Submitted to the Annals of Statistics

CLASSICAL LAPLACE ESTIMATION FOR $\sqrt[3]{n}$–CONSISTENT ESTIMATORS: IMPROVED CONVERGENCE RATES AND RATE–ADAPTIVE INFERENCE

BY SUNG JAE JUN* JORIS PINKSE YUANYUAN WAN

Center for the Study of Auctions, Procurements, and Competition Policy
Department of Economics
The Pennsylvania State University
and
Department of Economics
University of Toronto

We propose a classical (nonBayesian) Laplace estimator alternative for a large class of $\sqrt[3]{n}$–consistent estimators, including isotonic density and regression estimators, inverse density and regression estimators, the maximum score and mode regression estimators, and interval censoring and monotone hazard rate estimators. The proposed alternative provides a unified method of smoothing that applies to all examples mentioned above; easier computation is a byproduct in the maximum score case. Depending on the choice of input parameter and the degree of smoothness of a population function, the convergence rate of our estimator can be faster than $\sqrt[3]{n}$ and its limit distribution can be normal. With extreme smoothness, a rate close to $\sqrt{n}$ is achievable. We provide a bias reduction method and an inference procedure which automatically adapts to the correct convergence rate and limit distribution.

1. Motivation. This paper provides classical (i.e. nonBayesian) Laplace estimator alternatives for a large class of $\sqrt[3]{n}$–consistent estimators. Since the proposed alternatives are based on integration rather than optimization they can have computational advantages. More importantly, they allow for a unified method of smoothing that improves the rate of convergence (under extra smoothness conditions). The extremum estimators for which we provide alternatives can be divided into two classes: θ *class* estimators and η *class* estimators. We describe both classes below.

Let

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{e}}(\eta) = \operatorname*{argmax}_{\theta\in\Theta} \tilde{\boldsymbol{L}}_{\boldsymbol{n}}(\theta,\eta) \quad \text{with } \tilde{\boldsymbol{L}}_{\boldsymbol{n}}(\theta,\eta) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta,\eta), \tag{1}$$

*This paper is based on research supported by NSF grant SES–0922127. We thank the Human Capital Foundation (www.hcfoundation.ru), especially Andrey P. Vavilov, for their support of CAPCP (http://capcp.psu.edu) at Penn State University. We thank Don Andrews, Miguel Delgado, Jeremy Fox, Bo Honoré, Joel Horowitz, Roger Koenker, Runze Li, Oliver Linton, Peter Robinson, Neil Wallace, Haiqing Xu, Vicky Zinde–Walsh, and numerous departmental seminar and conference participants for helpful suggestions.

AMS 2000 subject classifications: Primary 62G05 (nonparametric estimation), 62G20 (nonparametrics and asymptotic properties); Secondary 62H12 (multivariate analysis and estimation), 62H10 (multivariate analysis and distributions of statistics)

where $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta, \eta) = \tilde{g}(\boldsymbol{\xi}_{\boldsymbol{i}}; \theta, \eta)$ for i.i.d. $\{\boldsymbol{\xi}_{\boldsymbol{i}}\}$ is such that $\hat{\boldsymbol{\theta}}_{\boldsymbol{e}}(\eta)$ converges at the $\sqrt[3]{n}$ rate. The first class of estimators, the θ class, can be expressed as $\hat{\boldsymbol{\theta}}_{\boldsymbol{e}}(\eta_0)$ for a chosen value of η_0; the case where $\tilde{\boldsymbol{L}}_{\boldsymbol{n}}$ does not contain the η parameter also belongs to this class. Asymptotic results for the θ class are provided by Cavanagh (1987); Kim and Pollard (1990). The second class of estimators, the η class, is given by $\hat{\boldsymbol{\eta}}_{\boldsymbol{e}} \in \operatorname{argmin}_{\eta \in N} \|\hat{\boldsymbol{\theta}}_{\boldsymbol{e}}(\eta) - \theta_0\|$ for a chosen value of θ_0.

Examples of the θ class of estimators are inverse density, regression, and hazard rate estimators, the maximum score estimator (Manski, 1975), a quantile estimator with interval censoring, and the mode regression estimator (Lee, 1989). Examples of the η class of estimators are the isotonic density and regression estimators (Grenander, 1956; Brunk, 1958), an interval censoring estimator (Ayer, Brunk, Ewing, Reid, and Silverman, 1955), and a monotone hazard rate estimator (Prakasa Rao, 1970). These examples will be explored in section 5.

We adapt the classical Laplace estimation technique of Chernozhukov and Hong (CH 2003), whose results do not apply to our case. Our alternative to $\hat{\boldsymbol{\theta}}_{\boldsymbol{e}}(\eta)$ is

$$\hat{\boldsymbol{\theta}}(\eta) = \frac{\int \theta \pi(\theta) \exp\{\alpha_n^2 \tilde{\boldsymbol{L}}_{\boldsymbol{n}}(\theta, \eta)\} d\theta}{\int \pi(\theta) \exp\{\alpha_n^2 \tilde{\boldsymbol{L}}_{\boldsymbol{n}}(\theta, \eta)\} d\theta}, \tag{2}$$

where α_n, π are input parameters. Our alternative to $\hat{\boldsymbol{\eta}}_{\boldsymbol{e}}$ is the value $\hat{\boldsymbol{\eta}}$ for which

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\eta}}) = \theta_0. \tag{3}$$

In CH the choice of α_n is largely immaterial for first order asymptotics as long as it diverges sufficiently fast; they use $\alpha_n = \sqrt{n}$. Here, however, both the convergence rate of $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}$ and their limit distributions depend on the divergence rate of α_n and the degree of smoothness of $\tilde{Q}(\theta, \eta) = \mathbb{E}\tilde{\boldsymbol{g}}_{\boldsymbol{1}}(\theta, \eta)$. Intuition for this is that the more slowly α_n diverges, the more the integrations in (2) smooth out $\tilde{\boldsymbol{L}}_{\boldsymbol{n}}$. We distinguish three cases: *(i)* α_n diverges faster than $\sqrt[3]{n}$ *(ii)* α_n diverges at the $\sqrt[3]{n}$ rate, *(iii)* α_n diverges more slowly than $\sqrt[3]{n}$. These cases correspond to the following convergence rates and limit distributions of $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}$: *(i)* a $\sqrt[3]{n}$ convergence rate and a Chernoff–like (limit) distribution identical to that of $\hat{\boldsymbol{\theta}}_{\boldsymbol{e}}, \hat{\boldsymbol{\eta}}_{\boldsymbol{e}}$, *(ii)* a $\sqrt[3]{n}$ rate and a distribution that can be characterized by a ratio of integrals over Gaussian processes, and *(iii)* a $\sqrt{n/\alpha_n}$ rate and a normal distribution.

There are other, optimization–based, techniques that also smooth out the objective function to obtain a better convergence rate under additional conditions, albeit that each method is only relevant for a specific $\sqrt[3]{n}$–consistent estimator. Examples include Barlow and Van Zwet (1969, 1970, 1971); Wright (1982); Friedman and Tibshirani (1984); Mukerjee (1988); Mammen (1991) for either of the isotonic estimators, Horowitz (1992) for the maximum score (MS) estimator, and Zinde-Walsh (2002); Jun, Pinkse, and Wan (2011b) for the least median of squares (LMS) estimator. A fundamental difference with the existing literature is that our method uses integration rather than optimization, which allows us to provide a unified theory of smoothing for a large class of estimators and which can have computational advantages, e.g. in the maximum score case.

Like other methods that smooth the objective function, the limiting normal distribution has non-negligible bias if a convergence rate of $n^{2/5}$ or faster is desired. In our case such bias arises if α_n

decreases at a rate no faster than $\sqrt[5]{n}$. The bias generally depends on $\tilde{Q}$ and π, and it has a convenient expansion provided that $\tilde{Q}$ is sufficiently smooth. The function π can be chosen to eliminate first order asymptotic bias. For instance, the first order bias can be removed by choosing π proportional to $\sqrt{\det\{-\partial_{\theta\theta^\top}\tilde{Q}(\theta,\eta)\}}$ in a neighborhood of (θ_0,η_0); this choice of π resembles the Jeffreys prior (Jeffreys, 1946) but our method is classical, $\tilde{Q}$ is not a loglikelihood, and π is an input parameter we shall, like CH, call a *quasi prior* since it resembles but serves a different purpose from a Bayesian prior. Higher order bias reductions can be achieved by a more complicated choice of π. The function π, then, serves a purpose similar to that of higher order kernels, albeit that its choice need not affect the asymptotic variance and it must satisfy different conditions. Like in the related literature if α_n is chosen based on an assumed degree of smoothness of $\tilde{Q}$ greater than its actual smoothness then the convergence rate will be affected adversely because of excess bias; see Pollard (1993).

The trichotomy of convergence rate and limit distribution raises the practical question of which limit distribution to use since in practice one chooses a value α_n, not a rate. We address this issue by providing a single inference procedure which is asymptotically valid in all three cases. This has the advantage that it provides robustness against choosing α_n 'too large,' in much the same way that Cattaneo, Crump, and Jansson (2011) provide robustness against choosing bandwidths that are too small in regular kernel estimation, albeit that there the limit distributions are always normal. To our knowledge, none of the existing literature for any of the estimation problems considered here provides such robustness. Robustness is of particular importance since several authors, including Sen, Banerjee, and Woodroofe (2010); Abrevaya and Huang (2005), have found the standard bootstrap to be inconsistent, although subsampling was shown to work for the MS estimator by Delgado, Rodrıguez-Poo, and Wolf (2001) and for the isotonic density and regression estimators by Sen, Banerjee, and Woodroofe (2010).

Which method to choose will in the end depend in large part on the performance of the estimator and inference procedure in practice. Since ours is the only method which works for many different estimation problems, a full comparison is beyond the scope of this paper. A detailed simulation comparison to assess the relative merits of each procedure is needed for each application, but the present paper focuses on establishing the theoretical properties of our procedure.

The remainder of this paper is laid out as follows. In section 2 we obtain results for the θ–class of estimators and in section 3 for the η–class of estimators. Section 4 describes our inference procedure and section 5 contains a discussion of potential applications.

2. Theta class. Our first set of results covers the case where $\eta_0 \in N \subset \mathbb{R}^d$ is given and $\theta_0 \in \Theta \subset \mathbb{R}^d$ is the unknown maximizer of $\tilde{Q}(\theta,\eta_0)$. The results are however stated more generally in that they cover convergence in a (shrinking) neighborhood of η_0 to prepare for the case in which θ_0 is given and η_0 is estimated; see section 3.

For any $\eta \in N$, let $\theta_0(\eta)$ be an arbitrary element of $\Theta_0(\eta) = \{\theta \in \Theta : \tilde{Q}(\theta,\eta) = \max_{\tilde{\theta}\in\Theta}\tilde{Q}(\tilde{\theta},\eta)$, and let $\boldsymbol{g_i}(\theta,\eta) = \tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta,\eta) - \tilde{\boldsymbol{g}}_{\boldsymbol{i}}\{\theta_0(\eta),\eta\}$. Then, by definition, $\Theta_0(\eta_0)$ is a singleton containing θ_0. Further, let Q and $\boldsymbol{L_n}$ be defined by $Q(\theta,\eta) = \mathbb{E}\boldsymbol{g_1}(\theta,\eta)$, $\boldsymbol{L_n}(\theta,\eta) = \sum_{i=1}^n \boldsymbol{g_i}(\theta,\eta)/n$, and let $\boldsymbol{S_n} = \boldsymbol{L_n} - Q$. So Q is a normalized version of $\tilde{Q}$ with the property that its maximum value equals zero.

In theorem 1 we derive asymptotic properties of

$$\hat{\boldsymbol{\theta}}^*(w^*,w) = \frac{\int \theta\pi(\theta)\exp\{\alpha_n^2 \boldsymbol{S_n}(\theta,\eta_0+w^*/\mathrm{r}_n) + \alpha_n^2 Q(\theta,\eta_0+w/\mathrm{r}_n)\}d\theta}{\int \pi(\theta)\exp\{\alpha_n^2 \boldsymbol{S_n}(\theta,\eta_0+w^*/\mathrm{r}_n) + \alpha_n^2 Q(\theta,\eta_0+w/\mathrm{r}_n)\}d\theta}, \tag{4}$$

as a process of w^* for fixed values of w, where $\{\alpha_n\}, \{\mathrm{r}_n\}$ are divergent sequences and w^*, w belong to some set $\mathscr{W}$. So for each $w \in \mathscr{W}$, $\hat{\boldsymbol{\theta}}^*(\cdot,w)$ is a random process with paths in $\mathbb{L}^\infty(\mathscr{W})$, where $\mathbb{L}^\infty(\mathscr{A})$ is the collection of bounded functions from $\mathscr{A}$ to a Euclidean space of implicit dimension. This will suffice for the θ class of estimators since there $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^*(0,0)$. The results of theorem 1 are used in theorem 3, in which we develop asymptotic results for $\hat{\boldsymbol{\eta}}$.

Let $\theta_{0n}(w) = \theta_0(\eta_0 + w/\mathrm{r}_n)$, such that for any divergent sequence $\{\tilde{\alpha}_n\}$ (used only for a technical purpose), the substitution $t = \tilde{\alpha}_n\{\theta - \theta_{0n}(w)\}$ yields

$$\tilde{\alpha}_n\{\hat{\boldsymbol{\theta}}^*(w^*,w) - \theta_{0n}(w)\} = \frac{\int t\pi_{nw}(t)\exp\Big\{\frac{\alpha_n^2}{\sqrt{n\tilde{\alpha}_n}}\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*) + \frac{\alpha_n^2}{\tilde{\alpha}_n^2}Q_{nw}(t)\Big\}dt}{\int \pi_{nw}(t)\exp\Big\{\frac{\alpha_n^2}{\sqrt{n\tilde{\alpha}_n}}\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*) + \frac{\alpha_n^2}{\tilde{\alpha}_n^2}Q_{nw}(t)\Big\}dt}, \tag{5}$$

where $\pi_{nw}(t) = \pi\{\theta_{0n}(w) + t/\tilde{\alpha}_n\}$, $Q_{nw}(t) = \tilde{\alpha}_n^2 Q\{\theta_{0n}(w) + t/\tilde{\alpha}_n, \eta_0 + w/\mathrm{r}_n\}$, and $\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*) = \sqrt{n\tilde{\alpha}_n}\boldsymbol{S_n}\{\theta_{0n}(w) + t/\tilde{\alpha}_n, \eta_0 + w^*/\mathrm{r}_n\}$. As part of the proof of theorem 1, we develop conditions under which the random process $\tilde{\boldsymbol{S}}^*_{\boldsymbol{nw}}$ given by $\tilde{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t,w^*) = \tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)/c_t$, with $c_t = \|t\| + 1$, converges weakly in $\mathbb{L}^\infty(\mathbb{R}^d \times \mathscr{W})$ to a limit process $\mathbb{G}^*$ which is flat in w^*.

The reasons for the trichotomous asymptotic distribution of theorem 1 can be inferred from (5). Ignore the dependence on w, w^* and distinguish three cases: *(i)* α_n diverges faster than $\sqrt[3]{n}$. Taking $\tilde{\alpha}_n = \sqrt[3]{n}$ the exponents in (5) then contain $\alpha_n^2 n^{-2/3}(\tilde{\boldsymbol{S}}_{\boldsymbol{nw}} + Q_{nw})$ which, together with a quadratic expansion, suggest a Chernoff–like limit distribution. Indeed, in this case integration is similar but not identical to maximization. *(ii)* α_n diverges at a $\sqrt[3]{n}$–rate; take $\tilde{\alpha}_n = \alpha_n$. Then both $\alpha_n^2/\sqrt{n\tilde{\alpha}_n}$ and $\alpha_n^2/\tilde{\alpha}_n^2$ are $O(1)$, but not $o(1)$, which suggests a ratio of two Gaussian integrals as the limit distribution. *(iii)* α_n diverges more slowly than $\sqrt[3]{n}$. Take $\tilde{\alpha}_n = \alpha_n$, such that the exponents in (5) become $\beta_n\tilde{\boldsymbol{S}}_{\boldsymbol{nw}} + Q_{nw}(t)$ for $\beta_n = \sqrt{\alpha_n^3/n} = o(1)$. Then, using an expansion of the exponential function and the fact that under our assumptions Q_{nw} has a quadratic expansion, a normal distribution follows as the limit distribution.

We now develop formal results. We start with some assumptions, which will be verified for a number of applications in section 5. Please note that for the θ case $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}$ does not depend on η and any conditions pertaining only to η (or w, w^*) can be safely ignored.

Let $g^*(\xi,t,w^*,w;\alpha,\mathrm{r}) = \tilde{g}\{\xi,\theta_0(\eta_0+w/\mathrm{r})+t/\alpha,\eta_0+w^*/\mathrm{r}\} - \tilde{g}\{\xi,\theta_0(\eta_0+w/\mathrm{r}),\eta_0+w^*/\mathrm{r}\}$, and let $\boldsymbol{g}^*_{\boldsymbol{i}}$ be g^* evaluated at $\xi = \boldsymbol{\xi_i}$,

ASSUMPTION. *For some neighborhood $\mathfrak{N}_0 \subset N$ of η_0,*

A. θ_0 *is in the interior of a compact set* Θ*;*
B. *for all* $\theta \neq \theta_0$, $\tilde{Q}(\theta,\eta_0) < \tilde{Q}(\theta_0,\eta_0)$*;*

C. *$\tilde{Q}$ is continuous on $\Theta \times \mathfrak{N}_0$, for some $q \geq 0$, $\tilde{Q}$ is $\Delta = q + 2$ times continuously differentiable in θ at (θ_0, η_0), $\partial_\eta \tilde{Q}$ is $q+2$ times continuously differentiable in θ at (θ_0, η_0), and the minimum and maximum eigenvalues of $-\partial_{\theta\theta^\intercal} \tilde{Q}(\theta_0, \eta_0) = V > 0$ satisfy $0 < \lambda^- \leq \lambda^+ < \infty$;*

D. *π is q times continuously differentiable in θ at θ_0, $\pi_0 = \pi(\theta_0) > 0$, and $\pi(\theta) = 0$ for all $\theta \notin \Theta$;*

E. *for any $w, w^*, \tilde{w}^* \in \mathscr{W}$, $H(t,s) = \lim_{\alpha\to\infty} \alpha \mathbb{E}\{\boldsymbol{g}_1^*(t, w^*, w; \alpha, \alpha) \boldsymbol{g}_1^*(s, \tilde{w}^*, w; \alpha, \alpha)\}$ does not depend on $w, w^*, \tilde{w}^*$;*

F. *for all $\xi \in \Xi$, $\tilde{g}(\xi; \theta, \eta)$ is right– (or left–) continuous in θ and η at (θ_0, η_0).* □

Assumptions A to D are standard except for the presence of the "prior" π. Note that by assumptions A to C and an implicit function theorem argument $\mathfrak{N}_0$ can be taken small enough to ensure that $\Theta_0(\eta)$ consists only of a singleton $\theta_0(\eta)$ for all $\eta \in \mathfrak{N}_0$, which we shall do from hereon. Assumption E is the cause of the $\sqrt[3]{n}$ convergence rate of $\hat{\boldsymbol{\theta}}_e$ (see Kim and Pollard, 1990); assumption F allows for the presence of discontinuous functions such as indicator functions.

Assumption E implies that H is a positive definite covariance kernel and that

$$\forall t, s \in \mathbb{R}^d : \begin{cases} \forall c > 0 : H(ct, cs) = cH(t,s), \\ H(t,t) + H(s,s) - 2H(t,s) = H(t-s, t-s), \end{cases} \tag{6}$$

where the second implication requires some simple but tedious manipulations. Assumption G below is relatively high level to maintain a desirable degree of generality, but will be shown to be verified for a range of applications in section 5.

Let $\{\tilde{\alpha}_n\}$ and $\{\mathfrak{r}_n\}$ be (positive) sequences with $0 \leq \tilde{\alpha}_n \leq \mathfrak{r}_n$ and $1/\tilde{\alpha}_n = o(1)$. Let further $g^*_{nw}(\xi; t, w^*) = g^*(\xi, t, w^*, w; \tilde{\alpha}_n, \mathfrak{r}_n)$ and $g^\circ_{nw}(\xi; t, w^*) = \sqrt{\tilde{\alpha}_n} g^*_{nw}(\xi; t, w^*)/c_t$, where $\xi \in \Xi$ the support of $\boldsymbol{\xi}_i$.

Assumption G. *Let $\mathscr{W} \subset \mathbb{R}^d$ be compact and let $\mathscr{F}_{nw} = \{g^\circ_{nw}(\cdot; t, w^*) : (t, w^*) \in \mathbb{R}^d \times \mathscr{W}\}$. Then*

(i) *there exists an envelope function $\mathfrak{F}_n$ such that for all $\xi \in \Xi$, $\sup_{(t,w^*)\in\mathbb{R}^d\times\mathscr{W}} |g^\circ_{nw}(\xi; t, w^*)| \leq \mathfrak{F}_n(\xi)$;*

(ii) *for $\boldsymbol{\mathfrak{F}}_{ni} = \mathfrak{F}_n(\boldsymbol{\xi}_i) : \mathbb{E}\boldsymbol{\mathfrak{F}}^2_{n1} = O(1)$;*

(iii) *for any $\epsilon > 0$, $\mathbb{E}\{\boldsymbol{\mathfrak{F}}^2_{n1} \mathbb{1}(\boldsymbol{\mathfrak{F}}_{n1} > \epsilon\sqrt{n})\} = o(1)$;*

(iv) *for any $0 < \epsilon_n = o(1)$,*

$$\sup_{\substack{\|(t,w^*)-(s,\tilde{w}^*)\| \leq \epsilon_n \\ (t,s,w^*,\tilde{w}^*)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{W}\times\mathscr{W}}} \mathbb{E}\{\boldsymbol{g}^\circ_{nw1}(t, w^*) - \boldsymbol{g}^\circ_{nw1}(s, \tilde{w}^*)\}^2 = o(1);$$

(v) *let $\mathscr{N}\{\epsilon, \mathscr{F}_{nw}, \mathbb{L}^2(\mathscr{P})\}$ be the ($\mathbb{L}^2$–) covering number for $\mathscr{F}_{nw}$ with respect to the probability measure $\mathscr{P}$. Then for every $0 < \epsilon_n = o(1)$,*

$$\sup_{\mathscr{P}^*} \int_0^{\epsilon_n} \sqrt{\log[\mathscr{N}\{\epsilon \|\boldsymbol{\mathfrak{F}}_{ni}\|_{\mathscr{P}^*,2}, \mathscr{F}_n, \mathbb{L}^2(\mathscr{P}^*)\}]} d\epsilon = o(1),$$

where $\sup_{\mathscr{P}*}$ *is the supremum taken over all finitely discrete probability measures* $\mathscr{P}*$ *with* $\|\mathfrak{F}_{\boldsymbol{n}i}\|_{\mathscr{P}*,2} > 0$. □

We make a few notational comments before we proceed. For a fixed value $w \in \mathscr{W}$, we will write $\mathbb{Y}_{\boldsymbol{n}}(w^*, w) \overset{w}{\to} \mathbb{Y}(w^*, w)$ in $\mathbb{L}^\infty(\mathscr{W})$ when the process $\mathbb{Y}_{\boldsymbol{n}}(\cdot, w)$ converges weakly to $\mathbb{Y}(\cdot, w)$, where $\mathbb{Y}_{\boldsymbol{n}}(\cdot, w)$ and $\mathbb{Y}(\cdot, w)$ have paths in $\mathbb{L}^\infty(\mathscr{W})$. When the limit process is flat in w^*, we will write $\mathbb{Y}_{\boldsymbol{n}}(w^*, w) \overset{w}{\to} \mathbb{Y}(w)$ in $\mathbb{L}^\infty(\mathscr{W})$. Similarly, we will use shorthand notation like $\mathbb{Y}_{\boldsymbol{n}}(t, w^*, w) \overset{w}{\to} \mathbb{Y}(t)$ in $\mathbb{L}^\infty(\mathbb{R}^d \times \mathscr{W})$ when the process $\mathbb{Y}_{\boldsymbol{n}}(\cdot, \cdot, w)$ converges weakly to $\mathbb{Y}(\cdot, \cdot, w)$, where $\mathbb{Y}_{\boldsymbol{n}}(\cdot, \cdot, w)$ and $\mathbb{Y}(\cdot, \cdot, w)$ have paths in $\mathbb{L}^\infty(\mathbb{R}^d \times \mathscr{W})$ and the limit process $\mathbb{Y}(\cdot, \cdot, w)$ depends neither on w^* nor on w.

We are now ready to state our first theorem. Let $C_V = \int \exp\bigl(-t^\intercal V t/2\bigr) dt$ and let ϕ_V be a normal density function with variance V^{-1}, where $t^\intercal$ denote the transpose of t.

THEOREM 1. *Suppose that assumptions A through F are satisfied. Then*

(i) if for some $c_\alpha > 0$, $\alpha_n/\sqrt[3]{n} - c_\alpha^2 = o(1)$ *and assumption G is satisfied for* $\mathfrak{r}_n \geq \tilde{\alpha}_n = \alpha_n$ *then for all* $w \in \mathscr{W}$,

$$\sqrt[3]{n}\bigl\{\hat{\boldsymbol{\theta}}^*(w^*, w) - \theta_{0n}(w)\bigr\} \overset{w}{\to} \frac{1}{c_\alpha^2} \frac{\int t \exp\bigl\{c_\alpha^3 \mathbb{G}(t)\bigr\}\phi_V(t)dt}{\int \exp\bigl\{c_\alpha^3 \mathbb{G}(t)\bigr\}\phi_V(t)dt}, \tag{7}$$

in $\mathbb{L}^\infty(\mathscr{W})$, *where* $\mathbb{G}$ *is a tight Gaussian process with covariance kernel* H*;*

(ii) if $\sqrt[3]{n}/\alpha_n = o(1)$ *and assumption G is satisfied for* $\mathfrak{r}_n \geq \tilde{\alpha}_n = \sqrt[3]{n}$ *then for all* $w \in \mathscr{W}$,

$$\sqrt[3]{n}\bigl\{\hat{\boldsymbol{\theta}}^*(w^*, w) - \theta_{0n}(w)\bigr\} \overset{w}{\to} \operatorname*{argmax}_{t \in \mathbb{R}^d} \mathbb{C}(t),$$

in $\mathbb{L}^\infty(\mathscr{W})$, *where* $\mathbb{C}(t) = \mathbb{G}(t) - t^\intercal V t/2$*;*

(iii) if α_n *diverges at a polynomial rate,* $\alpha_n = o(\sqrt[3]{n})$, *and assumption G is satisfied for* $\mathfrak{r}_n \geq \tilde{\alpha}_n = \alpha_n$ *then for all* $w \in \mathscr{W}$,

$$\sqrt{n/\alpha_n}\bigl\{\hat{\boldsymbol{\theta}}^*(w^*, w) - \theta_{0n}(w)\bigr\} - \frac{\mathbb{B}_{nw}}{C_V \pi_0 + o_p(1)} \overset{w}{\to} N(0, \mathscr{V}),$$

in $\mathbb{L}^\infty(\mathscr{W})$, *where* $\mathscr{V} = \iint t s^\intercal H(t,s)\phi_V(t)\phi_V(s)dtds$ *and* $\mathbb{B}_{nw}$ *(defined in (30)) has an expansion provided in part (iv);*

(iv) In part (iii), if moreover $q \geq 1$ *and* $\mathfrak{r}_n/\alpha_n^q = o(1)$ *then for all* $w \in \mathscr{W}$,

$$\mathbb{B}_{nw} = \frac{C_V}{\beta_n} \sum_{\tau=0}^{q} \frac{b_{q\tau}^*}{\alpha_n^\tau} + o\biggl(\frac{1}{\alpha_n^q \beta_n}\biggr), \tag{8}$$

where $b_{q\tau}^* = \sum_{j=0}^{\tau} b_{qj,\tau-j}$ *and*

$$b_{qjs} = \sum_{p=0}^{q} \sum_{m \in \mathscr{M}_{pqj}^*} \int D_{\pi s}(t) \biggl\{\prod_{\delta=1}^{q} \frac{D_{Q;\delta+2}^{m_\delta}(t)}{m_\delta!}\biggr\} t \phi_V(t) dt, \tag{9}$$

with $\mathscr{M}^*_{pqj}$ the collection of q–dimensional vectors consisting of nonnegative integers for which $\sum_{\delta=1}^{q} m_\delta = p$ and $\sum_{\delta=1}^{q} \delta m_\delta = j$ and where $D_{\pi s}$ and D_{Qs} are term s in a Taylor expansion of $\pi(\theta_0 + t)$ and $Q(\theta_0 + t, \eta_0)$ around θ_0, respectively.

PROOF. The proof of parts (i), (ii), (iii) and (iv) can be found in appendices B, C, D, and E, respectively. □

Theorem 1 implies that, as promised, in the θ case our Laplace estimator can have three different limit distributions, depending on the amount of smoothing. First, if little smoothing is applied (α_n is chosen large) then the limit distributions of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{e}}$ coincide, but even for $\alpha_n = \infty$ the estimators generally have different values; indeed, the Manski (1975) estimator is set–valued. The normal limit distribution and improved convergence rate only arise if a substantial amount of smoothing is applied (α_n diverges more slowly than $\sqrt[3]{n}$). Equations (8) and (9) plus some elbow grease reveal that $b^*_{q\tau} = 0$ for even values of τ and that for $q = 1$,

$$\mathbb{B}_{nw} = C_V \sqrt{n/\alpha_n^5} \int \big\{D_{\pi 1}(t)t + \pi_0 D_{Q3}(t)\big\}\phi_V(t)dt + o\big(\sqrt{n/\alpha_n^5}\big). \tag{10}$$

Hence for $q = 1$, if $\sqrt[5]{n} = o(\alpha_n)$ then the bias is negligible and if $\alpha_n = o(\sqrt[5]{n})$ then the bias dominates. If $\alpha_n = c_\alpha^2 \sqrt[5]{n}$ then theorem 1 implies that for $q = 1$ in the θ case,

$$n^{2/5}(\hat{\boldsymbol{\theta}} - \theta_0) \xrightarrow{d} N\left\{\frac{\int\{D_{\pi 1}(t) + \pi_0 D_{Q3}(t)\}t\phi_V(t)dt}{c_\alpha^4 \pi_0}, c_\alpha^2 \mathscr{V}\right\}. \tag{11}$$

The rate in (11) is the familiar nonparametric rate, which is no coincidence. Having chosen a convergence rate, the choice of input parameter is reduced to a choice of c_α, which can be made to minimize the asymptotic mean square error; data–dependent input parameter choices are discussed in Jun, Pinkse, and Wan (2011a). The main point of this paper, however, is rate–adaptive inference for a given value of the input parameter since in practice one chooses a value rather than a rate.

One way of removing the bias in (11) is to choose a prior for which the mean in (11) equals zero, for which it is required that

$$\partial_\theta \log \pi(\theta_0) = -\left\{\int tt^\intercal \phi_V(t)dt\right\}^{-1} \int D_{Q3}(t)t\phi_V(t)dt. \tag{12}$$

One way of satisfying (12) is for any ‘mother prior’ π^M with $\partial_\theta \pi^M(\theta_0) = 0$ to pick a matrix $\mathbb{A}$ such that (12) is satisfied for $\pi(\theta) \propto \pi^M(\mathbb{A}\theta)$. Alternatively, one can choose $\pi \propto \sqrt{\det(-\partial_{\theta\theta^\intercal} Q)}$ in a neighborhood of θ_0, which resembles the Jeffreys prior; see lemma F.1. Since $\partial_{\theta\theta^\intercal} Q$ is unknown, it must be estimated. Theorem 2 demonstrates addresses the issue of estimated priors for the case $q = 1$.

THEOREM 2. *Consider the θ–only case, i.e. $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}$ does not have an η argument. Suppose that assumptions A through F are satisfied for $q = 1$, that assumption G is satisfied for $\tilde{\alpha}_n = \min(\alpha_n, \sqrt[3]{n})$,* *and that* $1/\alpha_n = O(n^{-1/5})$. *Suppose moreover that* $\hat{\hat{\boldsymbol{\theta}}}$ *is identical to* $\hat{\boldsymbol{\theta}}$ *except that it uses a data–dependent prior* $\hat{\boldsymbol{\pi}}$ *in lieu of* π *and (i)* $\forall\theta \notin \Theta : \hat{\boldsymbol{\pi}}(\theta) = 0$*; (ii)* $\partial_\theta \hat{\boldsymbol{\pi}}$ *is continuous on* Θ *with probability approaching one; (iii) for some* $0 < \underline{\pi} < \bar{\pi} < \infty$, $\mathbb{P}\{\inf_{\theta\in\Theta} \hat{\boldsymbol{\pi}}(\theta) < \underline{\pi}\} = o(1)$ *and* $\mathbb{P}\{\sup_{\theta\in\Theta} \hat{\boldsymbol{\pi}}(\theta) > \bar{\pi}\} = o(1)$*; (iv) for some* $\bar{\pi}_1 < \infty$, $\mathbb{P}\{\sup_{\theta\in\Theta} \|\partial_\theta \hat{\boldsymbol{\pi}}(\theta)\| > \bar{\pi}_1\} = o(1)$*; (v)* $\hat{\boldsymbol{\pi}}(\theta_0) - \pi(\theta_0) = o_p(1)$*; (vi) for any* $c > 0$, $\lim_{t\downarrow 0} \lim_{n\to\infty} \mathbb{P}\big\{\sup_{\|\theta-\theta_0\|\le t} \|\partial_\theta \hat{\boldsymbol{\pi}}(\theta) - \partial_\theta \pi(\theta)\| > c\big\} = 0$. *Then*

$$\max(\sqrt{n/\alpha_n}, \sqrt[3]{n})(\hat{\hat{\boldsymbol{\theta}}} - \hat{\boldsymbol{\theta}}) = o_p(1).$$

Proof. See appendix F. □

If α_n diverges faster than $\sqrt[5]{n}$ then there is no asymptotic bias for any prior satisfying the conditions of theorem 1. So the most interesting case in theorem 2 arises when $\alpha_n = c_\alpha^2 \sqrt[5]{n}$. Indeed, then theorem 2 implies that $\hat{\hat{\boldsymbol{\theta}}} - \hat{\boldsymbol{\theta}} = o_p(n^{-2/5})$ such that if π is such that the asymptotic mean in (11) equals zero then

$$n^{2/5}(\hat{\hat{\boldsymbol{\theta}}} - \theta_0) \xrightarrow{d} N(0, c_\alpha^2 \mathscr{V}).$$

An attractive feature of theorem 2 is that the bias correction procedure does not change the asymptotic variance.

The conditions on $\hat{\boldsymbol{\pi}}$ in theorem 2 are mild and since $\hat{\boldsymbol{\pi}}$ can be chosen they are straightforward to satisfy with the possible exception of the uniform convergence of $\partial_\theta \hat{\boldsymbol{\pi}}$ in a neighborhood of θ_0. We chose not to provide the analogous result for the η class of estimators in theorem 2 to conserve space.

3. Eta class. We now turn our attention to the η case, in which θ_0 is chosen and η_0 satisfying $\theta_0(\eta_0) = \theta_0$ is the parameter of interest. The convergence results for $\hat{\boldsymbol{\eta}}$ are similar to those for $\hat{\boldsymbol{\theta}}$, but they do require additional assumptions.

Assumption.

- H. *the absolute values of the eigenvalues of* $\partial_{\theta\theta^\intercal} \tilde{Q}(\theta, \eta)$ *are no greater than* $\lambda^+(\theta, \eta)$ *which is uniformly bounded on* $\Theta \times N$*;*
- I. *for all* $\eta \in N$, $\tilde{Q}$ *is quasi–concave in* θ*, i.e. its sublevel sets are convex;*
- J. *the matrix* $\mathscr{C} = V^{-1}\partial_{\theta\eta^\intercal} \tilde{Q}(\theta_0, \eta_0)$ *is nonsingular;*
- K. *for all* $\eta \in N$, $\partial_\eta \tilde{Q}(\cdot, \eta)$ *is bounded and nondecreasing in every element of* θ.
- L. *for all* $\eta \neq \eta_0$ *in* N, $\theta_0 \notin \Theta_0(\eta)$*;*
- M. $\tilde{g}(\xi; \theta, \eta)$ *is continuous in* η. □

Assumption H is implied by continuity of $\partial_{\theta\theta^\intercal} Q$ and assumption I is weaker than concavity. Assumptions J and L are imposed to ensure identification of η_0. Assumption K removes the need to achieve uniformity in w in theorem 1. When θ and η are scalar–valued, assumption L is implied by assumptions J and K. See section 5 for a justification and examples.

Theorem 3. *Let assumptions A through J be satisfied. Then* *(i)* $\hat{\boldsymbol{\eta}} = \eta_0 + o_p(1)$;

(ii) $\hat{\boldsymbol{\eta}} = \eta_0 + O_p\big(n^{-1/3}\big)$;

(iii) *if the conditions of part (i) of theorem 1 are satisfied then*

$$\sqrt[3]{n}(\hat{\boldsymbol{\eta}} - \eta_0) \xrightarrow{d} \frac{1}{c_\alpha^2}\mathscr{C}^{-1}\frac{\int t \exp\{c_\alpha^3 \mathbb{G}(t)\}\phi_V(t)dt}{\int \exp\{c_\alpha^3 \mathbb{G}(t)\}\phi_V(t)dt};$$

(iv) *if the conditions of part (ii) of theorem 1 are satisfied then*

$$\sqrt[3]{n}(\hat{\boldsymbol{\eta}} - \eta_0) \xrightarrow{d} \mathscr{C}^{-1} \operatorname*{argmax}_{t\in\mathbb{R}^d} \mathbb{C}(t);$$

(v) *if the conditions of parts (iii) and (iv) of theorem 1 are satisfied and if $\tilde{\boldsymbol{\eta}}$ solves*

$$\hat{\boldsymbol{\theta}}(\tilde{\boldsymbol{\eta}}) - \theta_0 - \sqrt{\frac{\alpha_n}{n}}\frac{C_V \sum_{\tau=0}^{q}\big(b_{q\tau}^*/\alpha_n^\tau\big) + o_p(\alpha_n^{-q})}{\beta_n \alpha_n^d \int \pi(\theta)\exp\{\alpha_n^2 \boldsymbol{L_n}(\theta, \tilde{\boldsymbol{\eta}})\}d\theta} = 0, \tag{13}$$

then

$$\sqrt{n/\alpha_n}(\tilde{\boldsymbol{\eta}} - \eta_0) \xrightarrow{d} N\big\{0, \mathscr{C}^{-1}\mathscr{V}(\mathscr{C}^{-1})^\mathsf{T}\big\}.$$

Proof. See appendix G. □

So the results for $\hat{\boldsymbol{\eta}}$ (and its bias–reduced version $\tilde{\boldsymbol{\eta}}$) are thus largely the same as for $\hat{\boldsymbol{\theta}}$, albeit subject to additional assumptions. The extra $\mathscr{C}^{-1}$ matrix in theorem 3 are due to the fact that we are essentially estimating an inverse function here.

Our previous comments on the bias reduction of $\hat{\boldsymbol{\theta}}$ also apply to $\tilde{\boldsymbol{\eta}}$. For instance, for $q = 1$, if $\alpha_n = c_\alpha \sqrt[5]{n}$ then we only need to adjust the first order bias. If the Jeffreys prior is used, then the first order bias becomes zero, so $\tilde{\boldsymbol{\eta}}$ will simply solve $\hat{\boldsymbol{\theta}}(\tilde{\boldsymbol{\eta}}) = \theta_0$ as before.

4. Inference.

We now discuss our rate–adaptive inference procedure. We focus on the θ case with no (or fixed) η parameter, because it best highlights the idea with the least number of restrictions. We start by assuming the availability of consistent estimators for H and V.

Assumption.

N. $\hat{\boldsymbol{V}} = V + o_p(1)$;

O. $\forall t, s \in \mathbb{R}^d$, $\hat{\boldsymbol{H}}(t,s) = H(t,s) + o_p(1)$;

P. $\forall t, s \in \mathbb{R}^d$ *and* $c > 0$, $\hat{\boldsymbol{H}}(t-s, t-s) = \hat{\boldsymbol{H}}(t,t) - \hat{\boldsymbol{H}}(s,s) - 2\hat{\boldsymbol{H}}(t,s)$ *and* $\hat{\boldsymbol{H}}(ct, cs) = c\hat{\boldsymbol{H}}(t,s)$;

Q. $\mathbb{E}\big\{\sup_{\|t\|=1} \hat{\boldsymbol{H}}^{\boldsymbol{2}}(t,t)\big\} = O(1)$.

Assumptions N to Q are high–level, but they are satisfied by common estimators; assumption P is the sample counterpart of (6). Letting $\{\hat{\mathbb{G}}\}$ be a (sample–size–dependent) sequence of Gaussian processes with finite marginal distributions characterized by $\hat{\boldsymbol{H}}$, assumption O ensures that $\hat{\mathbb{G}}$ has the same marginal distributions as $\mathbb{G}$ in the limit and assumption P is needed to make $\exp\{\hat{\mathbb{G}}(\cdot)\}$ integrable with respect to a normal density with probability one. It is innocuous in that covariance kernel estimates satisfying assumption P are straightforward to construct. Assumption Q is needed for the weak convergence of $\hat{\mathbb{G}}$ to $\mathbb{G}$ in an $\mathbb{L}^2$ space, which is sufficient for convergence in distribution of (normal density–weighted) integrals of $\exp\{\hat{\mathbb{G}}(\cdot)\}$ to the corresponding integrals of $\exp\{\mathbb{G}(\cdot)\}$.

Let

$$\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}} = \frac{1}{\min\left(\sqrt[3]{\beta_n}, 1\right)} \frac{\int t \exp\left[\beta_n^{4/3}\{\hat{\mathbb{G}}(t) - t^\intercal \hat{V} t/2\}\right] dt}{\int \exp\left[\beta_n^{4/3}\{\hat{\mathbb{G}}(t) - t^\intercal \hat{V} t/2\}\right] dt},$$

which we can simulate with a given value of α_n and data. We now have the following theorem.

THEOREM 4. *Under assumptions N to Q, $\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}} \xrightarrow{d} \mathbb{J}_{\boldsymbol{\alpha}}$, where the distribution of $\mathbb{J}_{\boldsymbol{\alpha}}$ is equal to the limit distribution given in parts (i), (ii), or (iii) of theorem 1, depending on whether $\alpha_n/\sqrt[3]{n} \to c_\alpha^2 > 0$, $\sqrt[3]{n}/\alpha_n \to 0$, or $\alpha_n/\sqrt[3]{n} \to 0$, respectively.*

Theorem 4 shows that the limiting distribution of $\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}}$ automatically adapts to the rate of α_n. So, for all $w \in \mathscr{W}$, $\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}}$ approximates the distribution of $\max\left(\sqrt{n/\alpha_n}, \sqrt[3]{n}\right)\left\{\hat{\boldsymbol{\theta}}(w^*, w) - \theta_{0n}(w)\right\}$ uniformly in w^*, up to a "bias" term in the normality case. Therefore, using the limiting distributions of theorem 1 for inference does not require to translate a chosen value of α_n to a specific rate condition on it, which removes rate–related uncertainty encountered in practice.

5. Applications. We now provide an incomplete list of applications to which our results can be applied; further applications can be found in Groeneboom and Wellner (2001); Kim and Pollard (1990), albeit that for LMS Jun, Pinkse, and Wan (2011b) already provides a $\sqrt{n}$–consistent estimator.

We first list the applications with the formulas for $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}, H, V$ in each case before we verify our assumptions for a few representative cases. Please note that in a few instances, $\tilde{Q}$ is decreasing rather than increasing in η, but only the monotonicity in η is germane; η can always be replaced with $-\eta$.

[a] Inverse density: The object of interest is the inverse density function value $\theta_0 = f^{-1}(\eta_0)$ for specified η_0 where f is a (weakly) decreasing density function for which f' is continuous and negative at θ_0. Then one can use $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta) = \mathbb{1}(\boldsymbol{x}_{\boldsymbol{i}} \le \theta) - \theta\eta_0$, which produces $H(t,s) = \left|\mathrm{Med}(t,s,0)\right| f(\theta_0)$ and $V = -f'(\theta_0)$.

[b] Inverse regression: The object of interest is the inverse regression function value $\theta_0 = \mathfrak{m}^{-1}(\eta_0)$ for specified η_0 where $\mathfrak{m}$ with $\mathfrak{m}(x) = \mathbb{E}(\boldsymbol{y}_{\boldsymbol{1}}|\boldsymbol{x}_{\boldsymbol{1}} = x)$ is a (weakly) decreasing regression function of a continuously distributed regressor $\boldsymbol{x}_{\boldsymbol{i}}$ for which $\mathfrak{m}'$ is negative and continuous at θ_0. One can use $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta) = \mathbb{1}(\boldsymbol{x}_{\boldsymbol{i}} \le \theta)(\boldsymbol{y}_{\boldsymbol{i}} - \eta_0)$, which results in $H(t,s) = \left|\mathrm{Med}(t,s,0)\right| f(\theta_0)\mathbb{V}(\boldsymbol{y}_{\boldsymbol{1}}|\boldsymbol{x}_{\boldsymbol{1}} = \theta_0)$ and $V = -\mathfrak{m}'(\theta_0)f(\theta_0)$.

[c] Inverse monotone hazard: Observed are durations $\boldsymbol{x}_{\boldsymbol{i}}$ drawn from a distribution F for which the hazard rate $\mathscr{H} = f/(1-F)$ is (weakly) decreasing and strictly decreasing at the value of

interest $\theta_0 = \mathscr{H}^{-1}(\eta_0)$. Take $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta) = \mathbb{1}(\boldsymbol{x_i} \le \theta) - \eta_0 \min(\boldsymbol{x_i}, \theta)$, which yields $H(t,s) = |\mathrm{Med}(t,s,0)| f(\theta_0)$ and $V = -f'(\theta_0) - f^2(\theta_0)/\{1 - F(\theta_0)\}$.

[d] Quantiles with interval censoring: Observed are (continuously distributed) times $\boldsymbol{y_i}$ with distribution function $F_{\boldsymbol{y}}$ and the indicator $\mathbb{1}\{\boldsymbol{y_i} \ge \boldsymbol{x_i}\}$; $\boldsymbol{x_i}$ is continuous and independent of $\boldsymbol{y_i}$, but its value is not observed. The object of interest is the quantile $\theta_0 = F_{\boldsymbol{x}}^{-1}(\eta_0)$ of $\boldsymbol{x_1}$ at some specified value of η_0. Take $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta) = \mathbb{1}(\boldsymbol{y_i} \le \theta)\eta_0 - \mathbb{1}\big\{\mathbb{1}(\boldsymbol{x_i} \le \boldsymbol{y_i})\boldsymbol{y_i} \le \theta\big\}$, which results in $H(t,s) = |\mathrm{Med}(t,s,0)| F_{\boldsymbol{x}}(\theta_0)\{1 - F_{\boldsymbol{x}}(\theta_0)\} f_{\boldsymbol{y}}(\theta_0)$ and $V = f_{\boldsymbol{y}}(\theta_0) f_{\boldsymbol{x}}(\theta_0)$, where $f_{\boldsymbol{x}}, f_{\boldsymbol{y}}$ are the density functions corresponding to $F_{\boldsymbol{y}}, F_{\boldsymbol{x}}$, respectively.

[e] Mode regression (Lee, 1989): Observations $\boldsymbol{y_i} = \theta_0^\top \boldsymbol{x_i} + \boldsymbol{u_i}$ are available only when $\boldsymbol{y_i} \ge 0$, at least one of the regressors is continuous with nonzero coefficient, and the conditional distribution of $\boldsymbol{u_i}$ given $\boldsymbol{x_i}$ is assumed continuous and even, with the conditional density strictly decreasing at zero and weakly decreasing elsewhere. For given input parameter $\nu > 0$, let $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta) = \mathbb{1}\big\{|\boldsymbol{y_i} - \max(\theta^\top \boldsymbol{x_i}, \nu)| \le \nu\big\}$, resulting in $H(t,s) = 2\mathbb{E}\big\{\mathbb{1}(\theta_0^\top \boldsymbol{x_1} \ge \nu)\big|\mathrm{Med}(t^\top \boldsymbol{x_1}, s^\top \boldsymbol{x_1}, 0)\big| f_{\boldsymbol{u}|\boldsymbol{x}}(\nu|\boldsymbol{x_1})\big\}$ and $V = -2\mathbb{E}\big\{\mathbb{1}(\theta_0^\top \boldsymbol{x_1} \ge \nu)\boldsymbol{x_1}\boldsymbol{x_1}^\top f'_{\boldsymbol{u}|\boldsymbol{x}}(\nu|\boldsymbol{x_1})\big\}$. As a special case one can estimate the mode of a distribution by replacing $\boldsymbol{x_i}$ with a constant.

[f] Maximum score (Manski, 1975): The estimator is intended for the regression model with binary regressand $\boldsymbol{y_i}$ and regressors $\boldsymbol{x_i}$ under the assumption that $\mathrm{Med}(\boldsymbol{y_1}|\boldsymbol{x_1}) = \mathbb{1}(\tilde{\theta}_0^\top \boldsymbol{x_1} \ge 0)$ a.s., and where the regressor vector can be partitioned as $\boldsymbol{x_i} = [\boldsymbol{z}_{\boldsymbol{i}}^\top, \boldsymbol{a_i}]^\top$ for a continuous regressor $\boldsymbol{a_i}$, whose regression coefficient can be normalized to (plus or) minus one. Let $\tilde{\theta}_0 = [\theta_0^\top, -1]^\top$ and take $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta) = (2\boldsymbol{y_i} - 1)\mathbb{1}(\theta^\top \boldsymbol{z_i} - \boldsymbol{a_i} \ge 0)$, resulting in $H(t,s) = \mathbb{E}\big\{\mathrm{Med}(t^\top \boldsymbol{z_1}, s^\top \boldsymbol{z_1}, 0) f(\theta_0^\top \boldsymbol{z_1}|\boldsymbol{z_1})\big\}$ and $V = \mathbb{E}\big\{\boldsymbol{z_1}\boldsymbol{z_1}^\top \partial_a \mathbb{P}(\boldsymbol{y_1} = 1|\boldsymbol{a_1} = \theta_0^\top \boldsymbol{z_1}, \boldsymbol{z_1}) f(\theta_0^\top \boldsymbol{z_1}|\boldsymbol{z_1})\big\}$, where f is the conditional density of $\boldsymbol{a_1}$ given $\boldsymbol{z_1}$. The MS estimator is related to the *perceptron* (Rosenblatt, 1957) used in artificial neural networks. Other variations and extensions of the MS estimator include a multinomial version (in Manski, 1975) and panel data model with "fixed effects" (Manski, 1987).

[g] Isotonic density (Grenander, 1956): The object of interest is the density function value $\eta_0 = f(\theta_0)$ for chosen $\theta_0 > 0$ where f is a (weakly) decreasing density function for which $f'(\theta_0) < 0$. Then one can use $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta, \eta) = \mathbb{1}(\boldsymbol{x_i} \le \theta) - \theta\eta$. H and V are as in [a]. The formulation for $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}$ here is inspired by the characterization of the Grenander nonparametric maximum likelihood estimator (NPMLE) in Groeneboom (1985).

[h] Isotonic regression (Brunk, 1958, 1968): To be estimated is $\eta_0 = \mathfrak{m}(\theta_0)$ for chosen $\theta_0 > 0$ where m is as in [b]. One can use $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta, \eta) = \mathbb{1}(\boldsymbol{x_i} \le \theta)(\boldsymbol{y_i} - \eta)$. H, V are as in [b]. There is again a direct analogy to the corresponding NPMLE.

[i] Monotone hazard rate (Prakasa Rao, 1970): The objective is to estimate $\eta_0 = \mathscr{H}(\theta_0)$ where $\mathscr{H}, \boldsymbol{x_i}$ are as in [c]. Take $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta, \eta) = \mathbb{1}(\boldsymbol{x_i} \le \theta) - \eta \min(\boldsymbol{x_i}, \theta)$; H, V are is in [c].

[j] Interval censoring (Ayer, Brunk, Ewing, Reid, and Silverman, 1955): Using the notation of [d], the object of interest is $\eta_0 = F_{\boldsymbol{x}}(\theta_0)$ at some speicifed value θ_0. Take $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}(\theta, \eta) = \mathbb{1}(\boldsymbol{y_i} \le \theta)\eta - \mathbb{1}\big\{\mathbb{1}(\boldsymbol{x_i} \le \boldsymbol{y_i})\boldsymbol{y_i} \le \theta\big\}$; H, V are as in [d].

To illustrate our assumptions consider example [g]. Example [g] is similar to [h] and indeed [i] and [j] and will yield intuition for [a] to [d] as a byproduct. Example [f] is discussed in depth in Jun, Pinkse, and Wan (2011a).

For [g], $\tilde{Q}(\theta,\eta) = F(\theta) - \theta\eta$. In a neighborhood $\mathfrak{N}_0$ of η_0, $\theta_0(\eta) = f^{-1}(\eta)$ and since f' is assumed continuous and negative at θ_0, assumption B is satisfied. In assumption C, which is also satisfied, $q \geq 0$ corresponds to the number of derivatives f' posseses at θ_0. Further, for assumption G, note that $\mathscr{F}_{nw}$ consists of $g^{\circ}_{nw}(x;t,w^*) = \sqrt{\tilde{\alpha}_n}\{\mathbb{1}(\theta_n < x \leq \theta_n + t/\tilde{\alpha}_n) - (\eta_0 + w^*/\mathrm{r}_n)t/\tilde{\alpha}_n\}/c_t$, where $t \in \mathbb{R}$, $w^* \in \mathscr{W}$, and θ_n converges to θ_0. Therefore, $\mathscr{F}_{nw}$ is contained in a convex hull of the two classes $\mathscr{F}^1_{nw} = \{\sqrt{\tilde{\alpha}_n}\mathbb{1}\{\theta_n < x \leq \theta_n + t/\tilde{\alpha}_n\}/c_t : t \in \mathbb{R}\}$ and $\mathscr{F}^2_{nw} = \{(\eta_0 + w^*/\mathrm{r}_n)(t/c_t)/\sqrt{\tilde{\alpha}_n} : t \in \mathbb{R}, w^* \in \mathscr{W}\}$. Both $\mathscr{F}^1_{nw}$ and $\mathscr{F}^1_{nw}$ are polynomial classes with polynomial indices independent of n with well–behaved envelope functions $\sqrt{\tilde{\alpha}_n}/\big(1 + \tilde{\alpha}_n|x - \theta_n|\big)$ and $(|\eta_0| + \bar{w}/\mathrm{r}_n)/\sqrt{\tilde{\alpha}_n}$, respectively, where $\bar{w} = \sup \mathscr{W}$. In sum, the assumptions for theorem 1 are satisfied.

Assumption H requires the integration region to be restricted to an area in which f' is bounded; assumption I is implied by the assumed concavity of F (f' is decreasing). $\mathscr{C} = -1/f'(\theta_0)$, which is nonzero, $\partial_\eta Q(\theta) = \theta$ is bounded and nondecreasing, and $\tilde{g}$ is continuous in η, so assumptions J, K, and M hold. Assumption L follows from assumptions J and K, because θ and η are scalars.

If $q \geq 0$ then part (iii) of theorem 3 implies that if $\sqrt[3]{n}/\alpha_n = o(1)$ then our estimator produces the same limiting distribution as the Grenander estimator under the same conditions (Groeneboom, 1985) and if $\alpha_n = c_\alpha^2 \sqrt[3]{n}$ then it produces a $\sqrt[3]{n}$–rate but a different limiting distribution.

If $q \geq 1$ and $\alpha_n = c_\alpha^2 \sqrt[5]{n}$ then by (11) and theorem 3,

$$n^{2/5}\{\hat{f}(\theta_0) - f(\theta_0)\} \xrightarrow{d} N\left[\frac{\pi'(\theta_0) - \pi(\theta_0)f''(\theta_0)/2f'(\theta_0)}{c_\alpha^4 \pi(\theta_0)}, \frac{c_\alpha^2 f(\theta_0)\sqrt{-f'(\theta_0)}}{2\sqrt{\varpi}}\right], \tag{14}$$

where $\varpi = 3.14159....$ For flat π the assumptions needed for and asymptotic bias and variance of the proposed estimator are the same as those of a kernel density estimator using a normal kernel and bandwidth equal to $1/\sqrt[5]{n}\sqrt{-c_\alpha^4 f'(\theta_0)}$. For the Jeffreys–like prior $\pi(\theta) \propto \sqrt{-f'(\theta)}$ the asymptotic bias of our isotonic density estimator is zero.

If α_n is chosen 'too large' then the limit distribution in (14) will not be a good approximation for $\hat{f}$. Since it is hard to know what 'too large' means in this context, the inference procedure of theorem 4 is generally preferable.

References.

ABREVAYA, J., AND J. HUANG (2005): "On the bootstrap of the maximum score estimator," *Econometrica*, 73(4), 1175–1204.

AYER, M., H. BRUNK, G. EWING, W. REID, AND E. SILVERMAN (1955): "An empirical distribution function for sampling with incomplete information," *The Annals of Mathematical Statistics*, pp. 641–647.

BARLOW, R. E., AND W. R. VAN ZWET (1969): "Asymptotic properties of isotonic estimators for the generalized failure rate function. Part II: Asymptotic distributions," *Operations Research Center Report ORC 69-10, Univ. California*.

——— (1970): "Asymptotic properties of isotonic estimators for the generalized failure rate function. Part I: Asymptotic distributions," in *Nonparametric Techniques in Statistical Inference*, ed. by M. L. Puri, chap. 2, pp. 159–173. Wiley, New York.

——— (1971): "Comparison of several nonparametric estimators of the failure rate function," in *Operations Research and Reliability*, chap. 2, pp. 375–399. Gordon and Breach, New York.

BERGE, C. (1963): *Topological spaces*. Courier Dover.

BRUNK, H. (1958): "On the estimation of parameters restricted by inequalities," *The Annals of Mathematical Statistics*, pp. 437–454.

——— (1968): "Estimation of isotonic regression. Nonparametric techniques," *Statistical Inference*, pp. 177–195.

Cattaneo, M., R. Crump, and M. Jansson (2011): "Small bandwidth asymptotics for density-weighted average derivatives," *Econometric Theory*, Forthcoming.

Cavanagh, C. (1987): "The limiting behavior of estimators defined by optimization," Discussion paper, Harvard University.

Chernozhukov, V., and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.

Cremers, H., and D. Kadelka (1986): "On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in L_p^E," *Stochastic Processes and their Applications*, 21, 305–317.

Delgado, M., J. Rodriguez-Poo, and M. Wolf (2001): "Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator," *Economics Letters*, 73(2), 241–250.

Friedman, J., and R. Tibshirani (1984): "The monotone smoothing of scatterplots," *Technometrics*, 26(3), 243–250.

Grenander, U. (1956): "On the theory of mortality measurement, Part II," *Skandinavisk Aktuarietidskrift*, 39, 125–153.

Groeneboom, P. (1985): "Estimating a monotone density," in *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, II.*, ed. by L. M. L. Cam, and R. A. Olshen, pp. 535–555. Springer.

Groeneboom, P., and J. A. Wellner (2001): "Computing Chernoff's distribution," *Journal of Computational and Graphical Statistics*, 10(2), 388–400.

Horowitz, J. (1992): "A smoothed maximum score estimator for the binary response model," *Econometrica*, 60(3), 505–31.

Jeffreys, H. (1946): "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, pp. 453–461.

Jun, S., J. Pinkse, and Y. Wan (2011a): "Laplace estimation for maximum score," Discussion paper, Penn State.

——— (2011b): "$\sqrt{n}$–consistent robust integration-based estimation," *Journal of Multivariate Analysis*, 102, 828–846.

Kim, J., and D. Pollard (1990): "Cube root asymptotics," *Annals of Statistics*, 18(1), 191–219.

Lee, M. (1989): "Mode regression," *Journal of Econometrics*, 42(3), 337–349.

Mammen, E. (1991): "Estimating a smooth monotone regression function," *The Annals of Statistics*, 19(2), 724–740.

Manski, C. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3(3), 205–228.

——— (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica*, pp. 357–362.

Mukerjee, H. (1988): "Monotone nonparametric regression," *The Annals of Statistics*, 16(2), 741–750.

Pollard, D. (1993): "The asymptotics of a binary choice model," Discussion paper, Yale University.

Prakasa Rao, B. (1970): "Estimation for distributions with monotone failure rate," *The Annals of Mathematical Statistics*, pp. 507–519.

Rosenblatt, F. (1957): "The perceptron, a perceiving and recognizing automaton," Discussion paper, Cornell Aeronautical Laboratory.

Sen, B., M. Banerjee, and M. Woodroofe (2010): "Inconsistency of bootstrap: The Grenander estimator," *The Annals of Statistics*, 38(4), 1953–1977.

Steele, J. (2004): *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*. Cambridge Univ Pr.

van der Vaart, A., and J. Wellner (1996): *Weak convergence and empirical processes*. Springer.

Wright, F. T. (1982): "Monotone regression estimates for grouped observations," *The Annals of Statistics*, 10(1), 278–286.

Zinde-Walsh, V. (2002): "Asymptotic theory for some high breakdown point estimators," *Econometric theory*, 18(05), 1172–1196.

APPENDIX A: NOTATION

Throughout $\{\gamma_n\}$ is a sequence which diverges at a polynomial rate such that $\gamma_n = o\{\tilde{\alpha}_n^{1/(3(q+1))}\}$. Further, $\mathfrak{T}_n = \{t \in \mathbb{R}^d : \theta_0 + t/\tilde{\alpha}_n \in \Theta\}$, $\Gamma_n = \{t \in \mathbb{R}^d : \|t\| \le \gamma_n\}$, $\Gamma_n^c = \mathbb{R}^d - \Gamma_n$, and $\Gamma_n^{c*} = \mathfrak{T}_n - \Gamma_n$. Further, $R_{nw}(t) = Q_{nw}(t) + t^\intercal V t/2$. Additional notation will be introduced in the lemmas and proofs in the appendices below.

APPENDIX B: PART (i) OF THEOREM 1

LEMMA B.1. *For* $\mathscr{C} = V^{-1}\partial_{\theta\eta^\intercal} Q(\theta_0, \eta_0)$, $\mathfrak{r}_n\{\theta_{0n}(w) - \theta_0\} = \mathscr{C}w + o(1)$.

PROOF. Since Q is concave in θ at (θ_0, η_0) and continuous in both θ, η, it follows that $\theta_{0n}(w) - \theta_0 = o(1)$. Thus, by multiple use of the mean value theorem and using assumption C it follows that

$$0 = \partial_\theta Q\{\theta_{0n}(w), \eta_0 + w/\mathfrak{r}_n\} = \partial_{\theta\eta^\intercal} Q(\theta_0, \eta_0)w/\mathfrak{r}_n - V\{\theta_{0n}(w) - \theta_0\} + o\big\{\|\theta_{0n}(w) - \theta_0\| + 1/\mathfrak{r}_n\big\}. \quad \square$$

LEMMA B.2. $\tilde{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t, w^*) = \sum_{i=1}^n \{\boldsymbol{g}^\circ_{\boldsymbol{nwi}}(t, w^*) - \mathbb{E}\boldsymbol{g}^\circ_{\boldsymbol{nw1}}(t, w^*)\}/\sqrt{n} \overset{w}{\to} \mathbb{G}^*(t)$ *in* $\mathbb{L}^\infty(\Re^d \times \mathscr{W})$, *where* $\mathbb{G}^*$ *is a Gaussian process with the covariance kernel* $H^*(t, s) = H(t, s)/c_t c_s$. *Consequently,* $\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*) = \tilde{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t, w^*)c_t \overset{w}{\to} \mathbb{G}(t)$ *in* $\mathbb{L}^\infty_L(\Re^d \times \mathscr{W})$, *where* $\mathbb{L}^\infty_L$ *is a space of locally bounded functions on compacta.*

PROOF. Since $\Re^d \times \mathscr{W}$ is dense, assumption F ensures that for $j = 1, 2$, $\big\{\{g^\circ_{nw}(\cdot; t, w^*) - g^\circ(\cdot; s, \tilde{w}^*)\}^j\big\}_{\|(t,w^*)-(s,\tilde{w}^*)\|<\epsilon}$ is a pointwise measurable class, and hence $\mathscr{P}$–measurable for every $\mathscr{P}$; see van der Vaart and Wellner (1996, p.110). Note further that $\lim_{n\to\infty} \mathbb{E}\big\{\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(s, \tilde{w}^*)\big\} = H^*(t, s)$ for every $t, s \in \Re^d$ and every $w^*, \tilde{w}^* \in \mathscr{W}$ by assumption E. Therefore, the result follows from assumption G and van der Vaart and Wellner (1996, theorem 2.11.22). $\square$

LEMMA B.3. *For any* $|c| \le 1$, *any* b, *and any integer* $j \ge 0$, $\big|\exp(cb) - \sum_{s=0}^j (cb)^s/s!\big| \le |c|^{j+1}\exp(|b|)$.

PROOF. We have $\big|\exp(cb) - \sum_{s=0}^j (cb)^s/s!\big| \le \big|\sum_{s=j+1}^\infty (cb)^s/s!\big| \le |c|^{j+1} \sum_{s=j+1}^\infty |b|^s/s! \le |c|^{j+1}\exp(|b|)$. $\square$

LEMMA B.4. *For some* $0 < c_q < \infty$, *all sufficiently large* n, *all* t *for which* $\theta_{0n}(w) + t/\tilde{\alpha}_n \in \Theta$, *and all* $w \in \mathscr{W}$, $Q_{nw}(t) \le -\min(c_q\tilde{\alpha}_n^2, t^\intercal Vt/4)$.

PROOF. Choose neighborhoods $\mathfrak{N}_\eta, \mathfrak{N}_\theta$ of η_0, θ_0, such that for all $\eta \in \mathfrak{N}_\eta$ and some $0 < c_q < \infty$, *(i)* $\theta \in \mathfrak{N}_\theta \Rightarrow Q(\theta, \eta) \le -\{\theta - \theta_0(\eta)\}^\intercal V\{\theta - \theta_0(\eta)\}/4$, *(ii)* $\theta \notin \mathfrak{N}_\theta \Rightarrow Q(\theta, \eta) \le -c_q$. Such neighborhoods exist because of assumptions B and C. Take n large enough to ensure that $\eta_0 + w/\mathfrak{r}_n \in \mathfrak{N}_\eta$. Then, for all $\theta \in \Theta$, $Q(\theta, \eta_0 + w/\mathfrak{r}_n) \le -\min\big[\{\theta - \theta_{0n}(w)\}^\intercal V\{\theta - \theta_{0n}(w)\}/4, c_q\big]$, which implies the stated result. $\square$

LEMMA B.5. *For* $\bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t) = \sup_{w^*\in\mathscr{W}} |\tilde{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t, w^*)|$, *(i)* $\sup_{t\in\Gamma_n^c} \bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t) = o_p(1)$ *and (ii) for all* $\epsilon > 0$, $\lim_{\tilde{\gamma}\to\infty}\lim_{n\to\infty} \mathbb{P}\big\{\sup_{t\in\tilde{\Gamma}^c} \bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t) > \epsilon\big\} = 0$, *where* $\tilde{\Gamma} = \big\{t \in \mathbb{R}^d : \|t\| \le \tilde{\gamma}\big\}$.

PROOF. We show (i); (ii) is similar. For any $\epsilon > 0$ and any nonnegative integer j by B.2, the continuous mapping theorem, using the fact that $\mathbb{G}(st)$ and $\sqrt{s}\mathbb{G}(t)$ have the same distribution by assumption E, and by the Markov inequality, we have

$$(15)\quad \lim_{n\to\infty}\mathbb{P}\Big\{\Big|\sup_{\|t\|\geq\gamma_n}\bar{S}^*_{\boldsymbol{nw}}(t)\Big|>\epsilon\Big\}\leq\lim_{n\to\infty}\mathbb{P}\Big\{\sup_{\|t\|\geq j}|\bar{S}^*_{\boldsymbol{nw}}(t)|>\epsilon\Big\}=\mathbb{P}\Big\{\sup_{\|t\|\geq j}|\mathbb{G}^*(t)|>\epsilon\Big\}$$
$$=\sum_{s=j+1}^{\infty}\mathbb{P}\Big\{\sup_{s-1\leq\|t\|<s}|\mathbb{G}^*(t)|>\epsilon\Big\}\leq\sum_{s=j+1}^{\infty}\mathbb{P}\Big\{\sup_{s-1\leq\|t\|<s}|\mathbb{G}(t)|>\epsilon s\Big\}$$
$$\leq\sum_{s=j+1}^{\infty}\mathbb{P}\Big\{\sup_{\|t\|\leq 1}|\mathbb{G}(t)|>\epsilon\sqrt{s}\Big\}\leq\frac{\mathbb{E}\sup_{\|t\|\leq 1}\mathbb{G}^4(t)}{\epsilon^4}\sum_{s=j+1}^{\infty}\frac{1}{s^2}.$$

By van der Vaart and Wellner (1996, proposition A.2.4), we know that there exists some $C<\infty$ such that $\mathbb{E}\sup_{\|t\|\leq 1}\mathbb{G}^4(t)\leq C\{\mathbb{E}\sup_{\|t\|\leq 1}|\mathbb{G}(t)|\}^4$, which is finite by van der Vaart and Wellner (1996, corollary 2.2.8), such that the right hand side in (15) converges to zero as $j\to\infty$. □

LEMMA B.6. *For any $c>0$ and $\bar{S}_{\boldsymbol{nw}}(t)=\sup_{w^*\in\mathscr{W}}|\tilde{S}_{\boldsymbol{nw}}(t,w^*)|$, $\sup_{t\in\mathbb{R}^d}\{\bar{S}_{\boldsymbol{nw}}(t)-c\|t\|\}=O_p(1)$.*

PROOF. For any $1\leq t^*<\infty$, using B.2 and the fact that $\inf_{\|t\|>t^*}\|t\|/c_t=1/2$,

$$\limsup_{n\to\infty}\mathbb{P}\Big[\sup_{t\in\mathbb{R}^d}\{\bar{S}_{\boldsymbol{nw}}(t)-c\|t\|\}>C\Big]\leq$$
$$\limsup_{n\to\infty}\mathbb{P}\Big\{\sup_{\|t\|\leq t^*}\bar{S}_{\boldsymbol{nw}}(t)>C\Big\}+\lim_{n\to\infty}\mathbb{P}\Big\{\sup_{\|t\|>t^*}\bar{S}^*_{\boldsymbol{nw}}(t)>\frac{c}{2}\Big\}.$$

The first right hand side term converges to zero as $C\to\infty$ by B.2. The second right hand side term converges to zero as $t^*\to\infty$ by B.5. □

LEMMA B.7. *For any $c>0$ and all sufficiently large n, $\sup_{t\in\Gamma_n}\mathbb{1}\{|R_{nw}(t)|-ct^{\mathsf{T}}Vt/4>0\}=0$.*

PROOF. Note that for some $c^*=c^*(t,n)\in[0,1]$ and using the short hand $Q''_{nw}(c^*t)=\partial_{\theta\theta^{\mathsf{T}}}Q(\theta_{0n}+c^*t/\tilde{\alpha}_n,\eta_0+w/\mathfrak{r}_n)$,

$$\sup_{t\in\Gamma_n}\{|R_{nw}(t)|-ct^{\mathsf{T}}Vt/4\}\leq\frac{1}{2}\sup_{t\in\Gamma_n}\max\Big[t^{\mathsf{T}}\{Q''_{nw}(c^*t)+(1-c/2)V\}t,t^{\mathsf{T}}\{-Q''_{nw}(c^*t)-(1+c/2)V\}t\Big].$$

The stated result then follows from the fact that $c^*\in[0,1]$ and that all eigenvalues of $Q''_{nw}(c^*t)+(1-c/2)V$ and $-Q''_{nw}(c^*t)-(1+c/2)V$ are nonpositive for sufficiently large n, uniformly in Γ_n, by the continuity of $\partial_{\theta\theta^{\mathsf{T}}}Q$ in both θ and η; see assumption C. □

LEMMA B.8. *For any polynomial function P and any $0\leq c^*<\infty$,*

$$\sup_{w^*\in\mathscr{W}}\int_{\Gamma_n}\|P(t)\|\exp\{c^*\tilde{S}_{\boldsymbol{nw}}(t,w^*)\}\big|\exp\{R_{nw}(t)\}-1\big|\phi_V(t)dt=o_p(1).$$

PROOF. Let $\mathscr{I}_{\boldsymbol{n}}$ be the left hand side of the lemma statement and let $0 < c < 1$. We will show that $\mathscr{I}_{\boldsymbol{n}} = O_p(c)$. Taking $b = R_{nw}(t)/c$ and $j = 0$ in B.3 $C_V \mathscr{I}_{\boldsymbol{n}}/c$ is bounded by

$$\int_{\Gamma_n} \|P(t)\| \exp\{c^* \bar{\boldsymbol{S}}_{\boldsymbol{nw}}(t) + |R_{nw}(t)/c| - t^\top V t/2\} dt. \tag{16}$$

By B.7, for sufficiently large n, (16) is bounded above by $\int_{\Gamma_n} \|P(t)\| \exp\{c^* \bar{\boldsymbol{S}}_{\boldsymbol{nw}}(t) - t^\top V t/4\} dt$ with probability one, which does not depend on c and is $O_p(1)$ by B.6. Therefore, it follows that $\mathscr{I}_{\boldsymbol{n}} = O_p(c)$; let $c \downarrow 0$. □

LEMMA B.9. *For any polynomial function P,*

$$\sup_{w^* \in \mathscr{W}} \int_{\Gamma_n} \|P(t)\| \, |\pi_{nw}(t) - \pi_0| \exp\{c_\alpha^3 \tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\} \phi_V(t) dt = o_p(1). \tag{17}$$

PROOF. Note that by assumption D, for some $0 < C < \infty$, $|\pi_{nw}(t) - \pi_0| \le \big|\pi_0 - \pi\{\theta_{0n}(w)\}\big| + C\|t\|/\tilde{\alpha}_n$, where the first right hand side term is $o(1)$, and does not depend on t. Therefore, letting $\mathscr{J}_{\boldsymbol{n}}(t) = \|P(t)\| \exp\{c_\alpha^3 \bar{\boldsymbol{S}}_{\boldsymbol{nw}}(t)\} \phi_V(t) dt$, the left hand side in (17) is bounded above by $o(1) \int_{\Gamma_n} \mathscr{J}_{\boldsymbol{n}}(t) dt + \int_{\Gamma_n} \|t\| \mathscr{J}_{\boldsymbol{n}}(t) dt/\tilde{\alpha}_n$, which is $o(1) O_p(1) + O_p(1/\tilde{\alpha}_n) = o_p(1)$ by B.6. □

LEMMA B.10. *For any polynomial function P and constants $c_1, c_2 > 0$,*

$$\exp(\gamma_n) \int_{\Gamma_n^c} \|P(t)\| \exp\{c_1 \bar{\boldsymbol{S}}_{\boldsymbol{nw}}(t) - c_2 t^\top V t\} dt = o_p(1).$$

PROOF. Follows immediately from B.6. □

LEMMA B.11. *For any polynomial function P and constants $c_1, c_2, c_3 \ge 0$ with $c_2 + c_3 > 0$, and any sequences $\{\rho_n\}$ and $\{\tilde{\beta}_n\}$ with $1/\rho_n = O(1)$ and $\tilde{\beta}_n = O(1)$,*

$$\boldsymbol{I^c} = \exp(\gamma_n) \int_{\Gamma_n^{c*}} \|P(t)\| \exp\big[\rho_n^2 \{c_1 \tilde{\beta}_n \bar{\boldsymbol{S}}_{\boldsymbol{nw}}(t) + c_2 Q_{nw}(t) - c_3 t^\top V t\}\big] dt = o_p(1).$$

PROOF. Let $\bar{\boldsymbol{S}}^{\boldsymbol{c}}_{\boldsymbol{nw}} = \sup_{t \in \Gamma_n^c} \bar{\boldsymbol{S}}_{\boldsymbol{nw}}(t)/\|t\|$. Then for any $\epsilon^*, \epsilon > 0$,

$$\mathbb{P}(\boldsymbol{I^c} > \epsilon^*) \le \mathbb{P}(\bar{\boldsymbol{S}}^{\boldsymbol{c}}_{\boldsymbol{nw}} > \epsilon) + \mathbb{P}(\bar{\boldsymbol{S}}^{\boldsymbol{c}}_{\boldsymbol{nw}} \le \epsilon, \boldsymbol{I^c} > \epsilon^*). \tag{18}$$

RHS1 in (18) is $o(1)$ by B.5. For RHS2, note that if $\bar{\boldsymbol{S}}^{\boldsymbol{c}}_{\boldsymbol{nw}} \le \epsilon$ then for some $\bar{\beta} < \infty$,

$$\boldsymbol{I^c} \le \exp(\gamma_n) \int_{\Gamma_n^{c*}} \|P(t)\| \exp\big[\rho_n^2 \{c_1 \bar{\beta} \epsilon \|t\| + c_2 Q_{nw}(t) - c_3 t^\top V t\}\big] dt.$$

By B.4,

$$\sup_{t\in\Gamma_n^{c*}}\{c_1\bar\beta\epsilon\|t\|+c_2Q_{nw}(t)-c_3t^\top Vt\}\le\sup_{t\in\Gamma_n^{c*}}\{c_1\bar\beta\epsilon\|t\|-c_3t^\top Vt-c_2\min(c_q\tilde\alpha_{\boldsymbol n}^2,t^\top Vt/4)\}$$
$$\le\sup_{t\in\Gamma_n^{c*}}\{c_1\bar\beta\epsilon\|t\|-(c_3+c_2/4)\lambda^-\|t\|^2\}+\sup_{t\in\Gamma_n^{c*}}\big(c_1\bar\beta\epsilon\|t\|-c_3\lambda_-\|t\|^2-c_2c_q\tilde\alpha_n^2\big),$$

which, for some $c>0$ and all sufficiently large n, is bounded by $-c\gamma_n^2$. So,

$$\lim_{n\to\infty}\mathbb{P}(\bar{\boldsymbol S}^{\boldsymbol c}_{\boldsymbol{nw}}\le\epsilon,\boldsymbol I^{\boldsymbol c}>\epsilon^*)\le\lim_{n\to\infty}\mathbb{1}\Big\{\exp(\gamma_n-c\gamma_n^2)\int_{\Gamma_n^{c*}}\|P(t)\|dt>\epsilon^*\Big\}=0. \qquad\square$$

LEMMA B.12. *For any integer $j\ge0$,*

$$\sup_{w^*\in\mathscr W}\left\|\int t^j\exp\{c_\alpha^3\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)\}\big[\pi_{nw}(t)\exp\{Q_{nw}(t)\}-\pi_0\exp(-t^\top Vt/2)\}\big]dt\right\|=o_p(1).$$

PROOF. Let $\boldsymbol k_{\boldsymbol 1}(t,w^*)=t^j\pi_{nw}(t)\exp\{c_\alpha^3\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)+Q_{nw}(t)\}$ and $\boldsymbol k_{\boldsymbol 2}(t,w^*)=t^j\pi_0\times\exp\{c_\alpha^3\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)-t^\top Vt/2\}$. Noting that by assumption D $\boldsymbol k_{\boldsymbol 1}(t,w^*)=0$ for $t\in\Gamma_{\boldsymbol n}^c-\Gamma_{\boldsymbol n}^{c*}$, we have (omitting arguments)

$$\sup\left\|\int(\boldsymbol k_{\boldsymbol 1}-\boldsymbol k_{\boldsymbol 2})\right\|\le\sup\left\|\int_{\Gamma_n}(\boldsymbol k_{\boldsymbol 1}-\boldsymbol k_{\boldsymbol 2})\right\|+\sup\left\|\int_{\Gamma_n^{c*}}\boldsymbol k_{\boldsymbol 1}\right\|+\sup\left\|\int_{\Gamma_n^c}\boldsymbol k_{\boldsymbol 2}\right\|. \tag{19}$$

The first right hand side term in (19) is $o_p(1)$ by B.8 and B.9, the second term is $o_p(1)$ by B.11, and the last term is $o_p(1)$ by B.10. $\square$

PROOF OF PART (*i*) OF THEOREM 1. Note that for $\tilde\alpha_n=\alpha_n$ by (5) and B.12,

$$\alpha_n\{\hat{\boldsymbol\theta}^{\boldsymbol *}(w^*,w)-\theta_{0n}(w)\}=\frac{\int t\pi_{nw}(t)\exp\{c_\alpha^3\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)+Q_{nw}(t)\}dt}{\int\pi_{nw}(t)\exp\{c_\alpha^3\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)+Q_{nw}(t)\}dt}=$$
$$\frac{\int t\exp\{c_\alpha^3\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)\}\phi_V(t)dt}{\int\exp\{c_\alpha^3\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)\}\phi_V(t)dt}+o_p(1),$$

uniformly in $w^*\in\mathscr W$. Apply B.2 together with the continuous mapping theorem. Divide both sides by c_α^2. $\square$

APPENDIX C: PART (ii) OF THEOREM 1

LEMMA C.1. *Let $\tilde\gamma>0$, $\tilde\Gamma=\{t\in\mathbb R^d:\|t\|\le\tilde\gamma\}$, $\rho_n=\alpha_n^2/n^{2/3}$, and $\tilde{\boldsymbol L}_{\boldsymbol{nw}}(t,w^*)=\tilde{\boldsymbol S}_{\boldsymbol{nw}}(t,w^*)+Q_{nw}(t)$. Let further,*

$$\boldsymbol X_{\boldsymbol n}(w^*)=\frac{\int_{\tilde\Gamma}t\pi_{nw}(t)\exp\{\rho_n\tilde{\boldsymbol L}_{\boldsymbol{nw}}(t,w^*)\}dt}{\int_{\tilde\Gamma}\pi_{nw}(t)\exp\{\rho_n\tilde{\boldsymbol L}_{\boldsymbol{nw}}(t,w^*)\}dt},$$

and for all $w^*\in\mathscr W$, let $\mathbb X(w^*)=\mathbb X=\operatorname{argmax}_{t\in\tilde\Gamma}\mathbb C(t)$. Then $\boldsymbol X_{\boldsymbol n}\overset{w}{\to}\mathbb X$ in $\mathbb L^\infty(\mathscr W)$.

PROOF. We will use the Skorokhod representation theorem and the fact that $\mathbb{L}^2(\Omega, \mathscr{B}, \mu)$ is separable, where $(\Omega, \mathscr{B}, \mu)$ is a measure space. Let $\Omega = \tilde{\Gamma} \times \mathscr{W}$ and $\mathscr{B}$ be the Boreal sigma algebra on Ω. Let μ be the Lebesque measure. By B.2 and the fact that $\sup_{(t,w^*)\in\tilde{\Gamma}\times\mathscr{W}} \left|Q_{nw}(t) + t^\top V t/2\right| = o(1)$, we have the weak convergence of $\tilde{\boldsymbol{L}}_{\boldsymbol{nw}}$ to $\mathbb{C}$ in $\mathbb{L}^\infty(\tilde{\Gamma} \times \mathscr{W})$. Moreover, since the limit $\mathbb{C}$ is in $\mathbb{L}^2(\tilde{\Gamma} \times \mathscr{W}, \mathscr{B}, \mu)$ with probability one, the Skorokhod representation theorem implies that there exist $\tilde{\boldsymbol{L}}^*_{\boldsymbol{nw}}$ and $\mathbb{C}^*$ with the same distributions as $\tilde{\boldsymbol{L}}_{\boldsymbol{nw}}$ and $\mathbb{C}$ for which

$$\int_{\tilde{\Gamma}\times\mathscr{W}} \left|\tilde{\boldsymbol{L}}^*_{\boldsymbol{nw}}(t, w^*) - \mathbb{C}^*(t)\right|^2 dt dw^* = o_{\text{a.s.}}(1). \tag{20}$$

We will now establish that

$$\sup_{w^*\in\mathscr{W}} \left\|\boldsymbol{X}^*_{\boldsymbol{n}}(w^*) - \mathbb{X}^*\right\| = o_{\text{a.s}}(1), \tag{21}$$

where $\mathbb{X}^* = \operatorname{argmax}_{t\in\tilde{\Gamma}} \mathbb{C}^*(t)$ and

$$\boldsymbol{X}^*_{\boldsymbol{n}}(w^*) = \frac{\int_{\tilde{\Gamma}} t\pi_{nw}(t) \exp\{\rho_n \tilde{\boldsymbol{L}}^*_{\boldsymbol{nw}}(t, w^*)\} dt}{\int_{\tilde{\Gamma}} \pi_{nw}(t) \exp\{\rho_n \tilde{\boldsymbol{L}}^*_{\boldsymbol{nw}}(t, w^*)\} dt}.$$

First, let $\tilde{L}^*_{nw}, \mathbb{C}^*$ be sample paths of $\tilde{\boldsymbol{L}}^*_{\boldsymbol{nw}}, \mathbb{C}^*$ for which the $\mathbb{L}^2$–convergence in (20) holds. Let for arbitrary $c > 0$ and $\varkappa : \tilde{\Gamma} \times \mathscr{W} \mapsto \mathbb{R}$, $T(\varkappa, c) = \left\{(t, w^*) \in \tilde{\Gamma} \times \mathscr{W} : \left|\varkappa(t, w^*) - \bar{\mathbb{C}}^*\right| \le c\right\}$, where $\bar{\mathbb{C}}^* = \max_{t\in\tilde{\Gamma}} \mathbb{C}^*(t)$. Let further for arbitrary sets $T_1, T_2 \subset \tilde{\Gamma} \times \mathscr{W}$, $d^*(T_1, T_2) = \mu(T_1 - T_2) + \mu(T_2 - T_1)$. We first establish that

$$d^*\left\{T(\tilde{L}^*_{nw}, c), T(\mathbb{C}^*, c)\right\} = o(1). \tag{22}$$

Let $T_{1n}(c) = T(\tilde{L}^*_{nw}, c) - T(\mathbb{C}^*, c)$ and $T_{2n}(c) = T(\mathbb{C}^*, c) - T(\tilde{L}^*_{nw}, c)$. We show that $\mu\{T_{2n}(c)\} = o(1)$; establishing that $\mu\{T_{1n}(c)\} = o(1)$ is analogous.

Let for arbitrary $c^* > 0$, $T^*_n(c^*) = \left\{(t, w^*) \in \tilde{\Gamma} \times \mathscr{W} : \left|\tilde{L}^*_{nw}(t, w^*) - \mathbb{C}^*(t)\right| \le c^*\right\}$. Defining complements relative to $\tilde{\Gamma} \times \mathscr{W}$, we note that

$$\mu\left\{T_{2n}(c)\right\} = \mu\left\{T_{2n}(c) \cap T^*_n(c^*)\right\} + \mu\left\{T_{2n}(c) \cap T^{*c}_n(c^*)\right\} \le \\ \mu\left\{T_{2n}(c) \cap T^*_n(c^*)\right\} + \mu\left\{T^{*c}_n(c^*)\right\} = \mu\left\{T_{2n}(c) \cap T^*_n(c^*)\right\} + o(1),$$

by (20). Further,

$$T_{2n}(c) \cap T^*_n(c^*) \subset T^{**}_n(c, c^*) = \left\{(t, w^*) \in \tilde{\Gamma} \times \mathscr{W} : c \le |\tilde{L}^*_{nw}(t, w^*) - \bar{\mathbb{C}}^*| \le c + c^*\right\},$$

such that by (20),

$$\lim_{c^*\downarrow 0} \lim_{n\to\infty} \mu\left\{T^{**}_n(c, c^*)\right\} = \lim_{c^*\downarrow 0} \mu\left[\left\{(t, w^*) \in \tilde{\Gamma} \times \mathscr{W} : c \le \left|\mathbb{C}^*(t) - \bar{\mathbb{C}}^*\right| \le c + c^*\right\}\right] =$$

$$\mu\Big[\big\{(t,w^*)\in\tilde\Gamma\times\mathscr{W}:\big|\mathbb{C}^*(t)-\bar{\mathbb{C}}^*\big|=c\big\}\Big]=0,$$

because $\mathbb{C}^*$ is continuous and nowhere differentiable. So (22) holds.

Finally, note that for $j=0,1$, $\bar\pi$ defined in assumption D, and some $C<\infty$,

$$\sup_{w^*\in\mathscr{W}}\int_{T^c(\tilde L^*_{nw},c)}\|t\|^j|\pi_{nw}(t)|\exp\big[\rho_n\big\{\tilde L^*_{nw}(t,w^*)-\bar{\mathbb{C}}^*\big\}\big]dt\le C\exp(-\rho_n c)\bar\pi\int_{\tilde\Gamma}\|t\|^j dt=o(1),$$

by the compactness of $\tilde\Gamma$ and divergence of ρ_n. Thus,

$$\sup_{w^*\in\mathscr{W}}\frac{\int_{\tilde\Gamma}t\pi_{nw}(t)\exp\big\{\rho_n\tilde L^*_{nw}(t,w^*)\big\}dt}{\int_{\tilde\Gamma}\pi_{nw}(t)\exp\big\{\rho_n\tilde L^*_{nw}(t,w^*)\big\}dt}\le\text{ess sup}\,T(\tilde L^*_{nw},c)+o(1)=\text{ess sup}\,T(\mathbb{C}^*,c)+o(1).$$

Repeat the arguments to obtain

$$\inf_{w^*\in\mathscr{W}}\frac{\int_{\tilde\Gamma}t\pi_{nw}(t)\exp\big\{\rho_n\tilde L^*_{nw}(t,w^*)\big\}dt}{\int_{\tilde\Gamma}\pi_{nw}(t)\exp\big\{\rho_n\tilde L^*_{nw}(t,w^*)\big\}dt}\ge\text{ess inf}\,T(\tilde L^*_{nw},c)+o(1)=\text{ess inf}\,T(\mathbb{C}^*,c)+o(1).$$

Then note that $\lim_{c\downarrow 0}\text{ess inf}\,T(\mathbb{C}^*,c)=\lim_{c\downarrow 0}\text{ess sup}\,T(\mathbb{C}^*,c)=\text{argmax}_{t\in\tilde\Gamma}\,\mathbb{C}^*(t)$, where the last equality follows from Kim and Pollard (1990, lemma 2.6).

So we have established (21). Hence for any bounded and continuous function $\mathfrak{f}:\mathbb{L}^\infty(\mathscr{W})\mapsto\mathbb{R}$ we have by the dominated convergence theorem that $\mathbb{E}\mathfrak{f}(\boldsymbol{X_n})=\mathbb{E}\mathfrak{f}(\boldsymbol{X_n^*})\to\mathbb{E}\mathfrak{f}(\mathbb{X}^*)=\mathbb{E}\mathfrak{f}(\mathbb{X})$. □

LEMMA C.2. *Let $\rho_n,\tilde\Gamma$ be defined as in C.1 and $\tilde\Gamma_n^c=\mathfrak{T}_n-\tilde\Gamma$. Then for any $\epsilon>0$ and $j=0,1$,*

$$\lim_{\tilde\gamma\to\infty}\lim_{n\to\infty}\mathbb{P}\left(\frac{\sup_{w^*\in\mathscr{W}}\int_{\tilde\Gamma_n^c}\|t\|^j\pi_{nw}(t)\exp\big[\rho_n\big\{\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)+Q_{nw}(t)\big\}\big]dt}{\inf_{w^*\in\mathscr{W}}\int\pi_{nw}(t)\exp\big[\rho_n\big\{\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)+Q_{nw}(t)\big\}\big]dt}>\epsilon\right)=0.\tag{23}$$

PROOF. Denote the numerator and denominator in (23) $\boldsymbol{I_N}$ and $\boldsymbol{I_D}$, respectively. Note that for any $\bar c>0$,

$$\mathbb{P}\big(\boldsymbol{I_N}/\boldsymbol{I_D}>\epsilon\big)\le\mathbb{P}\big\{\boldsymbol{I_N}>\epsilon\exp(-\bar c\rho_n)\big\}+\mathbb{P}\big\{\boldsymbol{I_D}\le\exp(-\bar c\rho_n)\big\}.\tag{24}$$

We first work on the first right hand side term in (24). For $c>0$ let $\boldsymbol{Z_n}(\tilde\gamma,c)=\big\{t\in\tilde\Gamma_n^c:\sup_{w^*\in\mathscr{W}}|\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)|\le c\|t\|^2\big\}$. For $c^*=2\sup_{\theta\in\Theta}\|\theta\|<\infty$, we have $\sup_{t\in\boldsymbol{Z_n}(\tilde\gamma,c)}\|t\|\le c^*\sqrt[3]{n}$, such that by B.4 for sufficiently small c,

$$\begin{aligned}\sup_{t\in\boldsymbol{Z_n}(\tilde\gamma,c)}\sup_{w^*\in\mathscr{W}}\exp\big[\rho_n\big\{\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)+Q_{nw}(t)\big\}\big]\le\\ \sup_{t\in\boldsymbol{Z_n}(\tilde\gamma,c)}\exp\big[\rho_n\big\{c\|t\|^2-\min\big(c_q n^{2/3},\lambda_-\|t\|^2/4\big)\big\}\big]\le\end{aligned}\tag{25}$$

$$\exp\{\alpha_n^2(cc^{*2} - c_q)\} + \sup_{t \in \boldsymbol{Z_n}(\tilde{\gamma},c)} \exp\{\rho_n \|t\|^2 (c - \lambda_-/4)\} \leq$$
$$\exp\{\rho_n(c - \lambda_-/4)/\tilde{\gamma}^2\} + o\{\bar{c}\alpha_n^2\} = o\{\exp(-\bar{c}\rho_n)\},$$

if $\bar{c} > 0$ is chosen sufficiently small. Further,

$$\lim_{\tilde{\gamma}\to\infty} \lim_{n\to\infty} \mathbb{P}\{t \in \tilde{\Gamma}_n^c - \boldsymbol{Z_n}(\tilde{\gamma}, c)\} \leq \lim_{\tilde{\gamma}\to\infty} \lim_{n\to\infty} \mathbb{P}\left\{\sup_{\|t\|>\tilde{\gamma}} \frac{\bar{\boldsymbol{S}}_{\boldsymbol{nw}}(t)}{\|t\|^2} > c\right\} = 0, \tag{26}$$

by B.5. Combining (25) and (26) implies that the first right hand side term in (24) is $o(1)$.

Now the second right hand side term in (24). Let $0 < \tilde{\epsilon} \leq \sqrt{\bar{c}/4\lambda^+}$ be some constant to be manipulated later. Note that for $\bar{\boldsymbol{S}}_{\boldsymbol{nw}}(\tilde{\epsilon}) = \sup_{\|t\|\leq\tilde{\epsilon}} \sup_{w^*\in\mathscr{W}} |\tilde{\bar{\boldsymbol{S}}}_{\boldsymbol{nw}}(t, w^*)|$,

$$\mathbb{P}\{\boldsymbol{I_D} \leq \exp(-\bar{c}\rho_n)\} \leq \mathbb{P}\{\bar{\boldsymbol{S}}_{\boldsymbol{nw}}(\tilde{\epsilon}) > \bar{c}/2\} + \mathbb{P}\{\bar{\boldsymbol{S}}_{\boldsymbol{nw}}(\tilde{\epsilon}) \leq \bar{c}/2,\ \boldsymbol{I_D} \leq \exp(-\bar{c}\rho_n)\}. \tag{27}$$

Using assumption C, the second right hand side term in (27) is for sufficiently large n and $\underline{\pi} > 0$ bounded by

$$\mathbb{1}\left[\underline{\pi} \int_{\|t\|\leq\tilde{\epsilon}} \exp\{\rho_n(-\bar{c}/2 - \lambda^+\tilde{\epsilon}^2)\}dt \leq \exp(-\bar{c}\rho_n)\right] \leq \mathbb{1}\{\underline{\pi}\tilde{\epsilon}\exp(\rho_n\bar{c}/4) \leq 1\} = o(1).$$

Finally, for the first right hand side term in (27) note that by B.2,

$$\lim_{\tilde{\epsilon}\to 0} \lim_{n\to\infty} \mathbb{P}\{\bar{\boldsymbol{S}}_{\boldsymbol{nw}}(\tilde{\epsilon}) > \bar{c}/2\} = \lim_{\tilde{\epsilon}\to 0} \mathbb{P}\left\{\sup_{\|t\|\leq\tilde{\epsilon}} |\mathbb{G}(t)| > \bar{c}/2\right\} = 0,$$

since $\mathbb{G}(0) = 0$ by definition. □

Proof of part (*ii*) of theorem 1. For $\rho_n = \alpha_n^2/n^{2/3}$, $j = 0, 1$, and any set $\Gamma \in \mathbb{R}^d$, let $\boldsymbol{I_j}(\Gamma) = \int_\Gamma t^j \pi_{nw}(t) \exp\{\rho_n \tilde{\boldsymbol{L}}_{\boldsymbol{nw}}(t, w^*)\}dt$, and $\boldsymbol{I_j} = \boldsymbol{I_j}(\mathbb{R}^d)$. Then for $\tilde{\Gamma}, \tilde{\Gamma}_n^c$ defined in C.1 and C.2 by assumption D and (5) using $t = \sqrt[3]{n}\{\theta - \theta_{0n}(w)\}$,

$$\sqrt[3]{n}\{\hat{\boldsymbol{\theta}}^*(w^*, w) - \theta_{0n}(w)\} = \frac{\boldsymbol{I_1}}{\boldsymbol{I_0}} = \frac{\boldsymbol{I_1}(\tilde{\Gamma}) + \boldsymbol{I_1}(\tilde{\Gamma}_n^c)}{\boldsymbol{I_0}(\tilde{\Gamma}) + \boldsymbol{I_0}(\tilde{\Gamma}_n^c)} = \frac{\boldsymbol{I_1}(\tilde{\Gamma})}{\boldsymbol{I_0}(\tilde{\Gamma})}\left\{1 - \frac{\boldsymbol{I_0}(\tilde{\Gamma}_n^c)}{\boldsymbol{I_0}}\right\} + \frac{\boldsymbol{I_1}(\tilde{\Gamma}_n^c)}{\boldsymbol{I_0}}.$$

Apply C.1 and C.2. □

APPENDIX D: PART (iii) OF THEOREM 1

Lemma D.1. $\int \pi_{nw}(t)\exp\{Q_{nw}(t)\}dt = \pi_0 C_V + o(1)$.

Proof. The difference between left hand side and right hand side can be expanded as

$$(28)\quad C_V\int_{\Gamma_n}\pi_{nw}(t)\big[\exp\{R_{nw}(t)\}-1\big]\phi_V(t)dt+C_V\int_{\Gamma_n}\{\pi_{nw}(t)-\pi_0\}\phi_V(t)dt+\int_{\Gamma_n^c}\pi_{nw}(t)\exp\{Q_{nw}(t)\}dt-C_V\pi_0\int_{\Gamma_n^c}\phi_V(t)dt,$$

where all four terms are $o(1)$ by B.7, assumption D, B.11, and the fact that $1/\gamma_n=o(1)$, respectively. □

Lemma D.2. *For* $\mathscr{V}_N=\pi_0^2\iint ts^\top H(t,s)\phi_V(t)\phi_V(s)dtds$, $\int\pi_{nw}(t)t\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)\phi_V(t)dt$ *converges weakly in* $\mathbb{L}^\infty(\mathscr{W})$ *to a flat limit process whose marginals have a* $N(0,\mathscr{V}_N)$*–distribution.*

Proof. By B.2 and assumption D, as a process of w^*,

$$\int\pi_{nw}(t)t\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)\phi_V(t)dt\overset{w}{\to}\pi_0\int t\mathbb{G}(t)\phi_V(t)dt\sim N(0,\mathscr{V}_N).$$ □

Lemma D.3. *For* $j=0,1$, *if* $\beta_n=o(1)$,

$$\sup_{w^*\in\mathscr{W}}\left\|\int\pi_{nw}(t)t^j\left[\exp\{\beta_n\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)\}-\sum_{s=0}^{j}\{\beta_n\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)\}^s\right]\exp\{Q_{nw}(t)\}dt\right\|=O_p(\beta_n^{j+1}).$$

Proof. By B.3 with $c=\beta_n$ and $b=\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t,w^*)$, the left hand side is bounded by $\beta_n^{j+1}\boldsymbol{I}$, where $\boldsymbol{I}=\int|\pi_{nw}(t)|\,\|t\|^j\exp\{\bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t)c_t+Q_{nw}(t)\}dt$. So it suffices to show that $\boldsymbol{I}=O_p(1)$. Now, for any $0<C^*<\infty$,

$$(29)\quad \lim_{C\to\infty}\limsup_{n\to\infty}\mathbb{P}(\boldsymbol{I}>C)\le\lim_{C\to\infty}\limsup_{n\to\infty}\mathbb{P}\left\{\boldsymbol{I}>C,\ \sup_{t\in\mathbb{R}^d}\bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t)\le C^*\right\}+\limsup_{n\to\infty}\mathbb{P}\left\{\sup_{t\in\mathbb{R}^d}\bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t)>C^*\right\}.$$

The first right hand side term in (29) is by B.4 for some polynomial function P bounded by

$$\lim_{C\to\infty}\limsup_{n\to\infty}\mathbb{1}\left[\exp(C^*)\int_{\mathfrak{T}_n}\|P(t)\|\exp\{C^*c_t-\min(t^\top Vt/4,c_q\tilde{\alpha}_n^2)\}dt>C\right]=0,$$

since c_t cannot grow faster than $\tilde{\alpha}_n$ by the definition of $\mathfrak{T}_n$, and the normal distribution has infinitely many moments. Finally, by B.2, the second right hand side term in (29) equals $\mathbb{P}\{\sup_{t\in\mathbb{R}^d}\mathbb{G}^*(t)>C^*\}$. Let $C^*\to\infty$. □

Lemma D.4. *For any polynomial function* P, $\int_{\mathfrak{T}_n}\|P(t)\|\bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t)\big|\exp\{R_{nw}(t)\}-1\big|\phi_V(t)dt=o_p(1)$.

PROOF. Note that $\sup_{t\in\mathbb{R}^d} \bar{\boldsymbol{S}}^*_{\boldsymbol{nw}}(t) = O_p(1)$ by B.6, so it suffices to show that

$$\begin{cases} \int_{\Gamma_n} \|P(t)\| \left|\exp\{R_{nw}(t)\} - 1\right| \phi_V(t)dt &= o_p(1), \\ \int_{\Gamma_n^{c*}} \|P(t)\| \exp\{Q_{nw}(t)\}dt &= o_p(1), \\ \int_{\Gamma_n^{c*}} \|P(t)\| \phi_V(t)dt &= o_p(1), \end{cases}$$

which are established in B.8 and B.11. □

PROOF OF PART (*iii*) OF THEOREM 1. Define

$$\mathbb{B}_{nw} = \beta_n^{-1}\alpha_n^{d+1} \int \{\theta - \theta_{0n}(w)\}\pi(\theta)\exp\{\alpha_n^2 Q(\theta, \eta_0 + w/\mathrm{r}_n)\}d\theta, \tag{30}$$

$$\mathbb{D}_{\boldsymbol{nw}}(w^*) = \alpha_n^d \int \pi(\theta)\exp\Big[\alpha_n^2\{\boldsymbol{S_n}(\theta, \eta_0 + w^*/\mathrm{r}_n) + Q(\theta, \eta_0 + w/\mathrm{r}_n)\}\Big]d\theta. \tag{31}$$

By (5) we get

$$\sqrt{n/\alpha_n}\{\hat{\boldsymbol{\theta}}^*(w^*, w) - \theta_{0n}(w)\} = \frac{\beta_n^{-1}\int t\pi_{nw}(t)\exp\{\beta_n \tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*) + Q_{nw}(t)\}dt}{\int \pi_{nw}(t)\exp\{\beta_n \tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*) + Q_{nw}(t)\}dt}. \tag{32}$$

Denote the right hand side numerator in (32) $\mathbb{N}_{\boldsymbol{nw}}(w^*)$ and note that the denominator equals $\mathbb{D}_{\boldsymbol{nw}}(w^*)$.Then $\mathbb{D}_{\boldsymbol{nw}}(w^*) - C_V\pi_0$ is equal to

$$\int \pi_{nw}(t)\Big[\exp\{\beta_n \tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\} - 1\Big]\exp\{Q_{nw}(t)\}dt + \Big[\int \pi_{nw}(t)\exp\{Q_{nw}(t)\}dt - C_V\pi_0\Big],$$

where the first term is $o_p(1)$, uniformly in w^*, by D.3 and the second term is $o_p(1)$ by D.1. So $\mathbb{D}_{\boldsymbol{nw}}(w^*) = C_V\pi_0 + o_p(1)$, uniformly in w^*.

For $\mathbb{N}_{\boldsymbol{nw}}(w^*)$, noting that $\mathbb{B}_{nw} = \beta_n^{-1}\int t\pi_{nw}(t)\exp\{Q_{nw}(t)\}dt$, we have

$$\mathbb{N}_{\boldsymbol{nw}}(w^*) - \mathbb{B}_{nw} = \frac{1}{\beta_n}\int \pi_{nw}(t)t\big[\exp\{\beta_n\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\} - 1 - \beta_n\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\big]\exp\{Q_{nw}(t)\}dt +$$
$$C_V\int \pi_{nw}(t)t\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\big[\exp\{R_{nw}(t)\} - 1\big]\phi_V(t)dt + C_V\int \pi_{nw}(t)t\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\phi_V(t)dt,$$

where the first two right hand side terms are $o_p(1)$, uniformly in w^*, by D.3 and D.4. Hence, $\mathbb{N}_{\boldsymbol{nw}}(w^*) - \mathbb{B}_{nw} \overset{w}{\to} C_V\pi_0 N(0, \mathscr{V})$, as a process of w^*. □

APPENDIX E: PART (iv) OF THEOREM 1

LEMMA E.1.

$$\int_{\Gamma_n} t\pi_{nw}(t)\exp\Big[\exp\{R_{nw}(t)\} - \sum_{p=0}^{q}\frac{R_{nw}^p(t)}{p!}\Big]\phi_V(t)dt = O\{(\gamma_n^3/\alpha_n)^{q+1}\} = o(1). \tag{33}$$

PROOF. First note that for $s_n^* = \sup_{t\in\Gamma_n}\left|R_{nw}(t)\right|$ and by the definition of γ_n in appendix A,

$$s_n^* \le \gamma_n^2 \sup_{\|t\|\le\gamma_n} \left\|\partial_{\theta\theta^\intercal} Q\{\theta_{0n}(w)+t/\alpha_n, \eta_0 + w/\mathfrak{r}_n\} + V\right\| = O\left(\gamma_n^3/\alpha_n\right) = o(1),$$

by assumption C and the definition of γ_n in appendix A. The length of the left hand side in (33) is equal to

$$\left\|\sum_{p=q+1}^{\infty}\int_{\Gamma_n}\pi_{nw}(t)t\frac{R_{nw}^p(t)}{p!}\phi_V(t)dt\right\| \le (s_n^*)^{q+1}\exp(s_n^*)\int_{\Gamma_n}\left\|\pi_{nw}(t)t\right\|\phi_V(t)dt = O\left\{(s_n^*)^{q+1}\right\}.$$

□

PROOF OF PART (*iv*) OF THEOREM 1. Since $\mathbb{B}_{nw} = \beta_n^{-1}\int t\pi_{nw}(t)\exp\{Q_{nw}(t)\}dt$, it suffices to show that

$$\int t\pi_{nw}(t)\exp\{R_{nw}(t)\}\phi_V(t)dt = \sum_{\tau=0}^{q}\frac{b_{q\tau}^*}{\alpha_n^\tau} + o(\alpha_n^{-q}). \tag{34}$$

By B.11 and E.1, the left hand side in (34) equals

$$\sum_{p=0}^{q}\frac{1}{p!}\int_{\Gamma_n}\pi_{nw}(t)tR_{nw}^p(t)\phi_V(t)dt + o\left(\alpha_n^{-q}\right). \tag{35}$$

Let $D_{Q,\delta;nw}(t)/\alpha_n^\delta$ and $D_{\pi\delta;nw}(t)/\alpha_n^\delta$ be the order δ terms in the Taylor expansions of $Q\{\theta_{0n}(w)+t/\alpha_n, \eta_0+w^*/\mathfrak{r}_n\}$ and $\pi\{\theta_{0n}(w)+t/\alpha_n\}$ around $\left(\theta_{0n}(w), \eta_0+w^*/\mathfrak{r}_n\right)$ and $\theta_{0n}(w)$, respectively. Then, by assumptions C and D and B.1 and the mean value theorem, $R_{nw}(t)$ and $\pi_{nw}(t)$ have expansions such that

$$R_{nw}(t) = \sum_{\delta=1}^{q}\frac{D_{Q,\delta+2;nw}(t)}{\alpha_n^\delta} + \frac{D_{Q\Delta nw}^*(t)}{\alpha_n^q}, \qquad \pi_{nw}(t) = \sum_{\delta=0}^{q}\frac{D_{\pi\delta;nw}(t)}{\alpha_n^\delta} + \frac{D_{\pi qnw}^*(t)}{\alpha_n^q}, \tag{36}$$

where the remainders $D_{Q\Delta nw}^*(t)$ and $D_{\pi qnw}^*(t)$ satisfy

$$\sup_{t\in\Gamma_n\setminus\{0\}}\left\{\left|D_{Q\Delta;nw}^*(t)\right|/\|t\|^\Delta\right\} = o(1), \qquad \sup_{t\in\Gamma_n\setminus\{0\}}\left\{\left|D_{\pi q;nw}^*(t)\right|/\|t\|^q\right\} = o(1). \tag{37}$$

Further, by the definition of γ_n, assumptions C and D and B.1,

$$\begin{cases}\sup_{t\in\Gamma_n}\left|D_{Q;\delta+2;nw}(t) - D_{Q;\delta+2}(t)\right|/\alpha_n^\delta = O\left(\gamma_n^{\delta+2}/\mathfrak{r}_n\alpha_n^\delta\right) = o\left(\alpha_n^{-q}\right), & \delta = 1,\ldots,q,\\ \sup_{t\in\Gamma_n}\left|D_{\pi\delta;nw}(t) - D_{\pi\delta}(t)\right|/\alpha_n^\delta = O\left(\gamma_n^\delta/\mathfrak{r}_n\alpha_n^\delta\right) = o\left(\alpha_n^{-q}\right), & \delta = 0,\ldots,q,\end{cases} \tag{38}$$

Now, (36) to (38) and the fact that $\int\|P(t)\|\phi_V(t)dt < \infty$ for any polynomial function P imply that (35) equals

$$\sum_{p=0}^{q}\frac{1}{p!}\int_{\Gamma_n}\left\{\sum_{\delta=0}^{q}\frac{D_{\pi\delta}(t)}{\alpha_n^\delta}\right\}\left\{\sum_{\delta=1}^{q}\frac{D_{Q,\delta+2}(t)}{\alpha_n^\delta}\right\}^p t\phi_V(t)dt + o\left(\alpha_n^{-q}\right). \tag{39}$$

Let $\mathscr{M}_{pq}$ be the collection of vectors $m = (m_1, \cdots, m_q)$ consisting of nonnegative integers for which $\sum_{j=1}^{q} m_j = p$,such that $\mathscr{M}_{pqs}^* \subset \mathscr{M}_{pq}$ for all s. Then the first term in (39) is by the multinomial theorem equal to

$$\sum_{p=0}^{q} \int_{\Gamma_n} \Big\{ \sum_{\delta=0}^{q} \frac{D_{\pi\delta}(t)}{\alpha_n^\delta} \Big\} \sum_{m \in \mathscr{M}_{pq}} \prod_{\delta=1}^{q} \Big[\Big\{ \frac{D_{Q;\delta+2}}{\alpha_n^\delta} \Big\}^{m_\delta} \frac{1}{m_\delta!} \Big] t \phi_V(t) dt =$$
$$\sum_{p=0}^{q} \sum_{m \in \mathscr{M}_{pq}} \sum_{j=0}^{q} \frac{1}{\alpha_n^{j+\sum_{\delta=1}^{q} m_\delta \delta}} \int_{\Gamma_n} D_{\pi j}(t) \Big\{ \prod_{\delta=1}^{q} \frac{D_{Q,\delta+2}^{m_\delta}(t)}{m_\delta!} \Big\} t \phi_V(t) dt =$$
$$\sum_{j=0}^{q} \sum_{s=0}^{q-j} \frac{b_{qsj}}{\alpha_n^{s+j}} + O\big(\alpha_n^{-q-1}\big),$$

where $\int_{\Gamma_n^c} \cdot$ is negligible by B.10. □

APPENDIX F: THEOREM 2

Lemma F.1. $\int \{D_{\pi 1}(t) + \pi_0 D_{Q3}(t)\} t \phi_V(t) dt = 0.$

Proof. Note that $D_{\pi 1}(t) = -(\pi_0/2) \sum_{s=1}^{d} \{\partial_{\theta_s} \operatorname{vec}(\partial_{\theta\theta^\top} Q)\}^\top (\theta_0) \operatorname{vec}(V^{-1}) t_s$. Therefore, letting V^{pm} be the (p, m) element of V^{-1} and $\mathscr{K}_{pm\delta} = \partial_{\theta_p \theta_m \theta_\delta} Q(\theta_0)$, the j^{th} element of the left hand side of the lemma statement is

$$\int \pi_0 D_{Q3}(t) t_j \phi_V(t) dt + \int D_{\pi 1}(t) t_j \phi_V(t) dt =$$
$$\frac{\pi_0}{6} \sum_{p,m,\delta=1}^{d} \mathscr{K}_{pm\delta} \int t_p t_m t_\delta t_j \phi_V(t) dt - \frac{\pi_0}{2} \sum_{p,m,\delta=1}^{d} \mathscr{K}_{pm\delta} V^{m\delta} \int t_p t_j \phi_V(t) dt =$$
$$\frac{\pi_0}{6} \sum_{p,m,\delta=1}^{d} \mathscr{K}_{pm\delta} \big(V^{pm} V^{\delta j} + V^{p\delta} V^{mj} + V^{pj} V^{m\delta}\big) - \frac{\pi_0}{2} \sum_{p,m,\delta=1}^{d} \mathscr{K}_{pm\delta} V^{m\delta} V^{pj} = 0,$$

where the second equality is due to Isserlis's theorem for higher order moments of the multivariate normal distribution. □

Proof of theorem 2. Let $\hat{\boldsymbol{\pi}}_{\boldsymbol{nw}}(t) = \hat{\boldsymbol{\pi}}\{\theta_{0n}(w) + t/\tilde{\alpha}_n\}$, where $\tilde{\alpha}_n$ is chosen as in theorem 1. We first establish that if $q = 1$ then theorem 1 still holds if π is replaced with $\hat{\boldsymbol{\pi}}$, albeit that (since $q = 1$) $\mathbb{B}_{nw}$ is replaced with

(40) $$\hat{\mathbb{B}}_{\boldsymbol{nw}} = \int \{D_{\hat{\boldsymbol{\pi}} 1}(t) + \hat{\boldsymbol{\pi}}_{\boldsymbol{0}} D_{Q3}(t)\} \phi_V(t) dt / c_\alpha^4 \hat{\boldsymbol{\pi}}_{\boldsymbol{0}} + o_p(1).$$

For this, we need to allow for randomness of $\hat{\boldsymbol{\pi}}_{\boldsymbol{nw}}$ in the proofs involving π_{nw}, notably those of B.9, B.12, C.1, C.2 and D.1 to D.3 and appendix E. The argument is essentially the same in all cases, so we will use B.9 as a representative example. We need to show that

$$\sup_{w^*\in\mathscr{W}} \int_{\Gamma_n} \|P(t)\| \, |\hat{\boldsymbol{\pi}}_{\boldsymbol{nw}}(t) - \pi_0| \exp\{c_\alpha^3 \tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\}\phi_V(t)dt = o_p(1). \tag{41}$$

Denote the left hand side in (41) by $\hat{\boldsymbol{I}}$ and let $\hat{\boldsymbol{\sigma}} = \sup_{\theta\in\Theta}\left\|\partial_\theta \hat{\boldsymbol{\pi}}(\theta)\right\|$. Then for any $\epsilon > 0$,

$$\mathbb{P}(\hat{\boldsymbol{I}} > \epsilon) \le \mathbb{P}(\hat{\boldsymbol{I}} > \epsilon, \ \hat{\boldsymbol{\sigma}} \le \bar{\pi}_1) + \mathbb{P}(\hat{\boldsymbol{\sigma}} > \bar{\boldsymbol{\pi}}_{\boldsymbol{1}}) = \mathbb{P}(\hat{\boldsymbol{I}} > \epsilon, \ \hat{\boldsymbol{\sigma}} \le \bar{\pi}_1) + o(1). \tag{42}$$

Since for all $t \in \Gamma_n$, $|\hat{\boldsymbol{\pi}}_{\boldsymbol{nw}}(t) - \pi_0| \le \hat{\boldsymbol{\sigma}}\{\gamma_n/\tilde{\alpha}_n + \|\theta_{0n}(w) - \theta_0\|\}$, RHS1 in (42) is bounded by

$$\mathbb{P}\left[\bar{\pi}_1\{\gamma_n/\tilde{\alpha}_n + \|\theta_{0n}(w) - \theta_0\|\} \sup_{w^*\in\mathscr{W}} \int_{\Gamma_n} \|P(t)\| \, \exp\{c_\alpha^3 \tilde{\boldsymbol{S}}_{\boldsymbol{nw}}(t, w^*)\}\phi_V(t)dt > \epsilon\right] = o(1),$$

by B.9. Since both the numerator and the denominator on the left hand side in (40) are linear in $\hat{\boldsymbol{\pi}}_{\boldsymbol{0}}$, $\boldsymbol{\partial_\theta} \hat{\pi}(\boldsymbol{\theta_0})$, which are consistent for π_0, $\partial_\theta \pi(\theta_0)$, respectively, the stated result holds. □

APPENDIX G: THEOREM 3

LEMMA G.1. *Let $\mathfrak{d}_H^*(\theta, \eta) = \mathfrak{d}_H\{\theta, \Theta_0(\eta)\}$, where $\mathfrak{d}_H$ is the Hausdorff distance. The correspondence (mapping) $\bar{\Theta}^* : N \times \mathbb{R} \to \Theta$ given by $\bar{\Theta}^*(\eta, c) = \{\theta \in \Theta : \mathfrak{d}_H^*(\theta, \eta) \ge c\}$ is continuous in η in the sense of Berge (1963, page 109).*

PROOF. We first show that the correspondence $\Theta_0 : N \to \Theta$ given by $\Theta_0(\eta) = \operatorname{argmax}_{\theta\in\Theta} \tilde{Q}(\theta, \eta)$ is continuous in η. Upper hemicontinuity of Θ_0 follows by the maximum theorem (Berge, 1963, page 116) in view of the continuity of $\tilde{Q}$, because the correspondence $\Theta_0^* : N \to \Theta$ given by $\Theta_0^*(\eta) = \Theta$ is continuous due to the compactness of Θ. For lower hemicontinuity, let $C = \max_{\theta\in\Theta,\eta\in N}\left\|\partial_\eta \tilde{Q}(\theta, \eta)\right\| < \infty$ by assumption K such that $\sup_{\theta\in\Theta}\left|\tilde{Q}(\theta, \tilde{\eta}) - \tilde{Q}(\theta, \eta)\right| \le C\|\eta - \tilde{\eta}\|$. It then suffices to show that for all $\tilde{\eta} \in N$, all $\tilde{\theta} \in \Theta_0(\tilde{\eta})$, and all $\epsilon > 0$, $\|\eta - \tilde{\eta}\| < \epsilon/2C$ implies $Q(\tilde{\theta}, \eta) > -\epsilon$ (recall that $Q(\theta, \eta) = \tilde{Q}(\theta, \eta) - \max_{\theta\in\Theta} \tilde{Q}(\theta, \eta)$). Take $\theta \in \Theta_0(\eta)$ and note that

$$Q(\tilde{\theta}, \eta) = \tilde{Q}(\tilde{\theta}, \eta) - \tilde{Q}(\theta, \eta) = \{\tilde{Q}(\tilde{\theta}, \eta) - \tilde{Q}(\tilde{\theta}, \tilde{\eta})\} - Q(\theta, \tilde{\eta}) + \{\tilde{Q}(\theta, \tilde{\eta}) - \tilde{Q}(\theta, \eta)\} > -\epsilon.$$

So $\Theta_0(\eta)$ is a continuous correspondence and hence by the maximum theorem, $\mathfrak{d}_H^*(\theta, \eta)$ is continuous in η and $\bar{\Theta}^*(\eta, c)$ is an upper hemicontinuous correspondence in η. It remains to be shown that $\bar{\Theta}^*(\eta, c)$ is also lower hemicontinuous in η. Choose an arbitrary $\bar{\eta} \in N$ and $\bar{\theta} \in \bar{\Theta}^*(\bar{\eta}, c)$. Now, for any sequence $\{\eta_j\}$ converging to $\bar{\eta}$,

$$\min_{\breve{\theta}\in\Theta_0(\eta_j)} \|\bar{\theta} - \breve{\theta}\| \to \min_{\breve{\theta}\in\Theta_0(\bar{\eta})} \|\bar{\theta} - \breve{\theta}\| = c^* \ge c. \tag{43}$$

Take θ_j^* from the set of minimizers on the left hand side in (43) and set $\theta_j = \theta_j^* + c^*(\bar{\theta} - \theta_j^*)/\|\bar{\theta} - \theta_j^*\|$. By assumption I, $\Theta_0(\eta_j)$ is convex and hence $\theta_j \in \bar{\Theta}^*(\eta_j, c)$. Finally, $\|\theta_j - \bar{\theta}\| = \left|\|\theta_j^* - \bar{\theta}\| - c^*\right| \to 0$ by (43). □

LEMMA G.2.

$$\sup_{\eta^*,\eta\in N}\frac{\int \mathfrak{d}_H^*(\theta,\eta)\pi(\theta)\exp\{\alpha_n^2\boldsymbol{S_n}(\theta,\eta^*)+\alpha_n^2 Q(\theta,\eta)\}d\theta}{\int \pi(\theta)\exp\{\alpha_n^2\boldsymbol{S_n}(\theta,\eta^*)+\alpha_n^2 Q(\theta,\eta)\}d\theta}=o_p(1). \tag{44}$$

PROOF. Denote the left hand side in (44) by $\boldsymbol{I}$. Then for any $\epsilon^*,\epsilon>0$,

$$\mathbb{P}(\boldsymbol{I}>2\epsilon^*)\le\mathbb{P}\Big\{\sup_{\theta\in\Theta}\sup_{\eta^*\in N}|\boldsymbol{S_n}(\theta,\eta^*)|>\epsilon\Big\}+\mathbb{P}\Big\{\boldsymbol{I}>\epsilon^*,\ \sup_{\theta\in\Theta}\sup_{\eta^*\in N}|\boldsymbol{S_n}(\theta,\eta^*)|\le\epsilon\Big\},$$

where the first right hand side term is $o(1)$ by assumptions F, G, and M, and the uniform law of large numbers (e.g. van der Vaart and Wellner (1996, theorem 2.4.3)). We will deal with the second right hand side term below.

For any $c>0$ and $\eta\in N$, let $\bar\Theta(\eta,c)=\{\theta\in\Theta:\mathfrak{d}_H^*(\theta,\eta)\le c\}$, which is compact, and let $\bar\Theta^c(\eta,c)$ be its complement relative to Θ. Let further $c_q=c_q(c)=-\sup_{\eta\in N}\sup_{\theta\in\bar\Theta^c(\eta,c)}Q(\theta,\eta)>0$. Note that c_q is finite because $\bar\Theta^c(\eta,c)\subset\bar\Theta^*(\eta,c)$ and $\sup_{\theta\in\bar\Theta^*(\eta,c)}Q(\theta,\eta)$ is continuous in η by the maximum theorem and G.1. Also, c_q cannot be equal to zero since $\bar\Theta^c(\eta,c)\subset\bar\Theta^*(\eta,c/2)$, which is compact and continuous (as a correspondence) in η by G.1, such that for some $\eta^*\in N$ and some $\theta^*\in\bar\Theta^c(\eta^*,c)$, $Q(\theta^*,\eta^*)=0$, which is at odds with the definition of $\bar\Theta^c(\eta,c)$. Now, by assumption H, for $\lambda^\dagger=\max_{\eta\in N}\max_{\theta\in\Theta}\lambda^+(\theta,\eta)$, some $C<\infty$ and sufficiently large n,

$$\inf_{\eta\in N}\int_{\bar\Theta(\eta,c)}\pi(\theta)\exp\{-\alpha_n^2\epsilon+\alpha_n^2Q(\theta,\eta)\}d\theta\ge$$
$$\exp(-\alpha_n^2\epsilon)\inf_{\eta\in N}\int_{\bar\Theta(\eta,c)}\pi(\theta)\exp\{-\alpha_n^2\lambda^\dagger\|\theta-\theta_0(\eta)\|^2\}d\theta>\exp(-2\alpha_n^2\epsilon).$$

Finally, if $\sup_{\theta\in\Theta}\sup_{\eta^*\in N}|\boldsymbol{S_n}(\theta,\eta^*)|\le\epsilon$ then

$$\boldsymbol{I}\le c+\sup_{\eta\in N}\frac{\int_{\bar\Theta^c(\eta,c)}\mathfrak{d}_H^*(\theta,\eta)\pi(\theta)\exp\{\alpha_n^2\epsilon+\alpha_n^2Q(\theta,\eta)\}d\theta}{\int\pi(\theta)\exp\{-\alpha_n^2\epsilon+\alpha_n^2Q(\theta,\eta)\}d\theta}\le$$
$$c+\exp(3\alpha_n^2\epsilon)\sup_{\eta\in N}\int_{\bar\Theta^c(\eta,c)}\exp\{\alpha_n^2Q(\theta,\eta)\}d\theta\le c+C\exp\{\alpha_n^2(3\epsilon-c_q)\}=c+o(1),$$

provided that ϵ is chosen less than $c_q/3$. Finally, let $c\downarrow 0$. □

LEMMA G.3. *Suppose that for any fixed* $w\in\mathscr{W}$ *and some sequence* $\{\mathfrak{r}_n\}$,

$$\mathfrak{r}_n\{\hat{\boldsymbol{\theta}}^*(w^*,w)-\theta_0\}\xrightarrow{w}\mathbb{Z}+\mathscr{C}w, \tag{45}$$

in $\mathbb{L}^\infty(\mathscr{W})$, *where* $\mathbb{Z}$ *does not depend on* w^*,w. *Then* $\mathfrak{r}_n(\hat{\boldsymbol{\eta}}-\eta_0)\xrightarrow{d}-\mathscr{C}^{-1}\mathbb{Z}$.

Proof. Note that $\hat{\boldsymbol{w}} = \mathrm{r}_n(\hat{\boldsymbol{\eta}} - \eta_0)$ is for sufficiently large n with probability approaching one a solution to $\hat{\boldsymbol{\theta}}^{*}(\hat{\boldsymbol{w}}, \hat{\boldsymbol{w}}) = \theta_0$. Let $\hat{\boldsymbol{\psi}}(w^*, w) = \mathrm{r}_n\{\hat{\boldsymbol{\theta}}^{*}(w^*, w) - \theta_0\}$. By Chebyshev's order inequality (Steele, 2004, problem 5.2) and assumption K, $\hat{\boldsymbol{\theta}}^{*}$ is increasing in every element of w. Hence, for any closed hypercube $\mathscr{W}_c \subset \mathscr{W}$ using element–wise inequalities,

$$\mathbb{P}(\hat{\boldsymbol{w}} \in \mathscr{W}_c) = \mathbb{P}\{\exists w^* \in \mathscr{W}_c : \hat{\boldsymbol{\psi}}(w^*, w^*) = 0\} = \mathbb{P}\Big\{\inf_{w^* \in \mathscr{W}_c} \hat{\boldsymbol{\psi}}(w^*, w^*) \le 0 \le \sup_{w^* \in \mathscr{W}_c} \hat{\boldsymbol{\psi}}(w^*, w^*)\Big\}. \tag{46}$$

From (46) it follows that if $\underline{w}, \bar{w}$ represent respectively the vectors of minima and maxima of $\mathscr{W}_c$ then

$$\mathbb{P}\Big\{\sup_{w^* \in \mathscr{W}} \hat{\boldsymbol{\psi}}(w^*, \underline{w}) \le 0 \le \inf_{w^* \in \mathscr{W}} \hat{\boldsymbol{\psi}}(w^*, \bar{w})\Big\} \le \mathbb{P}(\hat{\boldsymbol{w}} \in \mathscr{W}_c) \le \mathbb{P}\Big\{\inf_{w^* \in \mathscr{W}} \hat{\boldsymbol{\psi}}(w^*, \underline{w}) \le 0 \le \sup_{w^* \in \mathscr{W}} \hat{\boldsymbol{\psi}}(w^*, \bar{w})\Big\}. \tag{47}$$

By (45) the limits of the majorant and minorant sides in (47) are equal and again by (47) hence

$$\lim_{n\to\infty} \mathbb{P}(\hat{\boldsymbol{w}} \in \mathscr{W}_c) = \lim_{n\to\infty} \mathbb{P}\{\hat{\boldsymbol{\psi}}(0, \underline{w}) \le 0 \le \hat{\boldsymbol{\psi}}(0, \bar{w})\} = \mathbb{P}(\mathbb{Z} \in -\mathscr{C}\mathscr{W}_c) - \mathbb{P}\big(-\mathscr{C}^{-1}\mathbb{Z} \in \mathscr{W}_c\big). \quad \square$$

Proof of theorem 3. For part (i) note that $\mathfrak{d}_H\{\theta_0(\eta_0), \Theta_0(\eta)\} = 0$ has a unique solution at η_0 by assumption L. Since $\mathfrak{d}_H\{\theta_0(\eta_0), \Theta_0(\eta)\}$ is continuous in $\eta \in N$ by the maximum theorem and G.1, where N is compact, we know that for any $\epsilon > 0$ there is an $\epsilon^* > 0$ such that

$$\mathbb{P}\big(\|\hat{\boldsymbol{\eta}} - \eta_0\| \ge \epsilon\big) \le \mathbb{P}\big[\mathfrak{d}_H\{\theta_0(\eta_0), \Theta_0(\hat{\boldsymbol{\eta}})\} \ge 2\epsilon^*\big] \le \mathbb{P}\big[\mathfrak{d}_H\{\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\eta}}), \Theta_0(\hat{\boldsymbol{\eta}})\} \ge \epsilon^*\big] + \mathbb{P}\{\|\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\eta}}) - \theta_0(\eta_0)\| \ge \epsilon^*\}. \tag{48}$$

The first right hand side term in (48) is $o(1)$ by G.2 and the second right hand side term can by definition be no greater than $\mathbb{P}\{\|\hat{\boldsymbol{\theta}}(\eta_0) - \theta_0(\eta_0)\| \ge \epsilon^*\} = o(1)$, again by G.2.

Now, observe that for any divergent sequence $\{\mathrm{r}_n\}$ and $\hat{\boldsymbol{w}} = \mathrm{r}_n(\hat{\boldsymbol{\eta}} - \eta_0)$ we have (with probability approaching one),

$$\mathrm{r}_n\{\hat{\boldsymbol{\theta}}^{*}(w^*, w) - \theta_0\} = \mathrm{r}_n\{\hat{\boldsymbol{\theta}}^{*}(w^*, w) - \theta_{0n}(w)\} + \mathrm{r}_n\{\theta_0(\eta_0 + w/\mathrm{r}_n) - \theta_0(\eta_0)\}. \tag{49}$$

For (ii), suppose that the convergence rate r_n of $\hat{\boldsymbol{\eta}}$ is slower than $\sqrt[3]{n}$. By G.3 it suffices to show that for any $w \in \mathscr{W}$, the second right hand side term in (49) converges to $\mathscr{C}w$ for any $w \in \mathscr{W}$, which was established in B.1, and that the first right hand side term converges to zero uniformly in $w^* \in \mathscr{W}$. For the second requirement, replace $\sqrt[3]{n}, \tilde{\alpha}_n$ in appendix C by r_n and replace $\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}$ by $\sqrt{\mathrm{r}_n^3/n}\tilde{\boldsymbol{S}}_{\boldsymbol{nw}}$, such that $\mathbb{X} = 0$ a.s., where $\mathbb{X}$ is as defined in C.1. Consequently, following the same steps as in (ii) of theorem 1 at the end of appendix C delivers the desired result.

Finally, the remaining results follow immediately from the corresponding parts of theorem 1 using $\mathrm{r}_n = \sqrt[3]{n}$ for parts (iii) and (iv) and $\mathrm{r}_n = \sqrt{n/\alpha_n}$ for (v). $\square$

APPENDIX H: THEOREM 4

In this section let μ be the probability measure $N(0, V^{-1})$ on $(\mathbb{R}^d, \mathscr{B})$, where $\mathscr{B}$ is the Borel σ–algebra.

LEMMA H.1. *For all sufficiently large n, $\limsup_{\|t\|\to\infty} |\hat{\mathbb{G}}(t)|/c_t = 0$ with probability one.*

PROOF. We will show that for any ϵ, $\mathbb{P}\{\limsup_{\|t\|\to\infty} |\hat{\mathbb{G}}(t)|/c_t > \epsilon\} = 0$. By the Borel–Cantelli lemma, it suffices to show that

$$\sum_{s=1}^{\infty} \mathbb{P}\Big\{\sup_{s-1\le\|t\|<s} |\hat{\mathbb{G}}(t)|/c_t > \epsilon\Big\} < \infty. \tag{50}$$

Repeating the latter part of the proof of B.5, the left hand side of (50) is by assumption P bounded above by

$$\sum_{s=1}^{\infty} \mathbb{P}\Big\{\sup_{\|t\|\le 1} |\hat{\mathbb{G}}(t)| > \epsilon\sqrt{s}\Big\} \le \frac{1}{\epsilon^4}\mathbb{E}\mathbb{E}\Big\{\sup_{\|t\|\le 1} \hat{\mathbb{G}}^{\mathbf{4}}(t)\big|\hat{\boldsymbol{H}}\Big\}\sum_{s=1}^{\infty}\frac{1}{s^2} \le \frac{C}{\epsilon^4}\mathbb{E}\Big[\mathbb{E}\Big\{\sup_{\|t\|\le 1} |\hat{\mathbb{G}}(t)|\big|\hat{\boldsymbol{H}}\Big\}\Big]^4, \tag{51}$$

where C is a generic constant; the last inequality follows from van der Vaart and Wellner (1996, proposition A.2.4) and the fact that $\sum_{s=1}^{\infty}(1/s^2) < \infty$. Now, by van der Vaart and Wellner (1996, theorem 2.2.8), for some $0 < \tilde{C} < \infty$, $\mathbb{E}\{\sup_{\|t\|\le 1} |\hat{\mathbb{G}}(t)|\big|\hat{\boldsymbol{H}}\} \le \tilde{C}\sup_{\|t\|=1}\sqrt{\hat{\boldsymbol{H}}(t,t)}$, because $\hat{\mathbb{G}}$ given $\hat{\boldsymbol{H}}$ is sub–Gaussian with respect to the semimetric $d^2_{VW}(t,s) = \hat{\boldsymbol{H}}(t,t)+\hat{\boldsymbol{H}}(s,s)-2\hat{\boldsymbol{H}}(t,s) = \hat{\boldsymbol{H}}(t-s,t-s) \le \|t-s\|\sup_{\|t\|=1}\hat{\boldsymbol{H}}(t,t)$ by assumption P. Therefore, the right hand side of (51) is bounded by $\tilde{C}^4\mathbb{E}\{\sup_{\|t\|=1}\hat{\boldsymbol{H}}^{\mathbf{2}}(t,t)\}$, which is finite by assumption Q. □

LEMMA H.2. $\hat{\mathbb{G}} \xrightarrow{w} \mathbb{G}$ *in* $\mathbb{L}^2(\mathbb{R}^d, \mathscr{B}, \mu)$.

PROOF. We use Cremers and Kadelka (1986, theorem 2). Note first that $\hat{\mathbb{G}}$ and $\mathbb{G}$ are contained in $\mathbb{L}^2(\mathbb{R}^d, \mathscr{B}, \mu)$ by H.1 and similar arguments using $\mathbb{E}\{\sup_{\|t\|\le 1}|\mathbb{G}(t)|\} < \infty$. Further, convergence in distribution of $\hat{\mathbb{G}}$ to $\mathbb{G}$ for finite marginals holds by construction. So, by theorem 2 of Cremers and Kadelka, it suffices to show that for $\boldsymbol{T} \sim N(0, V^{-1})$ independent of $\hat{\mathbb{G}}$, $\lim_{C\to\infty}\limsup_{n\to\infty}\mathbb{E}\big[|\hat{\mathbb{G}}(\boldsymbol{T})|\mathbb{1}\{|\hat{\mathbb{G}}(\boldsymbol{T})| > C\}\big] = 0$. For this purpose, note that

$$\limsup_{n\to\infty}\mathbb{E}\big[|\hat{\mathbb{G}}(\boldsymbol{T})|\mathbb{1}\{|\hat{\mathbb{G}}(\boldsymbol{T})| > C\}\big] \le \limsup_{n\to\infty}\mathbb{E}\hat{\mathbb{G}}^{\mathbf{2}}(\boldsymbol{T})/C = \limsup_{n\to\infty}\mathbb{E}\hat{\boldsymbol{H}}(\boldsymbol{T},\boldsymbol{T})/C,$$

where $\limsup\{\mathbb{E}\hat{\boldsymbol{H}}(\boldsymbol{T},\boldsymbol{T})\}^2 \le \mathbb{E}\|\boldsymbol{T}\|^2\limsup\mathbb{E}\{\sup_{\|t\|=1}\hat{\boldsymbol{H}}^{\mathbf{2}}(t,t)\} < \infty$ by assumption Q. □

LEMMA H.3. *For any polynomial P and for $j, \tilde{j} \in \{0,1\}$,*

$$\int \|P(t)\|\hat{\mathbb{G}}^{j}(t)\exp\{\tilde{j}\hat{\mathbb{G}}(t)\}\phi_V(t)dt \xrightarrow{d} \int \|P(t)\|\mathbb{G}^{j}(t)\exp\{\tilde{j}\mathbb{G}(t)\}\phi_V(t)dt.$$

PROOF. We only consider the case $j = \tilde{j} = 1$; the other cases are similar. Let $\mathscr{E} = \{\kappa \in \mathbb{L}^2(\mathbb{R}^d, \mathscr{B}, \mu) : \limsup_{\|t\|\to\infty} |\kappa(t)|/c_t = 0\}$. Since the functional $\mathfrak{M} : \mathscr{E} \to \mathbb{R}$ given by $\mathscr{M}(\kappa) = \int \|P(t)\|\kappa(t)\exp\{\kappa(t)\}d\mu(t)$ is continuous, the stated result then follows from H.1 and H.2 and the continuous mapping theorem (van der Vaart and Wellner, 1996, theorem 1.11.1). □

PROOF OF THEOREM 4. The first half of the theorem statement is a reformulation of parts of theorem 1 with implicit definition of $\mathbb{J}_{\boldsymbol{\alpha}}$ in each case, so we only need to establish that $\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}}$ has the same limiting distribution as the corresponding $\mathbb{J}_{\boldsymbol{\alpha}}(0)$ in all three cases.

First, suppose that $\sqrt[3]{n}/\alpha_n = o(1)$. Then by H.2 for $\hat{\mathbb{C}}(t) = \hat{\mathbb{G}}(t) - t^\intercal \hat{V} t/2$, $\hat{\mathbb{C}} \overset{w}{\to} \mathbb{C}$ in $\mathbb{L}^2(\mathbb{R}^d, \mathscr{B}, \mu)$. Then follow the same steps as in appendix C using $\hat{\mathbb{C}}$ in lieu of $\tilde{\boldsymbol{L}}_{\boldsymbol{nw}}$

Next, suppose that $\alpha_n = c_\alpha^2 \sqrt[3]{n}$. The maximum and minimum in the definition of $\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}}$ and the theorem statement have the effect of multiplying both sides by a constant, so suppose without loss of generality that $c_\alpha > 1$. Then carry out the substitution $t \leftarrow \beta_n^{2/3} t$ to obtain

$$\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}} = \int t\exp\{c_\alpha^3 \hat{\mathbb{G}}(t)\}\phi_{\hat{V}}(t)dt \Big/ c_\alpha^2 \int \exp\{c_\alpha^3 \hat{\mathbb{G}}(t)\}\phi_{\hat{V}}(t)dt.$$

Apply H.3 plus a minor correction for the fact that $\hat{V}$ depends on data to obtain the limit distribution indicated in (7).

Finally suppose that $\alpha_n = o(\sqrt[3]{n})$. Use the substitution $t \leftarrow \beta_n^{2/3} t$ to obtain

$$\hat{\boldsymbol{\Psi}}_{\boldsymbol{n}} = \int t\exp\{\beta_n \hat{\mathbb{G}}(t)\}\phi_{\hat{V}}(t)dt \Big/ \beta_n \int \exp\{\beta_n \hat{\mathbb{G}}(t)\}\phi_{\hat{V}}(t)dt.$$

Now, for $j = 0, 1$ and some $c \in [0, 1]$,

$$\int \left(\frac{t}{\beta_n}\right)^j \exp\{\beta_n \hat{\mathbb{G}}(t)\}\phi_{\hat{V}}(t)dt = \int \{t\hat{\mathbb{G}}(t)\}^j \phi_{\hat{V}}(t)dt + \frac{\beta_n}{(j+1)!}\int t^j \hat{\mathbb{G}}^{\boldsymbol{j+1}}(t)\exp\{c\beta_n \hat{\mathbb{G}}(t)\}\phi_{\hat{V}}(t)dt. \tag{52}$$

The second right hand side term in (52) is for some constant C with probability approaching one bounded in norm by $C\beta_n \int \|t\|^j |\hat{\mathbb{G}}(t)|^j \exp\{\hat{\mathbb{G}}(t) - \lambda_- \|t\|^2/4\}dt = o(1)O_p(1) = o_p(1)$ by H.1 and H.2. For $j = 0$ the first right hand side in (52) equals one and for $j = 1$ we have after a minor correction for the fact that $\hat{V}$ depends on data, using H.3 that $\int t\hat{\mathbb{G}}(t)\phi_{\hat{V}}(t)dt \overset{d}{\to} \int t\mathbb{G}(t)\phi_V(t)dt \sim N(0, \mathscr{V})$. □

DEPARTMENT OF ECONOMICS, THE PENNSYLVANIA STATE UNIVERSITY,
303 KERN GRADUATE BUILDING, UNIVERSITY PARK 16802,
E-MAIL: sjun@psu.edu; joris@psu.edu

DEPARTMENT OF ECONOMICS, UNIVERSITY OF TORONTO,
MAX GLUSKIN HOUSE, 150 ST. GEORGE STREET, 232, TORONTO, ON,
E-MAIL: yuanyuan.wan@utoronto.ca