hat{\theta})\right)\right]$$

where $\hat{C}_{w_{s1}} = n^{-1}\sum_{t=1}^n \mathbf{1}\{w_{t1} > w_{s1}\} \bar{l}(z_t,\hat{\theta})' h\left(z_t,\hat{\theta}\right) \bar{l}(z_t,\hat{\theta})$.

References

ANDREWS, D. W. (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, pp. 560–596.

ARCONES, M. A., AND B. YU (1994): "Central Limit Theorems for Empirical and U-Processes of Stationary Mixing Sequences," *Journal of Theoretical Probability*, pp. 47–71.

CHANG, M. N. (1990): "Weak Convergence of a Self-Consistent Estimator of the Survival Function with Doubly Censored Data," *Annals of Statistics*, 18, 391–404.

CHEN, X., AND Y. FAN (1999): "Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series," *Journal of Econometrics*, 91, 373–401.

HALL, P., AND C. HEYDE (1980): *Martingale Limit Theory and its Application.* Academic Press.

KALLENBERG, O. (1997): *Foundations of Modern Probability.* Springer Verlag, New York.

KHMALADZE, E. (1988): "An Innovation Approach to Goodness-of-Fit Tests in R^m," *Annals of Statistics*, pp. 1503–1516.

KHMALADZE, E. V. (1993): "Goodness of Fit Problem and Scanning Innovation Martingales," *The Annals of Statistics*, pp. 789–829.

KOUL, H. L., AND W. STUTE (1999): "Nonparametric Model Checks for Time Series," *Annals of Statistics*, pp. 204–236.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–1057.

POLLARD, D. (1984): *Convergence of Stochastic Processes.* Springer-Verlag New York, Inc.

VANDERVAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes.* Springer Verlag.