Econometric Theory, **23**, 2007, 281–308. Printed in the United States of America.
DOI: 10.1017/S0266466607070120

EFFICIENT SEMIPARAMETRIC ESTIMATION OF DURATION MODELS WITH UNOBSERVED HETEROGENEITY

PETER BEARSE
University of North Carolina at Greensboro

JOSÉ CANALS-CERDÁ
University of Colorado

PAUL RILSTONE
York University

This paper develops a new semiparametric approach for the estimation of hazard functions in the presence of unobserved heterogeneity. The hazard function is specified parametrically, whereas the distribution of the unobserved heterogeneity is indirectly estimated using the method of kernels. The semiparametric efficiency bounds are derived. The estimator obtains these bounds in large samples.

1. INTRODUCTION

Duration analysis is an integral component of empirical and theoretical econometrics. Of particular concern is the treatment of unobserved heterogeneity. Authors such as Lancaster (1979), Heckman and Singer (1984), and Baker and Melino (2000) have shown that ignoring unobserved heterogeneity can lead to biased estimates of the hazard function and that different specifications for the distribution of the unobserved heterogeneity can lead to very different estimates of the hazard function. As a consequence, several authors have considered estimation of these models without specifying any parametric form for the unobserved heterogeneity.

The Heckman and Singer (1984) approach consistently estimates the parameters of the hazard function. To our knowledge, no one has worked out the asymptotic distribution of their estimator. Most applied researchers base their inferences conditional on estimates of the unobserved heterogeneity. This can lead to incorrect inferences.[1] Gurmu, Rilstone, and Stern (1995) provide for

The authors thank Yongmiao Chen, James Heckman, Hidehiko Ichimura, Tony Lancaster, Qi Li, Adrian Pagan, Barry Smith, two anonymous referees, and the co-editor for helpful input. We particularly thank Steven Stern, who prompted us toward this line of research. Any errors are those of the authors. Research funding for Rilstone was provided by the Social Sciences and Humanities Research Council of Canada. Address correspondence to Paul Rilstone, Department of Economics, York University, Toronto, Ontario, Canada M3J 1P3; e-mail: pril@yorku.ca.

© 2007 Cambridge University Press 0266-4666/07 $12.00

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

consistent nonparametric estimation of both the distribution of the unobserved heterogeneity and the hazard function. However, the asymptotic distribution of their estimator is also unknown. Honoré (1990) and Van der Vaart (1996) have proposed semiparametric maximum likelihood estimators for some specific hazard rate models, and hence their asymptotic results are somewhat limited.

This paper introduces a new estimator of hazard rates without imposing functional form restrictions on the distribution of the unobserved heterogeneity. To our knowledge, this is the first efficient semiparametric estimator for the class of models considered. Unlike Heckman and Singer (1984), we derive the asymptotic distribution of our estimator and show that it obtains the semiparametric efficiency bounds. One drawback is with respect to the usual Weibull model, for which Hahn (1994) has shown that there exists no $\sqrt{N}$-consistent estimator. However, the asymptotic results that we develop do apply for a wide class of duration models, including non-Weibull and certain Weibull-like models.

Horowitz (1996) has proposed an $\sqrt{N}$-consistent estimator for a certain proportional hazards model with unobserved heterogeneity. This estimator is more general than ours in that it does not specify a functional form for the baseline hazard. On the other hand, because it leaves the baseline hazard unspecified, it will generally not be efficient. The Nielsen, Linton, and Bickel (1998) approach parameterizes the baseline hazard rate but leaves the dependency on observed covariates unspecified.

Our approach is quite straightforward. The likelihood function for continuous duration models with unobserved multiplicative heterogeneity is written as a function of the hazard rate, whose parametric form is known, and the density function of the integrated hazard rate. We propose using nonparametric (kernel) estimates of this density function in a two-step procedure to obtain an asymptotically efficient estimator.

The following section discusses the estimation procedure in greater detail. Section 3 presents the asymptotic properties of the estimator. The estimator and consistent standard errors can be calculated using the optimization procedures of most intermediate-level computer packages. Section 4 discusses additional estimation issues and the semiparametric efficiency bounds. Section 5 concludes. Proofs are collected in three Appendixes.[2]

2. THE ESTIMATOR

We focus on estimating mixed hazard rate models of the form $\lambda_i(\beta)\nu_i = \lambda(t_i, x_i;\beta)\nu_i$ where t_i is the duration variable, x_i is a vector of observable covariates, $\beta \in \Re^k$ is the parameter of interest, and ν_i is an unobservable nonnegative error with unit mean and is independent of x_i.

Consider first a benchmark estimator of β, supposing that we know the distribution of the unobserved heterogeneity. The survival function of t_i, conditional on x_i, is written

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

$$S_t(t_i|x_i;\beta) = \int e^{-\Lambda(t_i,x_i;\beta)\nu_i}\, dF_\nu(\nu_i) = \mathcal{L}[\Lambda_i(\beta)], \tag{2.1}$$

where $\Lambda_i(\beta) = \Lambda(t_i, x_i;\beta) = \int_0^{t_i} \lambda(u, x_i;\beta)\, du$ is the integrated hazard, F_ν is the distribution function of ν_i, and $\mathcal{L}$ is the corresponding Laplace transform. With F_ν (or $\mathcal{L}$) known, the conditional density function of t_i can be written as

$$f_t(t_i|x_i;\beta) = -\frac{d}{dt_i} S_t(t_i|x_i;\beta) = -\lambda(t_i, x_i;\beta)\mathcal{L}^{(1)}[\Lambda_i(\beta)], \tag{2.2}$$

where $\mathcal{L}^{(1)}$ is the derivative of $\mathcal{L}$ and the single observation log-likelihood function is

$$\log l_i(\beta) = \log \lambda_i(\beta) + \log - \mathcal{L}^{(1)}[\Lambda_i(\beta)]. \tag{2.3}$$

An efficient, albeit infeasible, estimator is obtained by maximizing the log-likelihood function based on (2.3). Alternatively, if a preliminary $\sqrt{N}$-consistent estimator is available, an asymptotically efficient two-step estimator can be obtained using the score function and Hessian based on the log-likelihood in (2.3).

To get an intuitive understanding of our approach, let S and f denote the survivor and density functions of Λ. Following Lancaster (1990), $f = -\mathcal{L}^{(1)}$ or $S = \mathcal{L}$. To see this, simply transform in the usual way from the conditional density of t to obtain the conditional density of Λ:

$$\begin{aligned} f(\Lambda_i(\beta)|x_i) &= f_t(t_i|x_i;\beta)\left(\frac{d\Lambda(t_i, x_i;\beta)}{dt}\right)^{-1} \\ &= -\lambda_i(\beta)L^{(1)}[\Lambda(t_i, x_i;\beta)]\,\frac{1}{\lambda_i(\beta)} \\ &= -\mathcal{L}^{(1)}[\Lambda_i(\beta)] \\ &= f(\Lambda_i(\beta)). \end{aligned} \tag{2.4}$$

The last expression follows because $\mathcal{L}^{(1)}$ is a function only of Λ. Alternatively, note that, conditional on ν_i, the integrated hazard has a unit exponential distribution with location parameter ν_i.[3] The survivor function is thus obtained by marginalizing out ν_i. This yields $\mathcal{L}[\Lambda_i]$ as in equation (2.1). The density function of Λ_i is thus given by $-\mathcal{L}^{(1)}[\Lambda_i]$.

Given this, a maximum likelihood approach could proceed by substituting nonparametric estimates of $f(\Lambda_i)$ into equations (2.3) and (2.4). This neglects several technical difficulties. The issues involved are similar to the Klein and Spady (1993) treatment of the binary choice model. Let F_ϵ denote the error term from a standard latent variable model. If F_ϵ were known, the single observation log-likelihood function would be written as $d_i \log F_\epsilon(x_i'\beta) + (1 - d_i) \times$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

$\log(1 - F_\epsilon(x_i'\beta))$ where d_i is an indicator function and x_i is a vector of covariates. The term F_ϵ cannot be estimated directly, but if $F_\epsilon(x_i'\beta) = \mathcal{E}[d_i|x_i'\beta]$ one can proceed by replacing $F_\epsilon(x_i'\beta)$ in the likelihood function by nonparametric estimates of $\mathcal{E}[d_i|x_i'\beta]$. In our case, although ν_i is unobservable, we can exploit the fact that $f = -\mathcal{L}^{(1)}$ and use nonparametric estimates of $f(\Lambda)$ in the likelihood function.

Rearranging (2.2) and (2.4), we can write

$$f_t(t_i|x_i;\beta) = \lambda(t_i, x_i;\beta) f(\Lambda_i(\beta)). \tag{2.5}$$

An interesting by-product of this result is that we can express the hazard function as

$$\lambda(t_i, x_i;\beta) = \frac{f_t(t_i|x_i;\beta)}{f(\Lambda_i(\beta))}. \tag{2.6}$$

Without unobserved heterogeneity, the hazard function is defined as the ratio of the density of the duration variable to its survival function. The presence of $f(\Lambda)$ in the denominator may seem odd until one notes, again, that in the absence of heterogeneity, the integrated hazard has an exponential distribution, in which case the survivor function and the density function are identical.

In essence, our estimator consists of replacing $-\mathcal{L}^{(1)}$ in (2.3) by kernel estimates of the density function of $\Lambda_i(\beta)$. However, there are several technical problems when one does this. First, the density of Λ_i is nonzero at the left edge of its support ($f(0) = \mathcal{E}[\nu] > 0$), and the usual kernel estimates will be biased.[4] To obtain an estimator with higher order derivatives and bias reduction we take a density having support $\Re$ and multiply it by a polynomial chosen to reduce bias. Thus, we estimate the density of Λ_i using

$$\begin{aligned}
\hat{f}(\Lambda_i) &= \frac{1}{N-1}\frac{1}{\gamma}\sum_{j\neq i} K_{ij}, \\
K_{ij} &= \mu(\Lambda_i/\gamma)^T P\left(\frac{\Lambda_i - \Lambda_j}{\gamma}\right) K^*\left(\frac{\Lambda_i - \Lambda_j}{\gamma}\right), \\
P(w)^T &= \left(1 \quad w \quad \frac{w^2}{2!} \quad \cdots \quad \frac{w^{s-1}}{(s-1)!}\right), \\
\mu(\Lambda_i/\gamma) &= \left(\int_{-\infty}^{\Lambda_i/\gamma} K^*(w)P(w)P(w)^T\,dw\right)^{-1} \iota_s,
\end{aligned} \tag{2.7}$$

where $\iota_s^T = (1 \quad 0 \quad 0 \quad \cdots \quad 0)$, N is the sample size, $\gamma = \gamma(N) \downarrow 0$ is the window width, and K^* is the "inner" kernel or smoothing function and a superscript T indicates matrix transposition.

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

The term $\hat{f}$ can be nonpositive, leading to analytical and numerical difficulties. Like Klein and Spady (1993), we circumvent this problem by using $\tilde{f} = (\hat{f}^2)^{1/2}$ rather than $\hat{f}$. Under the usual regularity conditions, $\hat{f}$ is, pointwise, $\sqrt{N\gamma}$-consistent. Although many large-sample results exist in the literature, we are unable to find some that are usefully applied to our problem given the dependence of $\hat{f}$ and its derivatives on β. As a result, Appendix C derives a number of results that are applicable in our context.[5]

A further issue is that f can get arbitrarily small so that the corresponding likelihood contribution need to be downweighted. Also, some adjustment must be made to the nonparametric estimates themselves to keep the score function bounded. We handle these problems using two smoothed trimming functions as in Klein and Spady (1993). The first downweights observations with small $\tilde{f}$; the second adds a small trimming factor to the denominator of the score function in this situation. See Klein and Spady (1993) for a more detailed explanation. A serious technical difficulty (which Klein and Spady and several authors with similar estimation problems do not have) is that our estimand can get arbitrarily close to zero. (In all cases $f \to 0$ as $t \to \infty$.) Klein and Spady and others suppose that the support of their functions is such that their estimands are strictly positive. The strict positivity assumption is not plausible in our case.

We presume that a $\sqrt{N}$-consistent estimate ($\tilde{\beta}$) is used for trimming. Many duration models lend themselves to $\sqrt{N}$-consistent estimation. For example, a mixed proportional hazard (MPH) model with $\Lambda_i(\alpha, \delta) = \Phi(t_i; \alpha)\exp(-x_i\delta)$ can be written as $\log \Phi(t_i; \alpha) = x_i\delta + u_i$ where Φ is the integrated baseline hazard and u_i is a residual independent of x_i (Kiefer, 1988). Method of moments can consistently estimate (α, δ) using functions of the x_i's as instruments. Similarly, models of the type in (2.1) can be written as $\log \Lambda(t_i, x_i; \beta) = u_i$ where u_i is a residual independent of x_i. As before, we can specify a vector of instruments using functions of the x_i's. Finally, a $\sqrt{N}$-consistent estimator of the vector of parameters β can be obtained based on the conditional moment restriction that $\mathcal{E}[u_i - \mu | x_i] = 0$ with $\mu = \mathcal{E}[u_i]$.[6] However, because method of moments uses only a finite number of moment restrictions, and does not make use of all the available information, it is not efficient.

Given this, the proposed estimator is written in two-step form:

$$\hat{\beta} = \tilde{\beta} - \hat{H}_N(\tilde{\beta})^{-1}\hat{S}_N(\tilde{\beta}), \tag{2.8}$$

$$\hat{S}_N(\beta) = \frac{1}{N}\sum \hat{\chi}_1(\hat{f}_i(\beta))\hat{S}_i(\beta), \qquad \hat{S}_i(\beta) = \frac{\dot{\lambda}(\beta)\hat{f}_i(\beta) + \nabla\hat{f}_i(\beta)}{\tilde{D}_i}, \tag{2.9}$$

$$\dot{\lambda}(\beta) = \frac{\nabla\lambda(\beta)}{\lambda(\beta)}, \qquad \tilde{D}_i = \tilde{f}_i(\beta) + 2\gamma^{\epsilon'/5}(1 - \chi_2(\hat{f}_i(\beta))); \tag{2.10}$$

$$\hat{H}_N(\beta) = \nabla\hat{S}_N(\beta), \tag{2.11}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

$$\chi_1(f) = \frac{1}{1 + \exp\left\{\dfrac{\gamma^{\epsilon/5} - f}{\gamma^{\epsilon/4}}\right\}}, \qquad \chi_2(f) = \frac{1}{1 + \exp\left\{\dfrac{\gamma^{\epsilon'/5} - f}{\gamma^{\epsilon'/4}}\right\}}. \tag{2.12}$$

The parameters ϵ and ϵ' are such that $0 < \epsilon < \epsilon'$ and satisfy some technical conditions spelled out in the following section. The term $\tilde{D}_i$ is defined to keep the denominator of $\hat{S}_i$ positive and bounded away from zero. In fact (see Remark A3 in Appendix A) $\tilde{D}_i \geq \gamma^{\epsilon'/5}$.

Matrix $\hat{H}_N(\tilde{\beta})$ could be replaced by any other weighting matrix that is consistent for (minus one times) the information matrix. Matrix $\hat{H}_N(\tilde{\beta})$ is a natural weighting matrix because we examine the properties of $\nabla \hat{S}_N(\tilde{\beta})$ in the demonstration of normality.[7]

We let $\mathcal{B} \subset \Re^k$ denote an open ball centered on the true value of β, β_0, with radius η for some $\eta > 0$. Regularity assumptions about functions of β are made to hold uniformly on $\mathcal{B}$. Note that, for large enough N, $\tilde{\beta} \in \mathcal{B}$ with probability approaching one. To avoid clutter, we leave this as understood in the derivations to follow.

3. ASYMPTOTIC RESULTS

In this section we establish the asymptotic normality of $\hat{\beta}$.[8] Let

$$\tilde{S}_i(\beta) = \nabla \log l_i(\beta) - \mathcal{E}[\nabla \log l_i(\beta) | \Lambda_i(\beta)], \qquad \Sigma(\beta) = \mathcal{E}[\tilde{S}_i(\beta)\tilde{S}_i(\beta)^T]. \tag{3.1}$$

As discussed subsequently, $\tilde{S}_i(\beta)$ is the "efficient score" for the class of models in this paper; $\Sigma(\beta_0)$ is the semiparametric efficiency bound; $0 \leq C < \infty$ denotes a generic constant; $B_i \geq 0$ refers to a generic dominating function. A random vector, $q_i(\beta)$, is B_i-dominated if, for all $\beta \in \mathcal{B}$, $\|q_i(\beta)\| \leq B_i$.

The following assumptions are used in this paper.

Assumption P.

(1) The observations, $\{z_i\}_{i=1}^N$, $z_i = (t_i, x_i)$, are independent and identically distributed (i.i.d.); $\lambda_i(\beta)$ and $\Lambda_i(\beta)$ are, respectively, twice and four times continuously differentiable with respect to β, at all points $\beta \in \mathcal{B}$; $\Sigma(\beta)$ is continuous in $\mathcal{B}$ and nonsingular at β_0; $\log \lambda_i$ and $\log f_i$ are Lipchitz-continuous with respect to β.

(2) For $j = 0,1,2$, the following functions (evaluated at $\beta = \beta_0$) vanish at 0 and ∞: $(\mathcal{E}[\nabla\Lambda_i|\Lambda_i]f_i)^{(j)}$, $(\mathcal{E}[\nabla^2\Lambda_i|\Lambda_i]f_i)^{(j)}$, $(\mathcal{E}[\nabla\Lambda_i\nabla\Lambda_i^T|\Lambda_i]f_i)^{(j)}$.

(3) The following functions (evaluated at $\beta = \beta_0$) have bounded derivatives with respect to Λ_i on $\Re^+$ to order $s + 2$: f_i, $\mathcal{E}[\nabla\Lambda_i|\Lambda_i]f_i$, $\mathcal{E}[\nabla^2\Lambda_i|\Lambda_i]f_i$, $\mathcal{E}[\nabla\Lambda_i\nabla\Lambda_i^T|\Lambda_i]f_i$.

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

(4) The following are B_i-dominated: $\|\nabla^{(j)}\Lambda_i(\beta)\|$, $j = 0,1,2,3,4$, $|\log \lambda_i|$, $|\dot{\lambda}_i|$, f_i, $|\log f_i|$, and $\|\nabla^{(j)}\Lambda_i/\Lambda_i\|$, $j = 1,2,3$.
(5) $\|\tilde{\beta} - \beta_0\| = O_P(1/\sqrt{N})$.

Assumption P(1) provides standard regularity conditions. A couple of remarks should be made with respect to identification. First, the identification of duration models with unobserved heterogeneity has received a large amount of attention. Elbers and Ridder (1982) and others have shown the identification of the MPH model under various regularity conditions such as $\mathcal{E}[\nu_i] = 1$. Ridder (1990) provides more general results on identification of generalized accelerated failure-time models. Neither of these papers uses any parametric restrictions. Because these results apply to a larger class of models than our own and we assume that an $\sqrt{N}$-consistent estimator exists we leave identification as implicit. Note that the density function of Λ_i is also the first derivative of the moment generating function of the unobserved heterogeneity. Because $\mathcal{E}[\nu_i] = f(0)$, this can be imposed by normalizing the estimates of $f(0)$ to one. Incorporating the equivalent constraint into the estimator becomes tedious analytically. An alternative is to impose a normalization on the integrated hazard. For example, for the proportional hazard case with $\Lambda(t,x;\beta) = \Phi(t;\alpha)\exp(x'\delta)$, it is often possible and sufficient to set $\Phi(1;\alpha) = 1$. With our identification condition we assume implicitly that this is done.

Assumption P(2) provides smoothness restrictions for bias reduction. Assumptions P(3) and (4) provide regularity conditions to simplify the asymptotic results. Assumption P(3) can be verified for most models using a Taylor expansion. It is used in the consistency proofs of Appendix C to put bounds on derivatives of kernel functions. Assumption P(5) has been discussed in Section 2.

Assumption K.

(1) $K^*: \Re \to \Re^+$ and $\int K^*(w)\,dw = 1$, and K^* is continuously differentiable to order 2.
(2) $K^*(w)|w|^q < C$, $|K^{*(1)}(w)||w|^q < C$, $|K^{*(2)}(w)||w|^q < C$, for all $q \leq 2s$.
(3) For all $w \geq 0$: $\|\mu(w)\| \leq C$, $\|w\mu(w)\| \leq C$, $\|w^j\mu^{(j)}(w)\| \leq C$, $j = 1,2,3$, and $\|\mu^{(1)}(w)\| \leq C$.
(4) The window width satisfies (a) $\sqrt{N}\gamma^{s-1-\epsilon'} \to 0$ and (b) $1/(N\gamma^{3+\epsilon'}) \to 0$.
(5) $\int|w|^qK^*(w)\,dw < C$, for all $q \leq 2s$, $\int|PK^*(w)|^2\,dw < C$, $\int|PK^*(w)^{(1)}|^2\,dw < C$.

Assumption K(1) implies that K^*, the "inner kernel," is a probability density on the real line. It allows one to easily verify that its partial moments exist. The restrictions in Assumption K(2) bound K^* and polynomial multiples of it. The restrictions on the tails of K^* are easily satisfied by normal kernels or any other density functions that decay at a suitably high geometric rate. Assumption K(3) is simple to verify when K^* is a probability density and will follow if

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

the sth-order partial moment matrices of K^* are nonsingular. This will follow for the probability densities typically used in kernel estimation. The conditions in Assumption K(4) are standard for semiparametric estimation. The $\sqrt{N}\gamma^{s-1}$ term in Assumption K(4)(a) is the order of bias for the score estimate. The term $(N\gamma^3)^{-1}$ is the order of the variance of density derivative estimators. As is shown in Remark A3 (Appendix A), $\gamma^{\epsilon'/5}$ represents a lower bound for the density estimator and hence the presence of the $\gamma^{-\epsilon'}$ term in Assumptions K(4)(a) and (b). The term ϵ' can be made arbitrarily small. Choosing $\gamma = CN^{-1/q'}$, the conditions will be satisfied if q' is slightly greater than 3 and s, the order of the kernel, is greater than $3/2 + 1$, so that a third- or higher order kernel should be used.

THEOREM 1. *Let Assumptions P and K hold. Then,* $\sqrt{N}(\hat{\beta} - \beta_0) \to N(0, \Sigma(\beta_0)^{-1})$ *and* $-\nabla^2 \log \hat{L}_N(\hat{\beta}) \to \Sigma(\beta_0)$ *in probability.*

The proof of this theorem is provided in Appendix A.

From Lemma A3 in Appendix A (minus one times) the derivative of the score function will be consistent for Σ. An inspection of the proof of Lemma A3 will show that various other approximations can also be used such as

$$\hat{\Sigma}(\hat{\beta}) = \frac{1}{N}\sum_{i=1}^{N} \chi_1(\hat{f}_i(\hat{\beta}))\hat{S}_i(\hat{\beta})\hat{S}_i(\hat{\beta})^T. \tag{3.2}$$

Under our assumptions this will be consistent for $\Sigma(\beta_0)$. The term $\hat{\Sigma}(\hat{\beta})$ can be calculated using the estimated score functions. It follows that likelihood ratio (LR), Lagrange multiplier (LM), and Wald tests and a Hausman–Wu test for correct specification of F_ν can be conducted using $\hat{\beta}$ (and/or restricted versions thereof).

4. EFFICIENCY ISSUES

The results of Section 3 rely implicitly on the usual identification condition that the expected log-likelihood function has a unique maximizer[9] and that $\Sigma(\beta_0)$ be a positive definite matrix. Hahn (1994) derived the efficiency bound for the MPH model without censoring and showed that for the mixed proportional Weibull model, $\lambda = \alpha t^{\alpha-1}\theta(x)$ with $\theta(x) = e^{-x^T\delta}$, $\Sigma(\beta_0)$ is singular, even though the model is identified.[10] This is unfortunate because it limits our distributional results somewhat; it also raises the question of whether the lack of $\sqrt{N}$-estimability of the Weibull model may be the rule, rather than the exception. It turns out that the issue is not that severe.

First, by fixing one of the elements of δ it is possible to estimate $(\alpha, \delta^T)^T$ at the usual $\sqrt{N}$ rate. This is discussed by Hahn and is done implicitly in the Monte Carlo experiments in Horowitz (1996). Furthermore, for choices of $\theta(x)$ other than $\exp(-x^T\delta)$, $\Sigma(\beta_0)$ often will be nonsingular. Finally, for the MPH family with $\theta(x) = e^{-x^T\delta}$ Ridder and Woutersen (2003) show that $\Sigma(\beta_0)$ is

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

singular only for those models with integrated baseline hazard of the form $\Lambda(t) = h(t)^{\alpha}$ and for which $h(0)$ is equal to zero or infinity. In the Weibull model the singularity of $\Sigma(\beta_0)$ can be avoided by considering $h(t) = \alpha(t + \epsilon)^{(\alpha-1)}$ for $\epsilon > 0$. Other examples of MPH models with a nonsingular $\Sigma(\beta_0)$ include the log-logistic specification (Trussel and Richards, 1985; Gritz, 1993), polynomial specifications (Kennan, 1985; Ham and Rea, 1987; Gurmu, Rilstone, and Stern, 1995), and the piecewise-constant hazard of Meyer (1990) and Han and Hausman (1990). In conclusion, there is a large class of non-Weibull, and nonstandard Weibull models that can be estimated at the $\sqrt{N}$ rate using our method.[11]

The Hahn (1994) efficiency bound results are derived for separable MPH models with time-invariant covariates and coefficients. The use of such nonseparable models in numerous papers such as Meyer (1990), Follmann, Goldberg, and May (1990), Katz and Meyer (1990), Jurajda (2002), and Canals-Cerdá (2003) underscores their importance. As a consequence, we present an intuitive way to derive the bounds of models of the form (2.1) without the separability requirement used in Hahn (1994).

Following Newey (1990), conjecture the true model as belonging to a family of submodels described by a parameter of interest, β, and a nuisance parameter, η, that describes the distribution of F_{ν}. The semiparametric efficiency bound is defined as the supremum of the Cramler–Rao lower bounds for all possible submodels. For a given submodel, the log-likelihood function is written

$$\log l_i(\beta, \eta) = \log \lambda_i(\beta) + \log f(\Lambda_i(\beta); \eta) \tag{4.1}$$

and the score function for this submodel as

$$\begin{pmatrix} S_{\beta i} \\ S_{\eta i} \end{pmatrix} = \begin{pmatrix} \nabla_{\beta} \log l_i(\beta, \eta) \\ \nabla_{\eta} \log l_i(\beta, \eta) \end{pmatrix} = \begin{pmatrix} \dfrac{\nabla_{\beta} \lambda_i(\beta)}{\lambda_i(\beta)} + \dfrac{\nabla_{\beta} f(\Lambda_i(\beta); \eta)}{f(\Lambda_i(\beta); \eta)} \\ \dfrac{\nabla_{\eta} f(\Lambda_i(\beta); \eta)}{f(\Lambda_i(\beta); \eta)} \end{pmatrix}. \tag{4.2}$$

Suppose that $f(\Lambda)$ and $\mathcal{E}[S_{\beta}|\Lambda]$ can be approximated using a basis series $\{\rho_j\}$ such that

$$f(\Lambda) = \lim_{M \to \infty} \sum_{j=1}^{M} \eta_j \rho_j(\Lambda) \quad \text{and} \quad \mathcal{E}[S_{\beta}|\Lambda] = \lim_{M \to \infty} \sum_{j=1}^{M} \gamma_j \frac{\rho_j(\Lambda)}{f(\Lambda)}, \tag{4.3}$$

where $\{\rho_j\}$ is normalized such that

$$\mathcal{E}\left[\frac{\rho_l(\Lambda)}{f(\Lambda)} \frac{\rho_j(\Lambda)}{f(\Lambda)} \right] = \begin{cases} 1, & l = j \\ 0, & l \neq j. \end{cases} \tag{4.4}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

For simplicity, assume that β is a scalar. As a consequence of these definitions the γ_j are such that

$$\gamma_j = \mathcal{E}\left[\mathcal{E}[S_\beta|\Lambda]\frac{\rho_j(\Lambda)}{f(\Lambda)}\right] = \mathcal{E}\left[S_\beta\frac{\rho_j(\Lambda)}{f(\Lambda)}\right] \quad \text{and} \quad \lim_{M\to\infty}\sum_{j=1}^{M}\gamma_j^2 = \mathcal{E}[\mathcal{E}[S_\beta|\Lambda]^2]. \tag{4.5}$$

Consider a sequence of likelihood functions for these models for different M. Putting

$$\delta_M = (\beta, \eta_M^T)^T \quad \text{and} \quad P_M = (\rho_1, \ldots, \rho_M)^T, \tag{4.6}$$

we can write down the corresponding log-likelihood function as

$$\log l(\delta_M) = \log\lambda(\beta) + \log\sum_{j=1}^{M}\eta_j\rho_j(\Lambda(\beta)). \tag{4.7}$$

The score function for this model is then

$$S_{\delta_M} = \begin{pmatrix} S_\beta \\ S_{\eta_M} \end{pmatrix} = \begin{pmatrix} \dfrac{\nabla_\beta\lambda}{\lambda} + \dfrac{\nabla_\beta f}{f} \\ \dfrac{P_M}{f} \end{pmatrix}, \tag{4.8}$$

and the information matrix is

$$\mathcal{I}_M = \mathcal{E}\begin{pmatrix} S_\beta^2 & S_\beta\dfrac{P_M^T}{f} \\ S_\beta\dfrac{P_M}{f} & \dfrac{\rho\rho^T}{f^2} \end{pmatrix} = \begin{pmatrix} \mathcal{E}[S_\beta^2] & \gamma^T \\ \gamma & I_m \end{pmatrix}. \tag{4.9}$$

The Cramler–Rao lower bound for β in such a model is simply

$$\mathcal{V}_M = (\mathcal{E}[S_\beta^2] - \gamma_M^T I_M^{-1}\gamma_M)^{-1} = \left(\mathcal{E}[S_\beta^2] - \sum_{j=1}^{M}\gamma_j^2\right)^{-1}. \tag{4.10}$$

This is increasing in M. In fact, from the properties of the γ's we have

$$\begin{aligned} \mathcal{V} = \lim_{M\to\infty}\mathcal{V}_M &= \left(\mathcal{E}[S_\beta^2] - \sum_{j=1}^{\infty}\gamma_j^2\right)^{-1} \\ &= (\mathcal{E}[S_\beta^2] - \mathcal{E}[\mathcal{E}[S_\beta|\Lambda]^2])^{-1} \\ &= (\mathcal{E}[(S_\beta - \mathcal{E}[S_\beta|\Lambda])^2])^{-1}. \end{aligned} \tag{4.11}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Here $\mathcal{V}$ corresponds to the asymptotic variance of $\hat{\beta}$, the estimator in this paper, and $\hat{\beta}$ is therefore first-order efficient. Furthermore, in the absence of covariates, $S_{\beta i}$ is simply a function of the duration variable and, because Λ_i is a monotonic transformation of the duration variable, we have $\mathcal{E}[S_{\beta i}|\Lambda_i] = \mathcal{E}[S_{\beta i}|t_i] = S_{\beta i}$. In this case the efficiency bound explodes. This is interesting because covariates are often used to demonstrate the identification of models with unobserved heterogeneity. Although the presence of observed covariates is not always a necessary condition for identification, the efficiency bounds indicate a key role for observed covariates to obtain an $\sqrt{N}$-consistent estimator.

5. CONCLUSION

This paper has introduced, to our knowledge, the first $\sqrt{N}$-consistent and efficient semiparametric maximum likelihood estimator for a class of duration models with unobserved heterogeneity. There are several directions for extending this research. First, we wish to allow for various censoring mechanisms. Second, we wish to generalize the results to competing risks and other multivariate models. Third, we wish to refine the results for more extensive specification testing and diagnostic checking.

NOTES

1. One might argue in defense of these approaches that the distributions can always be bootstrapped. However, this raises the question of whether the bootstrap is consistent in those cases.

2. Monte Carlo results demonstrating the viability of the technique can be obtained from the authors.

3. This remains true in the presence of time-dependent covariates, extending the applicability of our results.

4. A number of corrections for this "boundary" problem have been proposed in the statistics literature. Most of these (e.g., Gasser and Muller, 1979) are based on kernels with finite support and do not have the degree of smoothness or level of bias correction we require.

5. In the case of censoring it is necessary to estimate f by some other technique, for example, using some sort of smoothed Kaplan–Meier estimator. Although it is intuitive that the resulting estimator, with the corresponding modification to the likelihood function, will be consistent, we have found that this introduces many technical difficulties and is beyond the scope of the current paper.

6. Clearly, certain models may require additional identification assumptions.

7. The previous version of this paper defined $\hat{\beta}$ as the minimizer of a semiparametric likelihood function, showing consistency and asymptotic normality. In that version we also used an $\sqrt{N}$-consistent estimator to define the trimming functions. As the co-editor and a referee pointed out, given this reliance on a preliminary $\sqrt{N}$-consistent estimator, we might as well just examine a two-step version of the approach.

8. We denote the kth-order derivative of a function of Λ by a superscript (k). For a function $q(\beta)$, $\nabla q(\beta) = \partial q(\beta)/\partial\beta$, and $\nabla^2 q(\beta) = \partial^2 q(\beta)/\partial\beta\partial\beta^T$. For a matrix A, we define $\|A\| = (\mathrm{Vec}[A]'\mathrm{Vec}[A])^{1/2}$.

9. Elbers and Ridder (1982) and Heckman and Singer (1985) show that, apart from normalization restrictions, the presence of observable covariates is generally sufficient to identify mixed duration models.

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

10. See Horowitz (1996, pp. 148–150) for further discussion of this topic. Hahn (1994) also showed that the information matrix of the multispell MPH model is nonsingular even in the Weibull case.

11. An earlier version of this paper included numerical results indicating that the members of the family of Box–Cox hazards (Heckman and Singer, 1985; Lancaster, 1990), including hazard specifications such as the Weibull, log-logistic, and polynomial, have nonsingular $\Sigma(\beta_0)$ except in the Weibull case.

REFERENCES

Baker, M. & A. Melino (2000) Duration dependence and nonparametric heterogeneity: A Monte Carlo study. *Journal of Econometrics* 96, 357–393.

Canals-Cerdá, J. (2003) Pre-Marital Birth, Marriage and Welfare. Working paper, University of Colorado at Boulder.

Elbers, C. & G. Ridder (1982) True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies* 49, 403–409.

Follmann, D., M. Goldberg, & L. May (1990) Personal characteristics, unemployment insurance, and the duration of unemployment. *Journal of Econometrics* 45, 351–366.

Gasser, T. & H.G. Muller (1979) Kernel estimation of regression functions. In T. Gasser & M. Rosenblatt (eds.), *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, pp. 23–68. Springer.

Gritz, R.M. (1993) The impact of training on the frequency and duration of employment. *Journal of Econometrics* 57, 21–52.

Gurmu, S., P. Rilstone, & S. Stern (1995) Nonparametric Hazard Rate Estimation. Mimeo, Department of Economics, University of Virginia.

Hahn, J. (1994) The efficiency bound of the mixed proportional hazard model. *Review of Economic Studies* 61, 607–629.

Ham, J. & S. Rea (1987) Unemployment insurance and male unemployment duration in Canada. *Journal of Labor Economics* 5, 325–353.

Han, A. & J.A. Hausman (1990) Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics* 5, 1–28.

Heckman, J.J. & B. Singer (1984) A method for reducing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.

Heckman, J.J. & B. Singer (1985) Social science duration data. In J.J. Heckman & B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*, pp. 39–110. Cambridge University Press.

Honoré, B.E. (1990) Simple estimation of a duration model with unobserved heterogeneity. *Econometrica* 58, 453–474.

Horowitz, J.L. (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64, 103–138.

Jurajda, S. (2002) Estimating the effect of unemployment insurance compensation on the labor market histories of displaced workers. *Journal of Econometrics* 108, 227–252.

Katz, L. & B.D. Meyer (1990) The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of Public Economics* 41, 45–72.

Kennan, J. (1985) The duration of contract strikes in U.S. manufacturing. *Journal of Econometrics* 28, 5–28.

Kiefer, N. (1988) Economic duration data and hazard functions. *Journal of Economic Literature* 26, 646–679.

Klein, R.W. & R.H. Spady (1993) An efficient semiparametric estimator for binary response models. *Econometrica* 61, 387–421.

Lancaster, T. (1979) Econometric methods for the duration of unemployment. *Econometrica* 47, 939–956.

Lancaster, T. (1990) *The Analysis of Transition Data*. Cambridge University Press.

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Meyer, B.D. (1990) Unemployment insurance and unemployment spells. *Econometrica* 58, 757–782.
Newey, W.K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.
Nielsen, J.P., O. Linton, & P.J. Bickel (1998) On a semiparametic survival model with flexible covariate effect. *Annals of Statistics* 26, 215–241.
Ridder, G. (1990) The non-parametric identification of generalized accelerated failure time models. *Review of Economic Studies* 57, 167–182.
Ridder, G. & T.M. Woutersen (2003) The singularity of the information matrix of the mixed proportional hazard model. *Econometrica* 71, 1579–1589.
Trussel, J. & T. Richards (1985) Correcting for unobserved heterogeneity in hazard models: An application of the Heckman–Singer procedure to demographic data. In N. Tuma (ed.), *Sociological Methodology*, pp. 242–276. Jossey-Bass.
Van der Vaart, A. (1996) Efficient semiparametric estimation in semiparametric mixture models. *Annals of Statistics* 24, 862–879.

APPENDIX A

Throughout the Appendixes, Assumptions P and K are assumed to hold. A subscript q on a vector (matrix) denotes its qth (lqth) (arbitrary) element. The expectations operator, conditional on z_i, is indicated by $\mathcal{E}_i$. We first note several results that will be useful in the derivations to follow.

Remark A1. Λ_i and the covariates are statistically independent because, from equation (2.4), $f(\Lambda_i(\beta)|x_i) = f(\Lambda_i(\beta))$.

Remark A2. By the chain rule of calculus,

$$\frac{d}{dt_i}c(t_i,x_i) = \frac{d}{d\Lambda_i}c(t_i,x_i)\frac{d\Lambda_i}{dt_i} = c(t_i,x_i)^{(1)}\lambda_i, \tag{A.1}$$

so that $c(t_i,x_i)^{(1)} = (1/\lambda_i)(d/dt_i)c(t_i,x_i)$. In particular, $(\nabla\Lambda_i)^{(1)} = (1/\lambda_i)\nabla\lambda_i$ and $(\nabla^2\Lambda_i)^{(1)} = (1/\lambda_i)\nabla^2\lambda_i$.

Remark A3. Put $D = f + 2\gamma^{\epsilon'/5}(1-\chi_2(f))$ and note that, if $f < \gamma^{\epsilon'/5}$, then $1-\chi_2(f) > \frac{1}{2}$ and $D = D1[f \geq \gamma^{\epsilon'/5}] + D1[f < \gamma^{\epsilon'/5}] \geq f1[f \geq \gamma^{\epsilon'/5}] + 2\gamma^{\epsilon'/5}(1-\chi_2(f))1[f < \gamma^{\epsilon'/5}] \geq \gamma^{\epsilon'/5}$.

Remark A4. For any $\Delta < \infty$, $\chi_1(f)(1-\chi_2(f)) = O(N^{-\Delta})$ uniformly in $f \geq 0$. To see this note that, for large enough N, $\gamma^{\epsilon/5}/2 > 3\gamma^{\epsilon'/5}/2$ and

$$\begin{aligned}\chi_1(f)(1-\chi_2(f))N^\Delta &\leq \chi_1(f)1[f < \gamma^{\epsilon/5}/2]N^\Delta + (1-\chi_2(f))1[f \geq \gamma^{\epsilon/5}/2]N^\Delta \\ &\leq \frac{N^\Delta}{1+e^{\gamma^{-\epsilon/20}/2}} + e^{-\gamma^{-\epsilon'/20}/2}N^\Delta \\ &\to 0\end{aligned} \tag{A.2}$$

for any finite Δ.

Remark A5. From the definition of $\tilde{f}$, $\hat{f}$, and the positivity of f, we have

$$|\tilde{f}-f| = 1[\hat{f} \geq 0]|\hat{f}-f| + 1[\hat{f} < 0]|-\hat{f}-f|, \tag{A.3}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

and because

$$1[\hat{f}<0]|-\hat{f}-f| \le 1[\hat{f}<0]||\hat{f}|+|f||$$
$$= 1[\hat{f}<0]|-\hat{f}+f|$$
$$= 1[\hat{f}<0]|f-\hat{f}| \quad \textbf{(A.4)}$$

it follows that $|\tilde{f}-f| \le |\hat{f}-f|$, uniformly on z,β. Similarly $\hat{f}-\tilde{f} = 2\hat{f}1[\hat{f}<0]$, and because $\hat{f}1[\hat{f}<0]<0$,

$$|\hat{f}-\tilde{f}| = 2|\hat{f}|1[\hat{f}<0] \le 2|\hat{f}-f|1[\hat{f}<0] \le 2|\hat{f}-f|. \quad \textbf{(A.5)}$$

Remark A6. For any $\hat{f}$, any $\Delta<\infty$, $\chi_1(\hat{f})1[\hat{f}<0] = O(N^{-\Delta})$. To see this, note that $\chi_1(\hat{f})1[\hat{f}<0] \le \chi_1(0)1[\hat{f}<0] \le \chi_1(0)$ and $\chi_1(0)N^{-\Delta} \to 0$ for any $\Delta<\infty$.

Remark A7. Note that $(1-\chi_1(f))\chi_1(f)$ and $(1-\chi_2(f))\chi_2(f)$ are simply logistic density functions and

$$\frac{d}{df}\chi_1(f) \equiv \chi_1'(f) = -\frac{1}{\gamma^{\epsilon/4}}(1-\chi_1(f))\chi_1(f),$$
$$\frac{d}{df}\chi_2(f) \equiv \chi_2'(f) = -\frac{1}{\gamma^{\epsilon'/4}}(1-\chi_2(f))\chi_2(f). \quad \textbf{(A.6)}$$

Remark A8. Put $\tilde{D} = \tilde{f} + 2\gamma^{\epsilon'/5}(1-\chi_2(\hat{f}))$. From the mean value theorem

$$\tilde{D}-D = \tilde{f}-f+2\gamma^{\epsilon'/5-\epsilon'/4}((1-\chi_2(\bar{f}))\chi_2(\bar{f}))(\hat{f}-f), \quad \textbf{(A.7)}$$

where $\bar{f}$ is a mean value. From Remark A5 and the triangle inequality

$$|\tilde{D}-D| \le (1+2\gamma^{\epsilon'/5-\epsilon'/4})|\hat{f}-f|. \quad \textbf{(A.8)}$$

Remark A9. From the definition of $\hat{f}_i$ and the dominating conditions it follows that

$$\|\nabla^r f_i(\beta)\| \le B_i^r C\frac{1}{N\gamma^{r+1}}\sum_{j\neq i}B_j, \qquad r=0,1,2,3,4. \quad \textbf{(A.9)}$$

Put

$$N_{1i}(\beta) = \nabla f_i(\beta) - \mathcal{E}[\nabla f_i(\beta)|\Lambda_i(\beta)] - \mathcal{E}[\nabla\Lambda_i(\beta)|\Lambda_i(\beta)]f_i(\beta),$$
$$N_{2i}(\beta) = \nabla^2\Lambda_i(\beta)f_i^{(1)}(\beta) - (\mathcal{E}[\nabla^2\Lambda_i(\beta)f_i(\beta)|\Lambda_i(\beta)])^{(1)}$$
$$- \nabla\Lambda_i(\beta)(\mathcal{E}[\nabla\Lambda_i(\beta)^T|\Lambda_i(\beta)]f_i(\beta))^{(2)}$$
$$- (\mathcal{E}[\nabla\Lambda_i(\beta)|\Lambda_i(\beta)]f_i(\beta))^{(2)}\nabla\Lambda_i(\beta)^T$$
$$+ \nabla\Lambda_i(\beta)\nabla\Lambda_i(\beta)^T f_i^{(2)}(\beta) + (\mathcal{E}[\nabla\Lambda_i(\beta)\nabla\Lambda_i(\beta)^T|\Lambda_i(\beta)]f_i(\beta))^{(2)}. \quad \textbf{(A.10)}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Here $N_{1i}(\beta)$ and $N_{2i}(\beta)$ are, to first order, the expected values, conditional on z_i, of $\nabla\hat{f}_i(\beta)$ and $\nabla^2\hat{f}_i(\beta)$.

For the following two lemmas, all functions are evaluated at $\beta = \beta_0$. Put $D_i = f_i + 2\gamma^{\epsilon'/5}(1 - \chi_2(f_i))$, $\dot{\lambda}_i = \nabla \log \lambda_i$, $\tilde{S}_i^* = (\dot{\lambda}_i f_i + N_{1i})/D_i$, $\hat{N}_i = \dot{\lambda}_i \hat{f}_i + \nabla\hat{f}_i$.

LEMMA A1.

$$\frac{1}{N}\sum \chi_{1i}\hat{S}_i = \frac{1}{N}\sum \chi_{1i}\tilde{S}_i^* + o_P(N^{-1/2}).$$

Proof. We can decompose

$$\hat{S}_i = \frac{\mathcal{E}_i\hat{N}_i}{D_i} + \frac{(\hat{N}_i - \mathcal{E}_i\hat{N}_i)(D_i - \tilde{D}_i)}{\tilde{D}_i D_i} + \frac{(D_i - \tilde{D}_i)\mathcal{E}_i\hat{N}_i}{\tilde{D}_i D_i} + \frac{\hat{N}_i - \mathcal{E}_i\hat{N}_i}{D_i} \equiv \sum_{j=1}^{4} D_{ji}. \qquad \textbf{(A.11)}$$

Let

$$D_{lN} = \frac{1}{N}\sum \chi_{1i} D_{li}, \qquad l = 1,2,3,4. \qquad \textbf{(A.12)}$$

From Corollary C4 in Appendix C and the dominating conditions,

$$\sqrt{N}D_{3N} = \frac{1}{\sqrt{N}}\sum \chi_{1i}(D_i - \tilde{D}_i)\frac{f_i\tilde{S}_i}{\tilde{D}_i D_i} + O_p(1). \qquad \textbf{(A.13)}$$

The qth (arbitrary) element of the first term on the right-hand side of (A.13) has quadratic mean

$$\begin{aligned}
&\mathcal{E}\left[\left(\frac{1}{\sqrt{N}}\sum \chi_{1i}(D_i - \tilde{D}_i)\frac{f_i\tilde{S}_{iq}}{\tilde{D}_i D_i}\right)^2\right] \\
&\quad = \mathcal{E}\left[\left(\chi_{11}(D_1 - \tilde{D}_1)\frac{f_1\tilde{S}_{1q}}{\tilde{D}_1 D_1}\right)^2\right] \\
&\qquad + (N-1)\mathcal{E}\left[\left(\chi_{11}(D_1 - \tilde{D}_1)\frac{f_1\tilde{S}_{1q}}{\tilde{D}_1 D_1}\right)\left(\chi_{12}(D_2 - \tilde{D}_2)\frac{f_2\tilde{S}_{2q}}{\tilde{D}_2 D_2}\right)\right] \\
&\quad \equiv A + (N-1)B. \qquad \textbf{(A.14)}
\end{aligned}$$

From Remarks A4 and A8 and because $0 \le \chi_1 \le 1$,

$$\begin{aligned}
A &= O(\gamma^{-4\epsilon'/5})O(\gamma^{2\epsilon'/5}\gamma^{-2\epsilon'/4})\mathcal{E}[|f_1 - \hat{f}_1|^2\tilde{S}_{1q}^2] \\
&= O(\gamma^{-9\epsilon'/10})\mathcal{E}[\mathcal{E}_1|f_1 - \hat{f}_1|^2\tilde{S}_{1q}^2]. \qquad \textbf{(A.15)}
\end{aligned}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Using Corollary C4 and because the variance of $\hat{f}_i$ is bounded by $((N\gamma)^{-1})C$

$$A \leq O(\gamma^{-9\epsilon'/20})O\left(\frac{1}{N\gamma} + \gamma^{2s}\right)\mathcal{E}[\tilde{S}_{1q}^2] = o(1). \tag{A.16}$$

Now, with the exception of $\tilde{S}_1$ and $\tilde{S}_2$, each of the terms in B is a function only of $\Lambda_1, \ldots, \Lambda_N$. Thus

$$\begin{aligned} B &= \mathcal{E}\left[\left(\chi_{11}(D_1 - \tilde{D}_1)\frac{f_1\mathcal{E}[\tilde{S}_{1q}|\Lambda_1, \ldots, \Lambda_N]}{\tilde{D}_1 D_1}\right)\left(\chi_{12}(D_2 - \tilde{D}_2)\frac{f_2\mathcal{E}[\tilde{S}_{2q}|\Lambda_1, \ldots, \Lambda_N]}{\tilde{D}_2 D_2}\right)\right] \\ &= \mathcal{E}\left[\left(\chi_{11}(D_1 - \tilde{D}_1)\frac{f_1\mathcal{E}[\tilde{S}_{1q}|\Lambda_1]}{\tilde{D}_1 D_1}\right)\left(\chi_{12}(D_2 - \tilde{D}_2)\frac{f_2\mathcal{E}[\tilde{S}_{2q}|\Lambda_2]}{\tilde{D}_2 D_2}\right)\right] \\ &= 0. \end{aligned} \tag{A.17}$$

Therefore, $\sqrt{N}D_{3N} = o_p(1)$. From Corollary C4, $|\mathcal{E}_i\hat{f}_i - f_i| \leq \gamma^s C$ and $|\mathcal{E}_i\nabla\hat{f}_i - N_{1i}| \leq \gamma^{s-1}B_i$. The variances of $\hat{f}_i$ and $\nabla\hat{f}_i$ are of the usual orders $O(1/N\gamma)$ and $O(1/N\gamma^3)$. Consequently, using Remark A8 and the assumptions on the window width, D_{2N} is $o_P(N^{-1/2})$. It is trivial that

$$D_{1N} = \frac{1}{N}\sum \chi_{1i}\tilde{S}_i^* + o_P(N^{-1/2}). \tag{A.18}$$

Applying Lemma B1 in Appendix B to D_{4N} yields the stated result. ■

LEMMA A2.

$$\frac{1}{N}\sum \chi_{1i}\hat{S}_i = \frac{1}{N}\sum \tilde{S}_i + o_p(N^{-1/2}).$$

Proof. From Lemma A1 we have

$$\frac{1}{N}\sum \chi_{1i}\hat{S}_i = \frac{1}{N}\sum \chi_{1i}\tilde{S}_i^* + o_P(N^{-1/2}). \tag{A.19}$$

Note that $\chi_{1i}\tilde{S}_i^* = Q_i\tilde{S}_i$, $Q_i = \chi_{1i}f_i/D_i$, and $0 < Q_i \uparrow 1$ uniformly on f_i. From random sampling and dominated convergence

$$\begin{aligned} \mathcal{E}\left[\left(\frac{1}{\sqrt{N}}\sum Q_i\tilde{S}_{iq} - \frac{1}{\sqrt{N}}\sum \tilde{S}_{iq}\right)^2\right] &= \mathcal{E}\left[\left(\frac{1}{\sqrt{N}}\sum (Q_i - 1)\tilde{S}_{iq}\right)^2\right] \\ &= \mathcal{E}[(Q_i - 1)^2\tilde{S}_{iq}^2] \\ &\to 0. \end{aligned} \tag{A.20}$$

■

LEMMA A3. *Let $\hat{H}_N(\bar{\beta})$ denote the Hessian matrix $\hat{H}_N(\beta) = \nabla S_N(\beta)$, evaluated at (possibly different) points $\bar{\beta}_{lq}$, $l,q = 1, \ldots, k$. Suppose $\|\bar{\beta}_{lq} - \beta_0\| = O_P(N^{-1/2})$ for all l,q. Then, $\|-\hat{H}_N(\bar{\beta}) - \Sigma(\beta_0)\| = o_P(1)$.*

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Proof. We write the l,qth (arbitrary) element of $\hat{H}_N(\bar{\beta})$ as $\hat{h}_N(\bar{\beta}_{lq}) = (1/N)\sum \hat{h}_{iN}(\bar{\beta}_{lq})$. Using a second-order Taylor series we have

$$\hat{h}_N(\bar{\beta}_{lq}) = \hat{h}_N(\beta_0) + \nabla \hat{h}_N(\beta_0)^T(\bar{\beta}_{lq} - \beta_0) + \frac{1}{2}(\bar{\beta}_{lq} - \beta_0)^T \nabla^2 \hat{h}_N(\tilde{\beta}_{lq})(\bar{\beta}_{lq} - \beta_0), \tag{A.21}$$

where $\tilde{\beta}_{lq}$ is between $\bar{\beta}_{lq}$ and β_0. We will first show that the second and third terms in this Taylor series are $o_P(1)$. We will then derive the probability limit of the first term. Noting that the Hessian involves differentiating λ_i, $\hat{f}_i$, and the smoothing functions it is convenient to decompose the summands in $\hat{h}_N$: $\hat{h}_{iN} = \hat{h}_{i1N} + \hat{h}_{i2N} + \hat{h}_{i3N} + \hat{h}_{i4N}$ where

$$\hat{h}_{i1N} = -\frac{1}{\gamma^{\epsilon/4}} \frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i} \nabla_l \hat{f}_i \dot{\lambda}_{iq} \hat{f}_i, \qquad \hat{h}_{i2N} = -\frac{1}{\gamma^{\epsilon/4}} \frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i} \nabla_l \hat{f}_i \nabla_q \hat{f}_i,$$

$$\hat{h}_{i3N} = -\frac{2\gamma^{\epsilon'/5}}{\gamma^{\epsilon'/4}} \hat{\chi}_{1i} \hat{\chi}_{2i}(1 - \hat{\chi}_{2i}) \left(\frac{\nabla_l \hat{f}_i}{\tilde{D}_i} \frac{(\dot{\lambda}_{iq} \hat{f}_i + \nabla_q \hat{f}_i)}{\tilde{D}_i} \right),$$

$$\hat{h}_{i4N} = \hat{\chi}_{1i} \left(\frac{\hat{f}_i \nabla_l \dot{\lambda}_{iq} + \dot{\lambda}_{iq} \nabla_l \hat{f}_i + \nabla_l \nabla_q \hat{f}_i}{\tilde{D}_i} - \frac{\nabla_l \hat{f}_i}{\tilde{D}_i} \frac{(\dot{\lambda}_{iq} \hat{f}_i + \nabla_q \hat{f}_i)}{\tilde{D}_i} \right). \tag{A.22}$$

Note that the first two terms in the Taylor series in equation (A.21) are evaluated at $\beta = \beta_0$. Consider first $\hat{h}_{i1N}$. The rth element of $\nabla \hat{h}_{i1N}$ can be written as

$$-\gamma^{\epsilon/4} \nabla_r \hat{h}_{i1N} = \frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i} (\dot{\lambda}_{iq} \hat{f}_i (\nabla_r \nabla_l \hat{f}_i) + \nabla_l \hat{f}_i \hat{f}_i \nabla_r \dot{\lambda}_{iq} + \nabla_l \hat{f}_i \dot{\lambda}_{iq} \nabla_r \hat{f}_i) + \frac{\nabla_l \hat{f}_i \dot{\lambda}_{iq} \hat{f}_i}{\tilde{D}_i^2} \left(-\frac{\tilde{D}_i \chi_1''(\hat{f}_i)}{\gamma^{\epsilon/4}} - \chi_1'(\hat{f}_i) + \frac{2\chi_2'(\hat{f}_i)\chi_1'(\hat{f}_i)}{\gamma^{\epsilon'/4}} \right) \nabla_r \hat{f}_i \tag{A.23}$$

and

$$\begin{aligned}
\gamma^{\epsilon/4} |\nabla_r \hat{h}_{i1N}| &\leq \frac{C}{\gamma^{\epsilon'/5}} (|\hat{f}_i||\nabla_r \nabla_l \hat{f}_i| + |\nabla_l \hat{f}_i||\hat{f}_i| + |\nabla_l \hat{f}_i||\nabla_r \hat{f}_i|) B_i \\
&\quad + \frac{|\nabla_l \hat{f}_i| B_i |\hat{f}_i|}{\gamma^{2\epsilon'/5}} \frac{C}{\gamma^{\epsilon'/4}} |\nabla_r \hat{f}_i| \\
&\leq \frac{C}{\gamma^{2\epsilon'/5} \gamma^{\epsilon'/4}} (|\hat{f}_i||\nabla_r \nabla_l \hat{f}_i| + |\nabla_l \hat{f}_i||\hat{f}_i| + |\nabla_l \hat{f}_i||\nabla_r \hat{f}_i| + \gamma^{-1} |\nabla_l \hat{f}_i||\nabla_r \hat{f}_i|) B_i \\
&\leq \frac{C}{\gamma \gamma^{2\epsilon'/5} \gamma^{\epsilon'/4}} (|\hat{f}_i||\nabla_r \nabla_l \hat{f}_i| + |\nabla_l \hat{f}_i||\hat{f}_i| + |\nabla_l \hat{f}_i||\nabla_r \hat{f}_i|) B_i.
\end{aligned} \tag{A.24}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

By Cauchy–Schwarz

$$\begin{aligned}
&\mathcal{E}_i|\nabla_r \hat{h}_{i1N}| \\
&\quad \le \frac{CB_i}{\gamma^{1+\epsilon/4+2\epsilon'/5+\epsilon'/4}}((\mathcal{E}_i|\hat{f}_i|^2)^{1/2}(\mathcal{E}_i|\nabla_r\nabla_l\hat{f}_i|^2)^{1/2}) \\
&\qquad + \frac{CB_i}{\gamma^{\epsilon/4}\gamma\gamma^{2\epsilon'/5}\gamma^{\epsilon'/4}}((\mathcal{E}_i|\nabla_l\hat{f}_i|^2)^{1/2}(\mathcal{E}_i|\hat{f}_i|^2)^{1/2} + (\mathcal{E}_i|\nabla_l\hat{f}_i|^2)^{1/2}(\mathcal{E}_i|\nabla_r\hat{f}_i|^2)^{1/2}) \\
&\quad \le \frac{CB_i^3}{\gamma^{1+\epsilon/4+2\epsilon'/5+\epsilon'/4}}. \qquad \textbf{(A.25)}
\end{aligned}$$

Therefore $(1/N)\sum\nabla\hat{h}_{i1N}(\bar{\beta}_{lq}-\beta_0) = O_P((\gamma^{1+\epsilon/4+2\epsilon'/5+\epsilon'/4})^{-1})O_P(N^{-1/2}) = o_P(1)$ and $(1/N)\sum\nabla\hat{h}_{i1N} = O_P((\gamma^{1+\epsilon/4+2\epsilon'/5+\epsilon'/4})^{-1})$. Similarly, using Proposition C7 in Appendix C, we have $(1/N)\sum\nabla\hat{h}_{ijN}(\bar{\beta}_{lq}-\beta_0) = o_P(1), j = 2,3,4$. Regarding the third term in the Taylor series, the term $\nabla^2\hat{h}_N(\tilde{\beta}_{lq})$ is a function of the first four derivatives of $\Lambda_i(\beta)$, $\Lambda_j(\beta)$, and K_{ij}. The former are B_i-dominated; the latter are bounded by assumption. It is tedious but straightforward using Remark A9 to show that

$$\|\nabla^2\hat{h}_N(\tilde{\beta}_{lq})\| \le C\frac{1}{\gamma^{5+2\epsilon/5+2\epsilon/4}}\frac{1}{N}\sum B_i^4\frac{1}{N}\sum B_j^4 = O\left(\frac{1}{\gamma^{5+2\epsilon/5+2\epsilon/4}}\right)O_P(1), \qquad \textbf{(A.26)}$$

so that, because $\|\bar{\beta}_{lq}-\beta_0\| = O_P(N^{-1/2})$, $\hat{h}_N(\bar{\beta}_{lq}) = \hat{h}_N(\beta_0) + o_P(1)$. To derive the probability limit of $\hat{h}_N(\beta_0)$ it is useful to first decompose

$$\begin{aligned}
\hat{h}_{i1N} &= -\frac{1}{\gamma^{\epsilon/4}}\frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i}\dot{\lambda}_{iq}(N_{1li}+\nabla_l\hat{f}_i - N_{1li})(f_i+\hat{f}_i-f_i) \\
&= -\frac{1}{\gamma^{\epsilon/4}}\frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i}\dot{\lambda}_{iq}N_{1li}f_i - \frac{1}{\gamma^{\epsilon/4}}\frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i}\dot{\lambda}_{iq}N_{1li}(\hat{f}_i-f_i) \\
&\quad - \frac{1}{\gamma^{\epsilon/4}}\frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i}\dot{\lambda}_{iq}f_i(\nabla_l\hat{f}_i - N_{1li}) - \frac{1}{\gamma^{\epsilon/4}}\frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i}\dot{\lambda}_{iq}(\nabla_l\hat{f}_i - N_{1li})(\hat{f}_i-f_i). \qquad \textbf{(A.27)}
\end{aligned}$$

By applying Cauchy–Schwarz and Propositions C5 and C6 in Appendix C, it follows that each of the last three terms on the right-hand side has an average that is $o_P(1)$. Further, because

$$\frac{1}{\gamma^{\epsilon/4}}\frac{\chi_1'(\hat{f}_i)}{\tilde{D}_i} = \frac{1}{\gamma^{\epsilon/4}}\left(\chi_1'(f_i) - \frac{1}{\gamma^{\epsilon/4}}\chi_1''(\bar{f}_i)(f_i-\hat{f}_i)\right)\left(\frac{1}{D_i}+\frac{D_i-\tilde{D}_i}{\tilde{D}_iD_i}\right), \qquad \textbf{(A.28)}$$

where $\bar{f}_i$ is a mean value, it follows that $(1/N)\sum\hat{h}_{i1N} = -(1/N)\sum A_i + o_P(1)$ with $A_i \equiv -(1/\gamma^{\epsilon/4})(\chi_1'(f_i)/D_i)\dot{\lambda}_{iq}N_{1li}f_i$. By Remark A9, the dominating conditions, and the fact that $f_i \le D_i$,

$$|A_i| \le \frac{\chi_1'(f_i)}{\gamma^{\epsilon/4}}f_i\left|\frac{N_{1li}}{f_i}\frac{\dot{\lambda}_{iq}f_i}{D_i}\right| \le \frac{\chi_1'(f_i)}{\gamma^{\epsilon/4}}f_iB_i^2. \qquad \textbf{(A.29)}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

By Cauchy–Schwarz and because $0 \leq \chi_1' \leq 1$, $\mathcal{E}[\chi_1'(f_i)f_i B_i^2] \leq (\mathcal{E}[\chi_1'(f_i)f_i^2])^{1/2}$ $(\mathcal{E}B_i^4)^{1/2}$. Note that χ_1' is simply the logistic density function. Let $\mu(f)$ denote the density function of the random variable f. Using the change of variables, $\omega = (f - \gamma^{\epsilon/5})/\gamma^{\epsilon/4}$, we have

$$\mathcal{E}[\chi_1'(f)f^2] = \gamma^{\epsilon/4} \int \chi_1'(\omega)(\gamma^{\epsilon/5} + \gamma^{\epsilon/4}\omega)^2 \mu(\gamma^{\epsilon/5} + \gamma^{\epsilon/4}\omega)\, d\omega \leq \gamma^{2\epsilon/5}\gamma^{\epsilon/4}C. \quad \textbf{(A.30)}$$

We therefore see that the right-hand side of (A.30) is bounded by $\gamma^{\epsilon/5}\gamma^{\epsilon/8}C$. We thus see that the expected value of $|A_i|$ is bounded by $\gamma^{\epsilon/5}\gamma^{\epsilon/8}\gamma^{-\epsilon/4}C = O(\gamma^{3\epsilon/40}) = o(1)$. Therefore, the average of the A_i's converges to zero in probability, and hence $(1/N)\sum \hat{h}_{i1N} \xrightarrow{P} 0$. Similarly, $(1/N)\sum \hat{h}_{i2N} \xrightarrow{P} 0$, $(1/N)\sum \hat{h}_{i3N} \xrightarrow{P} 0$, and

$$\frac{1}{N}\sum \hat{h}_{i4N} = \frac{1}{N}\sum \chi_{1i}\left(\frac{f_i \nabla_l \dot{\lambda}_{iq} + \dot{\lambda}_{iq} N_{1li} + N_{2lqi}}{D_i} - \frac{N_{1li}}{D_i}\frac{(\dot{\lambda}_{iq} f_i + N_{1qi})}{D_i}\right) + o_P(1), \quad \textbf{(A.31)}$$

so that

$$\frac{1}{N}\sum \hat{H}_{iN} = \frac{1}{N}\sum H_{iN} + o_P(1), \quad \textbf{(A.32)}$$

where

$$H_{iN} \equiv \chi_{1i}\left(\frac{f_i}{D_i}\frac{\nabla^2\lambda_i}{\lambda_i} + \frac{N_{2i}}{D_i} - \frac{N_{1i}}{D_i}\frac{N_{1i}^T}{D_i} - \frac{f_i \dot{\lambda}_i \dot{\lambda}_i^T}{D_i} + \frac{\dot{\lambda}_i N_{1i}}{D_i}\left(1 - \frac{f_i}{D_i}\right)\right). \quad \textbf{(A.33)}$$

These terms can be bounded simply so that

$$\begin{aligned}\|H_{iN}\| &\leq \chi_{1i}\left(\frac{\|\nabla^2\lambda_i\|}{\lambda_i} + \frac{\|N_{2i}\|}{D_i} + \frac{\|N_{1i}\|}{D_i}\frac{\|N_{1i}\|}{D_i} + \|\dot{\lambda}_i\|\,\|\dot{\lambda}_i\| + \|\dot{\lambda}_i\|\frac{\|N_{1i}^T\|}{D_i}\right) \\ &\leq \left(\frac{\|\nabla^2\lambda_i\|}{\lambda_i} + \frac{\|N_{2i}\|}{f_i} + \frac{\|N_{1i}\|}{f_i}\frac{\|N_{1i}\|}{f_i} + \|\dot{\lambda}_i\|^2 + \|\dot{\lambda}_i\|\frac{\|N_{1i}\|}{f_i}\right) \\ &\leq B_i^2. \end{aligned} \quad \textbf{(A.34)}$$

Therefore, by the law of large numbers and Remark B1 in Appendix B we have that

$$\begin{aligned}\frac{1}{N}\sum \hat{H}_{i4N} &\to \mathcal{E}\left[\lim_{N\to\infty} H_{i4N}\right] \\ &= \mathcal{E}\left[\frac{\nabla^2\lambda_i}{\lambda_i} + \frac{N_{2i}}{f_i} - \frac{N_{1i}}{f_i}\frac{N_{1i}^T}{f_i} - \frac{\nabla\lambda_i}{\lambda_i}\frac{\nabla\lambda_i^T}{\lambda_i}\right] \\ &= -\Sigma(\beta_0). \end{aligned} \quad \textbf{(A.35)}$$

■

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Proof of Theorem 1. Using standard substitutions, $\hat{\beta} = \tilde{\beta} - \hat{H}_N(\tilde{\beta})^{-1}(\hat{S}_N(\beta_0) + \hat{H}_N(\bar{\beta})(\tilde{\beta} - \beta_0))$ where $\hat{H}_N(\bar{\beta})$ is the Jacobian of $\hat{S}_N(\tilde{\beta})$ with arguments evaluated at (possibly different) points between $\tilde{\beta}$ and β_0. Rearranging this we have

$$\sqrt{N}(\hat{\beta} - \beta_0) = \hat{H}_N(\tilde{\beta})^{-1}\sqrt{N}\hat{S}_N(\beta_0) + (I_k - \hat{H}_N(\tilde{\beta})^{-1}\hat{H}_N(\bar{\beta}))\sqrt{N}(\tilde{\beta} - \beta_0). \quad \textbf{(A.36)}$$

It is sufficient to show that (a) $\sqrt{N}\hat{S}_N(\beta_0) \xrightarrow{d} N(0, \Sigma(\beta_0))$ and (b) $\hat{H}_N(\bar{\beta}) \xrightarrow{P} -\Sigma(\beta_0)$ and $\hat{H}_N(\tilde{\beta}) \xrightarrow{P} -\Sigma(\beta_0)$.

(a) follows from Lemmas A1 and A2 and from Lemma B2 in Appendix B. (b) follows from Lemma A3. ■

APPENDIX B

LEMMA B1. $\sqrt{N}D_{4N} = (1/\sqrt{N})\Sigma(\chi_{1i}/D_i)(\hat{N}_i - \mathcal{E}_i\hat{N}_i) = o_P(1)$.

Proof. It is sufficient to show that $N\mathcal{E}[D_{4Nq}^2] = o_P(1)$. From the i.i.d. sampling assumption

$$N\mathcal{E}[D_{4Nq}^2] = \mathcal{E}\left[\left(\frac{\chi_{11}}{D_1}(\hat{N}_{1q} - \mathcal{E}_1\hat{N}_{1q})\right)^2\right]$$

$$+ (N-1)\mathcal{E}\left[\frac{\chi_{11}}{D_1}(\hat{N}_{1q} - \mathcal{E}_1\hat{N}_{1q})\frac{\chi_{12}}{D_2}(\hat{N}_{2q} - \mathcal{E}_2\hat{N}_{2q})\right]$$

$$\equiv A + (N-1)B. \quad \textbf{(B.1)}$$

Put

$$a_{ji} = \frac{\chi_{1i}}{D_i}(\dot{\lambda}_{iq}(K_{ij} - \mathcal{E}_i K_{ij}) + (\nabla_q K_{ij} - \mathcal{E}_i\nabla_q K_{ij})), \quad \textbf{(B.2)}$$

so that

$$\frac{\chi_{1i}}{D_i}(\hat{N}_{iq} - \mathcal{E}_i\hat{N}_{iq}) = \frac{1}{(N-1)\gamma}\sum_{j\neq i} a_{ji}, \quad \textbf{(B.3)}$$

where $\mathcal{E}[a_{ji}] = \mathcal{E}_i[a_{ji}] = 0, j \neq i$. From the i.i.d. sampling assumption

$$A = \frac{1}{(N-1)^2\gamma^2}\mathcal{E}\left[\left(\sum_{j=2}^{N} a_{j1}\right)^2\right]$$

$$= \frac{1}{(N-1)\gamma^2}(\mathcal{E}[a_{21}^2] + (N-2)\mathcal{E}[\mathcal{E}_1[a_{21}]\mathcal{E}_1[a_{31}]])$$

$$= o(1) + 0. \quad \textbf{(B.4)}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Next,

$$B = \frac{1}{(N-1)^2\gamma^2}\,\mathcal{E}\left[a_{12}a_{21} + \sum_{l=3}^{N}\sum_{m=3}^{N} a_{l1}a_{m2}\right]$$

$$= \frac{1}{(N-1)^2\gamma^2}\,(\mathcal{E}[a_{12}a_{21}] + (N-2)\mathcal{E}[a_{31}a_{32}] + (N-2)(N-3)\mathcal{E}[a_{31}a_{42}]). \quad \textbf{(B.5)}$$

It is straightforward that $[1/(N-1)\gamma^2]\mathcal{E}[a_{21}a_{12}] = o(1)$ and $\mathcal{E}[a_{31}a_{42}] = \mathcal{E}[a_{31}]\mathcal{E}[a_{42}] = 0$.

Put

$$a^*_{ji} = \frac{\chi_{1i}}{D_i}\,(\dot{\lambda}_{iq}(K_{ij} - f_i) + (\nabla_q K_{ij} - N_{1iq})). \quad \textbf{(B.6)}$$

It is useful to decompose

$$a_{31}a_{32} = a^*_{31}a^*_{32} + [a_{31} - a^*_{31}]a^*_{32} + [a^*_{31}][a_{32} - a^*_{32}] + [a_{31} - a^*_{31}][a_{32} - a^*_{32}], \quad \textbf{(B.7)}$$

whereby

$$\begin{aligned}\mathcal{E}[a_{31}a_{32}] &= \mathcal{E}[\mathcal{E}_3[a_{31}]\mathcal{E}_3[a_{32}]]\\ &= \mathcal{E}[\mathcal{E}_3[a^*_{31}]\mathcal{E}_3[a^*_{32}]] + \mathcal{E}[\mathcal{E}_3[a_{31} - a^*_{31}]\mathcal{E}_3[a^*_{32}]]\\ &\quad + \mathcal{E}[\mathcal{E}_3[a^*_{31}]\mathcal{E}_3[a_{32} - a^*_{32}]] + \mathcal{E}[\mathcal{E}_3[a_{31} - a^*_{31}]\mathcal{E}_3[a_{32} - a^*_{32}]]\\ &\equiv \mathcal{E}[\mathcal{E}_3[a^*_{31}]\mathcal{E}_3[a^*_{32}]] + o(\gamma^2). \quad \textbf{(B.8)}\end{aligned}$$

Were a standard, symmetric kernel being used, a^*_{31} would have mean zero, conditional on z_3. With the boundary modifications, slightly more delicate arguments are required to show that this is $o(\gamma)$. We have

$$\begin{aligned}\mathcal{E}_3[a^*_{31}] &= \int \mathcal{E}\left[\frac{\chi_{11}}{D_1}\,(\dot{\lambda}_{1q}(K_{13} - f_1) + (\nabla_q K_{13} - N_{11q}))|\,\Lambda_1\right] f(\Lambda_1)\,d\Lambda_1\\ &= \int \left(\mathcal{E}[\dot{\lambda}_{1q}|\,\Lambda_1]K_{13} + \nabla_q\Lambda_3\,\frac{\partial}{\partial\Lambda_3}\,K_{13} + \mathcal{E}[\nabla_q\Lambda_1|\,\Lambda_1]\,\frac{\partial}{\partial\Lambda_1}\,K_{13}\right) Q(\Lambda_1)\,d\Lambda_1.\end{aligned}$$

$$\textbf{(B.9)}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Using integration by parts and Remark A2

$$\begin{aligned}\mathcal{E}_3\left[\frac{1}{\gamma}a_{31}^*\right] &= \int \frac{1}{\gamma}\left(\mathcal{E}[\dot{\lambda}_{1q}|\Lambda_1]K_{13} - \frac{d}{d\Lambda_1}\mathcal{E}[\nabla_q\Lambda_1|\Lambda_1]K_{13}\right)Q(\Lambda_1)\,d\Lambda_1 \\ &\quad + \nabla_q\Lambda_3\frac{\partial}{\partial\Lambda_3}\int\frac{1}{\gamma}K_{13}Q(\Lambda_1)\,d\Lambda_1 \\ &\quad - \int\mathcal{E}[\nabla_q\Lambda_1|\Lambda_1]\frac{1}{\gamma}K_{13}\left(\frac{\partial}{\partial\Lambda_1}Q(\Lambda_1)\right)d\Lambda_1 \\ &= \nabla_q\Lambda_3\frac{\partial}{\partial\Lambda_3}\int\frac{1}{\gamma}K_{13}Q(\Lambda_1)\,d\Lambda_1 \\ &\quad - \int\mathcal{E}[\nabla_q\Lambda_1|\Lambda_1]\frac{1}{\gamma}K_{13}\left(\frac{\partial}{\partial\Lambda_1}Q(\Lambda_1)\right)d\Lambda_1. \qquad \textbf{(B.10)}\end{aligned}$$

Because $Q \uparrow 1$, note that, for any Λ_3, $\int(1/\gamma)K_{13}Q(\Lambda_1)\,d\Lambda_1 \to 1$ and

$$\int\mathcal{E}[\nabla_q\Lambda_1|\Lambda_1]\frac{1}{\gamma}K_{31}\left(\frac{\partial}{\partial\Lambda_1}Q(\Lambda_1)\right)d\Lambda_1 \to \mathcal{E}[\nabla_q\Lambda_3|\Lambda_3]\frac{\partial}{\partial\Lambda_3}1 = 0. \qquad \textbf{(B.11)}$$

Therefore, $\mathcal{E}[a_{31}a_{32}] = o(\gamma^2)$. Substituting (B.7), (B.8), and (B.16) into (B.6) we see that $B = o(N-1)$. Further substituting (B.5) and (B.17) into (B.1) yields the desired result. ■

The following lemma shows that using a nonparametric estimate of the density of Λ in the likelihood trimming function has no effect on the asymptotic distribution of the score function.

LEMMA B2.

$$\frac{1}{N}\sum\hat{\chi}_{1i}\hat{S}_i = \frac{1}{N}\sum\chi_{1i}\hat{S}_i + o_p(N^{-1/2}).$$

Proof. Second-order Taylor series expansions of $\chi_1(\hat{f}_i)$ around the points f_i, $i = 1,\ldots,N$, yields

$$\frac{1}{N}\sum\chi_1(\hat{f}_i)\hat{S}_i = \frac{1}{N}\sum\chi_{1i}\hat{S}_i + \bar{a} + \bar{b} + \bar{c}, \qquad \textbf{(B.12)}$$

where

$$\begin{aligned}\bar{a} &\equiv \frac{1}{N}\sum a_i, \qquad a_i \equiv \chi_{1i}(1-\chi_{1i})\frac{1}{\gamma^{\epsilon/4}}(f_i - \hat{f}_i)\tilde{S}_i^*, \\ \bar{b} &\equiv \frac{1}{N}\sum b_i, \qquad b_i \equiv \chi_{1i}(1-\chi_{1i})\frac{1}{\gamma^{\epsilon/4}}(f_i - \hat{f}_i)(\hat{S}_i - \tilde{S}_i^*), \\ \bar{c} &\equiv \frac{1}{N}\sum c_i, \qquad c_i \equiv \varphi_i\frac{1}{\gamma^{2\epsilon/4}}(f_i - \hat{f}_i)^2\hat{S}_i, \\ \varphi_i &\equiv (1-\chi_1(\bar{f}_i))\chi_1(\bar{f}_i)(2\chi_1(\bar{f}_i) - 1) \qquad \textbf{(B.13)}\end{aligned}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

and where $\bar{f}_i$, $i = 1, \ldots, N$, are mean values and we note that $|\varphi_i| \leq 1$. The proof consists of showing that $\bar{a} = o_p(\sqrt{N})$, $\bar{b} = o_p(\sqrt{N})$, and $\bar{c} = o_p(\sqrt{N})$.

$$\mathcal{E}[(\sqrt{N}\bar{a}_q)^2] = \frac{1}{\gamma^{\epsilon/2}}\,\mathcal{E}[\varphi_1^2(f_1 - \hat{f}_1)^2\tilde{S}_{1q}^{*2}] + \frac{1}{\gamma^{\epsilon/2}}\,(N-1)\mathcal{E}[\varphi_1\varphi_j(f_1 - \hat{f}_1)(f_2 - \hat{f}_2)\tilde{S}_{1q}^{*}\tilde{S}_{2q}^{*}], \tag{B.14}$$

$$\frac{1}{\gamma^{\epsilon/2}}\,\mathcal{E}[\varphi_1^2(f_1 - \hat{f}_1)^2\tilde{S}_{1q}^{*2}] = \frac{1}{\gamma^{\epsilon/2}}\,\mathcal{E}[\varphi_1^2\mathcal{E}_1(f_1 - \hat{f}_1)^2\tilde{S}_{1q}^{*2}] \leq \frac{1}{\gamma^{\epsilon/2}}\,\mathcal{E}[\tilde{S}_{1q}^{*2}]\left(\gamma^{2s} + \frac{1}{N\gamma}\right)C = o(1). \tag{B.15}$$

Because $\hat{f}_1$ and $\hat{f}_2$ are functions of $\Lambda_i, i = 1, \ldots, N$, by iterated expectations and random sampling,

$$\mathcal{E}[\varphi_1\varphi_2(f_1 - \hat{f}_1)(f_2 - \hat{f}_2)\tilde{S}_{1q}^{*}\tilde{S}_{1q}^{*}] = \mathcal{E}[\varphi_1\varphi_2(f_1 - \hat{f}_1)(f_2 - \hat{f}_2)\mathcal{E}[\tilde{S}_{1q}^{*}\tilde{S}_{2q}^{*}|\Lambda_1, \ldots, \Lambda_N]]. \tag{B.16}$$

Noting that $\mathcal{E}[\tilde{S}_{1q}^{*}\tilde{S}_{2q}^{*}|\Lambda_1, \ldots, \Lambda_N] = \mathcal{E}[\tilde{S}_{1q}^{*}|\Lambda_1]\mathcal{E}[\tilde{S}_{2q}^{*}|\Lambda_2]$ and

$$\mathcal{E}[\tilde{S}_{1q}^{*}|\Lambda_1] = \mathcal{E}\left[\frac{\dot{\lambda}_{1q}f_1 + \nabla_q f_1 - \mathcal{E}[\nabla_q f_1|\Lambda_1] - \mathcal{E}[\dot{\lambda}_{1q}|\Lambda_1]f_1}{D_1}\,\middle|\,\Lambda_1\right] = 0, \tag{B.17}$$

it follows that $\bar{a} = o_p(\sqrt{N})$. The qth (arbitrary) element of b_i is written

$$b_{iq} = \chi_{1i}(1 - \chi_{1i})\,\frac{1}{\gamma^{\epsilon/4}}\,(f_i - \hat{f}_i)\left(\frac{D_i(\nabla_q\hat{f}_i - N_{1iq})}{\tilde{D}_i D_i} + \frac{(D_i - \tilde{D}_i)N_{1iq}}{\tilde{D}_i D_i}\right). \tag{B.18}$$

Now, because $\chi_{1i}(1 - \chi_{1i}) \leq 1$, $D_i \geq \gamma^{\epsilon'/5}$, and $\tilde{D}_i \geq \gamma^{\epsilon'/5}$

$$|b_{iq}| \leq \frac{1}{\gamma^{\epsilon/4}}\,\frac{1}{\gamma^{2\epsilon'/5}}\,|f_i - \hat{f}_i|(D_i|\nabla_q\hat{f}_i - N_{1iq}| + |D_i - \tilde{D}_i||N_{1iq}|) \leq \frac{1}{\gamma^{\epsilon/4}}\,\frac{1}{\gamma^{2\epsilon'/5}}\,|f_i - \hat{f}_i|(|\nabla_q\hat{f}_i - N_{1iq}|C + |D_i - \tilde{D}_i|B_i). \tag{B.19}$$

Let $\bar{f}_i$ denote a mean value.

$$|D_i - \tilde{D}_i| = |f_i - \tilde{f}_i + 2\gamma^{\epsilon'/5}(\chi_2(\hat{f}_i) - \chi_2(f_i))| = |f_i - \tilde{f}_i + 2\gamma^{\epsilon'/5}\gamma^{-\epsilon'/4}((1 - \chi_2(\bar{f}_i))\chi_2(\bar{f}_i))(f_i - \hat{f}_i)| \leq \gamma^{-\epsilon'/4}|f_i - \hat{f}_i|C, \tag{B.20}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

so

$$|b_{iq}| \leq \frac{1}{\gamma^{\epsilon/4}} \frac{1}{\gamma^{2\epsilon'/5}} |f_i - \hat{f}_i||\nabla \hat{f}_{il} - N_{1il}|C + \frac{1}{\gamma^{\epsilon/4}} \frac{1}{\gamma^{2\epsilon'/5}} \gamma^{-\epsilon'/4} |f_i - \hat{f}_i|^2 B_i. \tag{B.21}$$

Applying the Cauchy–Schwarz inequality to the first term

$$\begin{aligned}
\mathcal{E}_i|b_{iq}| &\leq \frac{1}{\gamma^{\epsilon/4}} \frac{1}{\gamma^{2\epsilon'/5}} (\mathcal{E}_i|f_i - \hat{f}_i|^2 \mathcal{E}_i|\nabla_q \hat{f}_i - N_{1iq}|^2)^{1/2} C \\
&\quad + \frac{1}{\gamma^{\epsilon/4}} \frac{1}{\gamma^{2\epsilon'/5}} \gamma^{-\epsilon'/4} \mathcal{E}_i|f_i - \hat{f}_i|^2 B_i \\
&\leq \frac{1}{\gamma^{\epsilon/4}} \frac{1}{\gamma^{2\epsilon'/5}} \left(\left(\gamma^{2s} + \frac{1}{N\gamma} \right) \left(\gamma^{2s} + \frac{1}{N\gamma^2} \right) \right)^{1/2} B_i \\
&\quad + \frac{1}{\gamma^{\epsilon/2}} \frac{1}{\gamma^{2\epsilon'/5}} \left(\gamma^{2s} + \frac{1}{N\gamma} \right) B_i.
\end{aligned} \tag{B.22}$$

We see then that $\mathcal{E}|\sqrt{N}\bar{b}_q| \leq \sqrt{N}\mathcal{E}[(1/N)\sum \mathcal{E}_i|b_{iq}|] = o(1)$. Because $\sqrt{N}\bar{b}_q$ converges in mean absolute value to zero, $\sqrt{N}\bar{b}_q = o_P(1)$. Next,

$$|c_{iq}| \leq \frac{1}{\gamma^{2\epsilon/4}} (f_i - \hat{f}_i)^2 |\hat{S}_{iq}| \leq \frac{1}{\gamma^{\epsilon/2}} (f_i - \hat{f}_i)^2 \frac{1}{\gamma^2 \gamma^{\epsilon'/5}} B_i \tag{B.23}$$

and

$$\begin{aligned}
\mathcal{E}_i|c_{iq}| &\leq \frac{1}{\gamma^{\epsilon/2}} \mathcal{E}_i (f_i - \hat{f}_i)^2 \frac{1}{\gamma^2 \gamma^{\epsilon'/5}} B_i \\
&\leq \frac{1}{\gamma^{\epsilon/2}} \left(\gamma^{2s} + \frac{1}{N\gamma} \right) C \frac{1}{\gamma^2 \gamma^{\epsilon'/5}} B_i,
\end{aligned} \tag{B.24}$$

so that $\mathcal{E}|\sqrt{N}\bar{c}_q| \leq \sqrt{N}\mathcal{E}[(1/N)\sum \mathcal{E}_i|c_{iq}|] = o(1)$ and the result holds. ■

Put

$$H_{i4} = \frac{\nabla^2 \lambda_i}{\lambda_i} + \frac{N_{2i}}{f_i} + \frac{\nabla \lambda_i}{\lambda_i} \frac{N_{1i}^T}{f_i} + \frac{N_{1i}}{f_i} \frac{\nabla \lambda_i^T}{\lambda_i} - \tilde{S}_i \tilde{S}_i^T. \tag{B.25}$$

Remark B1. $\mathcal{E}[H_{i4}] = -\Sigma(\beta_0)$. By differentiation, integration by parts, and Remark A2

$$\begin{aligned}
\mathcal{E}[H_{i4}] &= \int \mathcal{E}[\dot{\lambda}_i|\Lambda_i](\mathcal{E}[\nabla \Lambda_i^T|\Lambda_i]f_i)^{(1)}\, d\Lambda_i + \int (\mathcal{E}[\nabla \Lambda_i|\Lambda_i]f_i)^{(1)} \mathcal{E}[\dot{\lambda}_i^T|\Lambda_i]\, d\Lambda_i \\
&\quad - \int \mathcal{E}[\dot{\lambda}_i \nabla \Lambda_i^T|\Lambda_i] f_i^{(1)}\, d\Lambda_i - \int \mathcal{E}[\nabla \Lambda_i \dot{\lambda}_i^T|\Lambda_i] f_i^{(1)}\, d\Lambda_i \\
&\quad + \int \mathcal{E}[\dot{\lambda}_i N_{1i}^T|\Lambda_i]\, d\Lambda_i + \int \mathcal{E}[N_{1i} \dot{\lambda}_i^T|\Lambda_i]\, d\Lambda_i - \Sigma(\beta_0).
\end{aligned} \tag{B.26}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Noting that we can write $N_{1i} = [\nabla\Lambda_i|\Lambda_i]f_i^{(1)} - (\mathcal{E}[\nabla\Lambda_i^T|\Lambda_i]f_i)^{(1)}$ allows us to confirm the result.

APPENDIX C

For a function $c(z_i,\beta)$, put $\bar{c}(\Lambda_i(\beta)) = \mathcal{E}[c(z_i,\beta)|\Lambda_i(\beta)]$, $cf(\Lambda_i(\beta)) = c(\Lambda_i(\beta))f(\Lambda_i(\beta))$ and $\bar{cf}^{(l)}(\Lambda) = d^l(\bar{cf}(\Lambda))/d\Lambda^l$. In this Appendix all functions are evaluated at $\beta = \beta_0$.

Lemma C1. *Suppose that $\bar{cf}(\Lambda_i)$ is differentiable to order s with respect to Λ_i on $\Re$ and these derivatives are bounded. Then for some finite C $|(1/\gamma)\mathcal{E}_i[K_{ij}(\beta)c(z_j)] - \bar{cf}(\Lambda_i)| \leq \gamma^s C$.*

Proof. The density function of Λ_j is $f(\Lambda_j)$, and using an sth-order Taylor series we can write

$$\begin{aligned}
\frac{1}{\gamma}\mathcal{E}_i[K_{ij}c(z_j)] &= \mu(\Lambda_i/\gamma)^T\frac{1}{\gamma}\int_0^\infty P\left(\frac{\Lambda_i-\Lambda_j}{\gamma}\right)K^*\left(\frac{\Lambda_i-\Lambda_j}{\gamma}\right)\bar{cf}(\Lambda_j)\,d\Lambda_j \\
&= \mu(\Lambda_i/\gamma)^T\int_{-\infty}^{\Lambda_i/\gamma} P(w)K^*(w)\bar{cf}(\Lambda_i - w\gamma)\,dw \\
&= \mu(\Lambda_i/\gamma)^T\left(\int_{-\infty}^{\Lambda_i/\gamma} K^*(w)P(w)P(w)^T\,dw\right)\bar{cf}^{(0,s-1)}(\Lambda_i) \\
&\quad + \mu(\Lambda_i/\gamma)^T\int_{-\infty}^{\Lambda_i/\gamma} K^*(w)P(w)\bar{cf}^{(s)}(\bar{\Lambda}_i)\frac{(-w\gamma)^s}{s!}\,dw \\
&= \bar{cf}(\Lambda_i) + \gamma^s\mu(\Lambda_i/\gamma)^T\int_{-\infty}^{\Lambda_i/\gamma} P(w)K^*(w)\bar{cf}^{(s)}(\bar{\Lambda}_i)\frac{(-w)^s}{s!}\,dw, \qquad \textbf{(C.1)}
\end{aligned}$$

where

$$\bar{cf}^{(0,s-1)}(\Lambda_i)^T = \left(\bar{cf}(\Lambda_i) \quad \bar{cf}^{(1)}(\Lambda_i)\frac{(-1)}{1!} \quad \cdots \quad \bar{cf}^{(s-1)}(\Lambda_i)\frac{(-1)^{s-1}}{(s-1)!}\right) \qquad \textbf{(C.2)}$$

and $\bar{\Lambda}_i = \Lambda_i - \kappa w\gamma$ $\kappa \in (0,1)$ is a mean value. The result follows from the moment conditions on K^* and the boundedness of $\bar{cf}^{(s)}$. ■

Lemma C2. *Suppose that $\bar{cf}(\Lambda_i)$ is differentiable with respect to Λ_i on $\Re$ to order $s+1$, these derivatives are bounded, and $\bar{cf}(0) = \bar{cf}(\infty) = 0$. Then, for some finite $C \geq 0$: (a) $|\mathcal{E}_i[(1/\gamma)(\partial/\partial\Lambda_i)K_{ij}c(z_j)] - (\bar{cf}(\Lambda_i))^{(1)}| \leq \gamma^{s-1}C$ and (b) $|\mathcal{E}_i[(1/\gamma)((\partial/\partial\Lambda_j)K_{ij})c(z_j)] - (-\bar{cf}(\Lambda_i))^{(1)}| \leq \gamma^s C$.*

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Proof. To prove (a), using the proof of Lemma C1 we have

$$\begin{aligned}\frac{1}{\gamma}\,\mathcal{E}_i\left[\frac{\partial}{\partial\Lambda_i}K_{ij}\,c(z_j)\right] &= \frac{\partial}{\partial\Lambda_i}\,\frac{1}{\gamma}\,\mathcal{E}_i[K_{ij}\,c(z_j)]\\ &= (\bar{c}f(\Lambda_i))^{(1)} + \gamma^{s-1}\mu^{(1)}(\Lambda_i/\gamma)^T\\ &\quad\times\int_{-\infty}^{\Lambda_i/\gamma}P(w)K^*(w)\bar{c}f^{(s)}(\bar{\Lambda}_i)\,\frac{(-w)^s}{s!}\,dw\\ &\quad+\gamma^s\mu(\Lambda_i/\gamma)^T\int_{-\infty}^{\Lambda_i/\gamma}P(w)K^*(w)\bar{c}f^{(s+1)}(\bar{\Lambda}_i)\,\frac{(-w)^s}{s!}\,dw\\ &\quad+\gamma^{s-1}\mu(\Lambda_i/\gamma)^TP(\Lambda_i/\gamma)K^*(\Lambda_i/\gamma)\bar{c}f^{(s)}((1-\kappa)\Lambda_i)\,\frac{(-\Lambda_i/\gamma)^s}{s!}\\ &= (\bar{c}f(\Lambda_i))^{(1)} + \gamma^{s-1}O(1). \end{aligned}\tag{C.3}$$

To prove (b), note that $\bar{c}f(0) = \bar{c}f(\infty) = 0$, so that using integration by parts we have $\mathcal{E}_i[((\partial/\partial\Lambda_j)K_{ij})\bar{c}(\Lambda_j)] = -\int_0^\infty K_{ij}\,\bar{c}f^{(1)}(\Lambda_j)\,d\Lambda_j$. The rest of the steps are as in the proof of Lemma C1. ■

Lemma C3. *Suppose that $\bar{c}f(\Lambda_i)$ is differentiable with respect to Λ_i on $\Re$ to order $s+2$, these derivatives are bounded, $\bar{c}f(0) = \bar{c}f(\infty) = 0$, and $\bar{c}f^{(1)}(0) = \bar{c}f^{(1)}(\infty) = 0$. Then, for some finite $C \geq 0$: (a) $|\mathcal{E}_i[(1/\gamma)(\partial^2/\partial\Lambda_i^2)K_{ij}\,c(z_j)] - (\bar{c}f(\Lambda_i))^{(2)}| \leq \gamma^{s-2}C$, (b) $|\mathcal{E}_i[(1/\gamma)((\partial^2/\partial\Lambda_j^2)K_{ij})c(z_j)] - (\bar{c}f(\Lambda_i))^{(2)}| \leq \gamma^sC$, and (c) $|\mathcal{E}_i[(1/\gamma)((\partial^2/\partial\Lambda_j\,d\Lambda_i)K_{ij})c(z_j)] - (-\bar{c}f(\Lambda_i))^{(2)}| \leq \gamma^{s-1}C$.*

Proof. Follow the same steps as used in the proof of Lemma C2 except that, in proving (b), integration by parts is used twice. ■

Corollary C4. *For some finite $C \geq 0$: (a) $|\mathcal{E}_i[\hat{f}_i] - f_i| \leq \gamma^sC$, (b) $\|\mathcal{E}_i[\nabla\hat{f}_i] - N_{1i}\| \leq \gamma^{s-1}B_i$, and (c) $\|\mathcal{E}_i[\nabla^2\hat{f}_i] - N_{2i}\| \leq \gamma^{s-2}B_i^2$.*

Proof. Part (a) follows directly from Lemma C1. To prove (b) note that

$$\nabla\hat{f}_i = \frac{1}{(N-1)}\sum_{j\neq i}\frac{1}{\gamma}\left(\frac{\partial K_{ij}}{\partial\Lambda_i}\nabla\Lambda_i + \frac{\partial K_{ij}}{\partial\Lambda_j}\nabla\Lambda_j\right)\tag{C.4}$$

and apply Lemma C2 and the dominating conditions on $\nabla\Lambda_i$. To prove (c) note that

$$\begin{aligned}\nabla^2\hat{f}_i &= \frac{1}{(N-1)}\sum_{j\neq i}\frac{1}{\gamma}\left(\frac{\partial K_{ij}}{\partial\Lambda_i}\nabla^2\Lambda_i + \frac{\partial K_{ij}}{\partial\Lambda_j}\nabla^2\Lambda_j\right)\\ &\quad+\frac{1}{(N-1)}\sum_{j\neq i}\frac{1}{\gamma}\left(\frac{\partial^2K_{ij}}{\partial\Lambda_i^2}\nabla\Lambda_i\nabla\Lambda_i^T + \frac{\partial^2K_{ij}}{\partial\Lambda_j\partial\Lambda_i}\nabla\Lambda_j\nabla\Lambda_i^T\right)\\ &\quad+\frac{1}{(N-1)}\sum_{j\neq i}\frac{1}{\gamma}\left(\frac{\partial^2K_{ij}}{\partial\Lambda_i\partial\Lambda_j}\nabla\Lambda_i\nabla\Lambda_j^T + \frac{\partial^2K_{ij}}{\partial\Lambda_j^2}\nabla\Lambda_j\nabla\Lambda_j^T\right)\end{aligned}\tag{C.5}$$

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

and apply Lemma C3 and the dominating conditions on $\nabla\Lambda_i$ and $\nabla^2\Lambda_i$. ■

Proposition C5. $\mathcal{E}[(\hat{f}_i - \mathcal{E}_i[\hat{f}_i])^2] = O((N\gamma)^{-1})$.

Proof. Using iterated expectations and the random sampling assumption

$$\mathcal{E}\mathcal{E}_i[(\hat{f}_i - \mathcal{E}_i\hat{f}_i)^2] = \frac{1}{(N-1)\gamma^2}\,\mathcal{E}\mathcal{E}_1[(K_{12} - \mathcal{E}_1 K_{12})^2]$$

$$= \frac{1}{(N-1)\gamma^2}\,\mathcal{E}\mathcal{E}_1[K_{12}^2] + O(N^{-1}); \tag{C.6}$$

$$\gamma^{-1}\mathcal{E}_1[K_{12}^2] \le \|\mu(\Lambda_1/\gamma)\|^2 \int \|PK^*(\omega)\|^2 f(\Lambda_1 + \gamma\omega)\,d\omega \le C. \tag{C.7}$$

■

Proposition C6. $\mathcal{E}\|\nabla\hat{f}_i - \mathcal{E}_i[\nabla\hat{f}_i]\|^2 = O((N\gamma^3)^{-1})$.

Proof. Without loss of generality we consider the qth (arbitrary) element of $\nabla\hat{f}_i$. As in Proposition C5

$$\mathcal{E}[(\nabla_q\hat{f}_i - \mathcal{E}_i[\nabla_q\hat{f}_i])^2] = \frac{1}{(N-1)\gamma^2}\,\mathcal{E}\mathcal{E}_1[(\nabla_q K_{12})^2] + O(N^{-1}). \tag{C.8}$$

The first term on the right-hand side can be bounded as follows:

$$\mathcal{E}_1[(\nabla_q K_{12})^2] = \mathcal{E}_1[(\nabla_q(\mu_1^T PK_{12}^*))^2]$$

$$\le \|\nabla_q\mu_1\|^2\mathcal{E}_1[\|PK_{12}^*\|^2] + \mathcal{E}_1[(\mu_1^T\nabla_q PK_{12}^{*(1)})^2]$$

$$\le \left(\frac{\nabla_q\Lambda_1}{\Lambda_1}\right)^2 \left\|\frac{\Lambda_1}{\gamma}\,\mu_1^{(1)}\right\|^2 C + \frac{1}{\gamma^2}\,\|\mu_1\|^2(\nabla_q\Lambda_1)^2\mathcal{E}_1[\|PK_{12}^{*(1)}\|^2]$$

$$\le B_1^2 C + \frac{1}{\gamma}\,B_1^2 C, \tag{C.9}$$

where the last inequality follows from the dominating conditions and a standard change of variables. Inserting this last expression into the previous equation gives the desired result. ■

Proposition C7. *Let $\nabla_r\nabla_l\nabla_q\hat{f}_i$ denote an arbitrary third-order derivative of $\hat{f}_i$. Then $|\mathcal{E}_i\nabla^3\hat{f}_i| \le B_i C$.*

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press

Proof. By random sampling $\mathcal{E}_i[\nabla_r\nabla_l\nabla_q\hat{f}_i] = (1/\gamma)\mathcal{E}_i[\nabla_r\nabla_l\nabla_q\mu_i^T PK^*(\omega)]$, say, where $\mu_i = \mu(\Lambda_i/\gamma)$ and $\omega = (\Lambda_i - \Lambda_j)/\gamma$.

$$
\begin{aligned}
&\nabla_r\nabla_l\nabla_q\mu_i^T PK^*(\omega)\\
&\quad= \left(\frac{\nabla_r\nabla_l\nabla_q\Lambda_i}{\Lambda_i}\frac{\Lambda_i}{\gamma}\mu_i^{(1)T} + \frac{\nabla_r\Lambda_i}{\Lambda_i}\frac{\nabla_l\Lambda_i}{\Lambda_i}\frac{\nabla_q\Lambda_i}{\Lambda_i}\left(\frac{\Lambda_i}{\gamma}\right)^3\mu_i^{(3)T}\right)PK^*(\omega)\\
&\qquad+ \left(\frac{\nabla_r\nabla_l\Lambda_i}{\Lambda_i}\frac{\nabla_q\Lambda_i}{\Lambda_i} + \frac{\nabla_l\Lambda_i}{\Lambda_i}\frac{\nabla_r\nabla_q\Lambda_i}{\Lambda_i} + \frac{\nabla_r\Lambda_i}{\Lambda_i}\frac{\nabla_l\nabla_q\Lambda_i}{\Lambda_i}\right)\\
&\qquad\times\left(\frac{\Lambda_i}{\gamma}\right)^2\mu_i^{(2)T}PK^*(\omega)\\
&\qquad+ \mu_i^T PK^{*(2)}(\omega)(\nabla_r\nabla_l\omega\nabla_q\omega + \nabla_l\omega\nabla_r\nabla_q\omega + \nabla_r\omega\nabla_l\nabla_q\omega)\\
&\qquad+ \mu_i^T(PK^{*(3)}(\omega)\nabla_r\omega\nabla_l\omega\nabla_q\omega + PK^{*(1)}(\omega)\nabla_r\nabla_l\nabla_q\omega)\\
&\quad\equiv \rho_i PK^*(\omega) + \mu_i^T PK^{*(1)}(\omega)\gamma^{-1}A_1(Z_i,Z_j)\\
&\qquad+ \mu_i^T PK^{*(2)}(\omega)\gamma^{-2}A_2(Z_i,Z_j) + \mu_i^T PK^{*(3)}(\omega)\gamma^{-3}A_3(Z_i,Z_j),
\end{aligned}
\tag{C.10}
$$

$$
\begin{aligned}
\mathcal{E}_i[\nabla^3\mu_i^T PK^*(\omega)] &= \rho_i\mathcal{E}_i[PK^*(\omega)] + \mu_i^T\mathcal{E}_i[PK^{*(1)}(\omega)\gamma^{-1}A_1(Z_i,Z_j)]\\
&\quad+ \mu_i^T\mathcal{E}_i[PK^{*(2)}(\omega)\gamma^{-2}A_2(Z_i,Z_j)]\\
&\quad+ \mu_i^T\mathcal{E}_i[PK^{*(3)}(\omega)\gamma^{-3}A_3(Z_i,Z_j)].
\end{aligned}
\tag{C.11}
$$

By inspection, $|\rho_i\mathcal{E}_i[PK^*(\omega)]| \le B_i^3$. Using integration by parts as in the proof of Lemma C2, $\mu_i^T\mathcal{E}_i[PK^{*(r)}(\omega)\gamma^{-r}A_r(Z_i,Z_j)] = (-1)^r\mu_i^T\int PK^*(\omega)(A_r(Z_i,\omega - \gamma\Lambda_i)f(\omega - \gamma\Lambda_i))^{(r)}\,d\omega$, $r = 1,2,3$. The A_r functions are simply products of the derivatives of Λ_i and conditional expectations of these derivatives. The former factor out of the preceding integral. The latter, multiplied by $f(\Lambda_i)$ and its derivatives, are bounded, and $|\mu_i^T\mathcal{E}_i[PK^{*(j)}(\omega)\gamma^{-j}A_j(Z_i,Z_j)]| \le B_i^3 C$. ■

https://doi.org/10.1017/S0266466607070120 Published online by Cambridge University Press