

Consistency and Asymptotic Normality of Sieve Estimators Under Weak and Verifiable Conditions

Herman J. Bierens*

Department of Economics and CAPCP†
Pennsylvania State University
University Park, PA 16802

December 11, 2011

Abstract

This paper considers sieve estimation of semi-nonparametric (SNP) models with an unknown density function as non-Euclidean parameter, next to a finite-dimensional parameter vector. The density function involved is modeled via an infinite series expansion, so that the actual parameter space is infinite-dimensional. It will be shown that under weak and verifiable conditions the sieve estimators of these parameters are consistent, and the estimators of the Euclidean parameters are \sqrt{N} asymptotically normal, given a random sample of size N . The latter result is derived in a different way than in the sieve estimation literature. It appears that this asymptotic normality result is in essence the same as for the finite dimensional case. This approach is motivated and illustrated by an SNP discrete choice model.

*The helpful comments of Xiaohong Chen are gratefully acknowledged. This paper has been presented at Penn State University, the University of Rochester, Ohio State University, and LAMES 2011, Chile.

†Support for research within the Center for the Study of Auctions, Procurements, and Competition Policy (CAPCP) at Penn State has been provided by a gift from the Human Capital Foundation (<http://www.hcfoundation.ru/en/>).

JEL codes: C13, C14, C21, C25

Keywords: Sieve estimation, Semi-nonparametric models, Consistency, Asymptotic normality, Discrete choice models

1. Introduction

Semi-nonparametric (SNP) models (also called semi-parametric models) are models that are only partially parametrized, and the non-specified part is an unknown function. See Chen (2007) for a recent review of these models and the various ways to estimate them. If this unknown function is approximated by a series expansion, the standard method to estimate these models is the method of sieves proposed by Grenander (1981). There is also a substantial literature on estimation of semi-parametric models using nonparametric kernel density and/or regression estimators (see for example Horowitz 1998), but these approaches are beyond our scope.

Starting with Geman and Hwang (1982), the consistency of sieve estimators is well-established in the sieve estimation literature, albeit under rather restrictive conditions. In this literature the asymptotic normality of a finite subset of parameter estimates is proved by using a functional derivative, such as the Frechet, Hadamard or Gateaux derivatives, together with a related δ -method and Lipschitz and equicontinuity conditions. See for example Andrews (1994), Bickel et al. (1998), Gill (1989), Newey (1997) and Shen (1997), among others. However, the asymptotic normality conditions involved are very high-level and therefore difficult to verify, in particular the conditions A through D in Shen (1997). The latter paper has set the standard for the asymptotic normality of sieve estimators of smooth functionals. See also Chen and Shen (1998) and Chen (2011) for the time series case and Chen et al. (2003) for the non-smooth case.

Therefore, the purpose of this paper is to establish weak, verifiable and/or implementable conditions for the consistency of sieve estimators in general, and the \sqrt{N} asymptotic normality of the sieve estimator of the Euclidean parameter vector in particular, given a random sample of size N . The i.i.d. assumption is merely made to keep the paper focused on its essentials. It can easily be relaxed to the time series case, for example by assuming stationarity and ergodicity, and using a martingale difference central limit theorem. See McLeish (1974) for the latter.

In general, SNP models involve a (true) Euclidean parameter vector θ_0 , say, of structural parameters (or parameters of interest) and infinitely many (true)

nuisance parameters $\delta^0 = \{\delta_{0,i}\}_{i=1}^\infty$, say, combined as $\xi^0 = (\theta_0, \delta^0) = \{\xi_{0,k}\}_{k=1}^\infty \in \Xi$, where Ξ is the infinite-dimensional parameter space involved. As noted by Shen (1997) and others, the usual assumption for finite dimensional parametric models that the parameter space is a compact metric space containing the true parameter vector in its interior may not hold for Ξ in some cases. The general conditions in Shen (1997) assume these problems away. In contrast, the approach in the current paper will be confined to the case $\Xi = \mathbf{X}_{k=1}^\infty[-\bar{\xi}_k, \bar{\xi}_k]$, where $\bar{\xi}_k$ is an a priori chosen positive sequence converging to zero for $k \rightarrow \infty$, endowed with an appropriate metric to make Ξ compact. Moreover, to prove the asymptotic normality of the sieve estimator of θ_0 it will be assumed that $\bar{\xi}_k$ is chosen such that $\xi^0 = (\theta_0, \delta^0) \in \mathbf{X}_{k=1}^\infty(-\bar{\xi}_k, \bar{\xi}_k)$. It appears that the sieve order for which this asymptotic normality result holds depends on the rate at which $\bar{\xi}_k \rightarrow 0$. Admittedly, this is more restrictive than the general conditions considered in the SNP literature, but given the appropriate choice of the sequence $\bar{\xi}_k$ the other conditions for consistency of the sieve estimator of ξ^0 and the asymptotic normality of the sieve estimator of θ_0 are weak and/or verifiable.

As to consistency, it will be shown that there is no need to assume that the expectation of the log-likelihood function is finite for all admissible parameter values. Next to weak standard regularity conditions it suffices that this expectation is finite in the true (infinite-dimensional) parameter ξ^0 and a single other one. Moreover, it will be shown that asymptotic normality of the sieve estimators of the Euclidean parameter vector θ_0 can be established much easier, and under verifiable primitive conditions, than by the functional derivative approach. The basic idea is actually quite simple. Starting from the standard system of mean value equations for the first-order conditions, this system is converted to a single mean value equation in random function form in order to handle the increasing dimension, by taking a weighted sum of the standard mean value equations with weight functions an orthogonal sequence of cosine functions on the unit interval. Then under mild regularity conditions, one side of this function equation converges weakly to a Gaussian process, whereas the other side contains the parameter estimates in deviation of their true values, times \sqrt{N} , weighted by random functions. The nuisance parameters involved can be eliminated from this equation by using the residuals of the projections of the weight functions corresponding to the parameters of interest on the space spanned by the weight functions corresponding to the nuisance parameters. It appears that the resulting asymptotic variance matrix is the limit of the corresponding variance matrix in the finite dimensional case. The latter confirms similar conclusions by Newey (1994), Ai and Chen (2007),

Ackerberg et al. (2010) and Ichimura and Lee (2010).

My focus will be on SNP models where the non-Euclidean parameter is a density or distribution function which is modeled via an infinite series expansion, similar to the approach of Gallant and Nychka (1987). The latter authors consider an SNP version of Heckman's (1979) sample selection model, where the bivariate error distribution of the latent variable equations is modeled via a Hermite expansion of the error density. Another example of an SNP model is the mixed proportional hazard (MPH) model proposed by Lancaster (1979). In their seminal paper, Elbers and Ridder (1982) have shown that under some mild conditions and normalizations the MPH model is nonparametrically identified. Heckman and Singer (1984) propose to estimate the distribution function of the unobserved heterogeneity variable by a discrete distribution. Bierens (2008) and Bierens and Carvalho (2007) use orthonormal Legendre polynomials to model semi-nonparametrically the unobserved heterogeneity distribution of interval-censored mixed proportional hazard models and bivariate mixed proportional hazard models, respectively.

Of course, there are many more examples of SNP models. However, I will use an SNP discrete choice index model as benchmark model to motivate and illustrate the approach in this paper.

Any density function can be converted one-to-one via an a priori chosen mapping to a density function on the unit interval. See for example Bierens (2008) and the next section. Therefore, without loss of generality we may assume that the non-Euclidean parameter in the SNP models under review takes the form of a density function $h(u)$ on the unit interval. Because $\sqrt{h(u)}$ is an element of the Hilbert space $L^2(0, 1)$ of square-integrable functions, $\sqrt{h(u)}$ can be represented by a countable infinite linear combination of a complete orthogonal sequence in $L^2(0, 1)$, similar to the approach of Gallant and Nychka (1987) for densities on \mathbb{R} with Hermite polynomials as orthonormal sequence. The Legendre polynomials used in Bierens (2008) form such a complete orthogonal sequence in $L^2(0, 1)$. However, as is well-known, the same applies to the well-known Fourier series and the related cosine series. The cosine series has the advantage that it is easy to impose smoothness conditions on $h(u)$. Another advantage is that distribution functions on the unit interval have a closed-form expression in terms of the sine series. Therefore, the approach in this paper will be based on series expansions in terms of the cosine series, although the results carry over straightforwardly to the Fourier series.

Gallant (1981) was the first econometrician to proposed Fourier series expansions as a way to model unknown functions. Gallant's approach is actually

nonparametric in that no Euclidean parameters are involved. See also Eastwood and Gallant (1991) and the references therein. However, the use of Fourier series expansions to model unknown functions has been proposed earlier in the statistics literature, for example in Kronmal and Tarter (1968).

The outline of the paper is as follows. In section 2 the SNP discrete choice model will be introduced and conditions for its identification will be established. Next to standard conditions on the covariates, identification requires a normalization of the unknown distribution function involved to fix its location and scale. This will be done by imposing two quantile restrictions. The version of this model that will be used to motivate and illustrate the sieve estimation approach is the SNP Logit model, which is an SNP generalization of the standard Logit model. In section 3 it will be shown how densities on the unit interval can be represented by a series expansion in terms of the cosine series, and how smoothness and compactness conditions can be imposed. In section 4 the SNP Logit model will be reformulated as an infinite parameter model, in two forms: a penalized least squares (PLS) form and a penalized maximum likelihood (PML) form, where the role of penalty function is to enforce the aforementioned quantile restrictions. Conditions will be set forth for the strong consistency of the sieve estimators in the PLS case, and weak consistency in the PML case. Section 5 provides weak conditions for the consistency of sieve estimators of general SNP models. In particular, it will be shown that the usual condition that the expectation of the objective function is finite can be relaxed. Section 6 deals with the asymptotic normality of the sieve estimators of the Euclidean parameters, using a different approach than in the sieve estimation literature. In section 7 it will be shown that the general asymptotic normality conditions in section 6 apply to the SNP Logit model. The concluding section 8 summarizes the main contribution of this paper to the sieve estimation literature, and indicates further applications.

The proofs are given in either Appendix A (section 9), or in a separate appendix, Bierens (2011). The latter contains proofs that are not too difficult, or are variations of published results, or are too tedious, together with a brief review of some well-known Hilbert space results used in this paper. The lemmas and theorems for which the proofs are given in Bierens (2011) are indicated by an asterix: (*). Appendix B (section 10) deals with convergence in probability of projections of a random element of a Hilbert space on the space spanned by an array of random elements in this Hilbert space, and their corresponding residuals. These results play a key-role in the asymptotic normality proof in section 6.

Throughout the paper I will use the following notations. The indicator func-

tion is denoted by $I(\cdot)$, and \mathbb{N} and \mathbb{N}_0 denote the sets of positive and nonnegative integers, respectively. The partial derivative to a parameter with index k will be denoted by ∇_k , and $\nabla_{k,m}$ denotes the second partial derivatives to parameters with indices k and m .

2. The SNP discrete choice index model

2.1. The benchmark model

As benchmark model, I will focus on the SNP discrete choice index model

$$\Pr [Y = 1|X] = F_0(\alpha_0 + \beta_0'X) \quad (2.1)$$

where $Y \in \{0, 1\}$, $X \in \mathbb{R}^q$, $q \geq 1$, is a (vector) of observable covariates, $F_0(x)$ is an unknown absolutely continuous distribution function on \mathbb{R} with density $f_0(x)$, and $\alpha_0 \in \mathbb{R}$ and $\beta_0 \in \mathbb{R}^q$ are parameters to be estimated. Moreover, similar to the Logit and Probit cases it will be assumed that

$$f_0(x) \text{ is continuous and positive on } \mathbb{R}. \quad (2.2)$$

Given an a priori chosen absolutely continuous distribution function $G(x)$ with support \mathbb{R} we can write model (2.1) as

$$\Pr [Y = 1|X] = H_0(G(\alpha_0 + \beta_0'X)) = H_0(G((1, X')\theta_0)), \quad (2.3)$$

where $\theta_0 = (\alpha_0, \beta_0)' \in \mathbb{R}^p$, $p = q + 1$, and $H_0(u)$ is a distribution function on $[0, 1]$, i.e., $H_0(u) = F_0(G^{-1}(u))$, with G^{-1} the inverse of G . The corresponding density takes the form

$$h_0(u) = f_0(G^{-1}(u))/g(G^{-1}(u)), \quad (2.4)$$

where g is the density function of G .

Note that by (2.4) and condition (2.2), $h_0(u)$ is continuous and positive on $(0, 1)$. Moreover, if

$$\lim_{|x| \rightarrow \infty} f_0(x)/g(x) < \infty \quad (2.5)$$

then $h_0(0) < \infty$ and $h_0(1) < \infty$, so that then $h_0(u)$ is uniformly continuous on $[0, 1]$.

In general, the role of the a priori chosen distribution function G is three-fold: 1. G specifies the support of the unknown distribution function F_0 in the SNP model;

2. G maps one-to-one the parameter space of absolutely continuous candidate distributions for F_0 onto a space of density functions on the unit interval, which enables us to develop a unified inference approach for a wide range of SNP models;

3. G serves as an initial guess for $F_0(x) = H_0(G(x))$. If the guess is right then $H_0(u) = u$. A related interpretation is that $F_0 = G$ represents a standard parametric model of which the SNP model is a generalization.

As to the latter, let for example $G(x)$ be the logistic distribution function,

$$G(x) = (1 + \exp(-x))^{-1}. \quad (2.6)$$

Then the standard Logit model corresponds to $H_0(u) = u$, and the general SNP discrete choice index model (2.3) corresponds to

$$h_0(u) = \frac{f_0(\ln(u) - \ln(1 - u))}{u(1 - u)}, \quad (2.7)$$

as follows straightforwardly from (2.4) and the fact that in the case (2.6), $g(x) = G(x)(1 - G(x))$ and $G^{-1}(u) = \ln(u/(1 - u))$. Thus, in the case (2.6) model (2.3) becomes a generalization of the standard Logit model. However, the choice of (2.6) does not exclude other standard parametric discrete choice models. For example, with G chosen as (2.6), the standard Probit model corresponds to

$$h_0(u) = \frac{\exp(-(\ln(u) - \ln(1 - u))^2/2)}{u(1 - u)\sqrt{2\pi}}.$$

2.2. Identification

As pointed out by Manski (1988), model (2.1) is not identified without further conditions. One of the reasons is that for any pair of constants μ and $\sigma > 0$ we can find a distribution function $H(u)$ on $[0, 1]$ such that $H_0(G(x)) = H(G(\mu + \sigma x))$, namely $H(u) = H_0(G((G^{-1}(u) - \mu)/\sigma))$, so that then $\Pr[Y = 1|X] = H(G(\mu + \sigma\alpha_0 + \sigma\beta'_0 X)) = H_0(G(\alpha_0 + \beta'_0 X))$. A possible solution is to set $\alpha_0 = 0$ and normalize one of the components of β_0 to 1, as in Manski (1988), but then we lose the desirable property that the case $F_0 = G$ corresponds to the uniform distribution $H_0(u) = u$.

Alternatively, we can achieve identification by imposing two quantile restrictions, for example

$$H(u_1) = H_0(u_1) = u_1, \quad H(u_2) = H_0(u_2) = u_2, \quad (2.8)$$

for an a priori chosen pair $u_1 < u_2$ in $(0, 1)$, together with some regularity conditions on the distribution of X . The reason for choosing the quantiles and the corresponding values of H and H_0 the same is to accommodate the uniform distribution. Then similar to the identification conditions for the interval censored mixed proportional hazard model considered in Bierens (2008) it can be shown that the SNP discrete choice index model (2.3) is identified under the following conditions.

Assumption 2.1. *Let $X \in \mathbb{R}^q$ be the vector of covariates in the SNP discrete choice model (2.3). The following conditions hold.*

- (a) $E[X'X] < \infty$.
- (b) If $q = 1$ then the distribution of X has support \mathbb{R} , and $\beta_0 \neq 0$.
- (c) If $q \geq 2$ then we can partition X as $X = (X_1, X_2)'$, with $X_2 \in \mathbb{R}^{q-1}$, such that the conditional distribution of X_1 given X_2 has support \mathbb{R} . Moreover, the coefficient of X_1 is nonzero and the variance matrix of X_2 is nonsingular.
- (d) The distribution function H_0 in (2.3) is absolutely continuous with density h_0 satisfying $h_0(u) > 0$ on $(0, 1)$.
- (e) The distribution function H_0 is confined to a space of absolutely continuous distribution functions H on $[0, 1]$ satisfying the quantile restrictions $H(u_1) = u_1$, $H(u_2) = u_2$ for an a priori chosen pair $u_1 \neq u_2$ in $(0, 1)$.

Thus,

Lemma 2.1.^(*) *Under Assumption 2.1, $H_0(G(\alpha_0 + \beta_0'X)) = H(G(\alpha + \beta'X))$ a.s. implies $\alpha = \alpha_0$, $\beta = \beta_0$ and $H \equiv H_0$.*

2.3. The SNP Logit model

From now onwards it will be assumed that the logistic distribution (2.6) has been chosen as initial guess $G(x)$. The corresponding SNP model (2.3) will be referred to as the SNP Logit model. Moreover, in addition to part (d) of Assumption 2.1 I will adopt condition (2.2) together with tail condition (2.5), so that

Assumption 2.2. *The density $h_0(u)$ in (2.7) is uniformly continuous on $[0, 1]$.*

I will propose to estimate the Euclidean parameter vector θ_0 and the distribution function H_0 consistently by a penalized sieve least squares method as well

as by a penalized sieve maximum likelihood method, on the basis of a random sample of size N from (Y, X) , where the role of the penalty function involved is to enforce the quantile restriction in part (e) of Assumption 2.1. In the penalized maximum likelihood case, however, we need to augment Assumption 2.2 with the conditions that

Assumption 2.3. $h_0(0) > 0$ and $h_0(1) > 0$,

in order to deal with the effect of the logs in the log-likelihood function.

Note that by (2.4), Assumption 2.3 implies $\lim_{|x| \rightarrow \infty} f_0(x)/g(x) > 0$, hence by (2.5), Assumptions 2.2 and 2.3 together require to choose G and its density g such that

$$\lim_{|x| \rightarrow \infty} f_0(x)/g(x) \in (0, \infty). \quad (2.9)$$

In particular, in the logistic case (2.6) the tail condition (2.9) reads

$$\lim_{|x| \rightarrow \infty} (\ln(f_0(x)) - |x|) \in \mathbb{R}, \quad (2.10)$$

which, admittedly, is restrictive. On the other hand, in the SNP Logit case $f_0(x)$ can be estimated consistently via penalized sieve least squares without requiring Assumptions 2.2 and 2.3 (see the remark following Theorem 5.1), so that the tail condition (2.10) is empirically verifiable. Moreover, if condition (2.10) does not seem to hold the plot of the estimate of $f_0(x)$ can be used to select a distribution function G for which tail condition (2.9) is plausible.

3. Series expansions of densities on the unit interval

3.1. Cosine series representation

In Bierens (2008) I have proposed a series representation of a density $h(u)$ on $[0, 1]$ based on orthonormal Legendre polynomials, because these polynomials form a complete orthonormal sequence in the Hilbert space $L^2(0, 1)$. However, the main problem with this representation is that the Legendre polynomials have to be computed recursively so that $h(u)$ has no closed form expression, and neither has the corresponding distribution function $H(u) = \int_0^u h(v)dv$. The same applies to the density and distribution function representations on the basis of Hermite polynomials advocated by Gallant and Nychka (1987).

The sequence of Legendre polynomials is not the only complete orthonormal sequence in $L^2(0, 1)$. As is well-known, the Fourier series $\rho_0(u) \equiv 1$, $\rho_k(u) = \sqrt{2} \sin(2k\pi u)$ if $k \in \mathbb{N}$ is odd, $\rho_k(u) = \sqrt{2} \cos(2k\pi u)$ if $k \in \mathbb{N}$ is even, is complete in $L^2(0, 1)$, and the same applies to the related cosine series $\rho_0(u) \equiv 1$, $\rho_k(u) = \sqrt{2} \cos(k\pi u)$ for $k \in \mathbb{N}$. See for example Kronmal and Tarter (1968) and Bierens (2011) for the latter. The advantage of using the cosine series instead of the Legendre polynomials is that then the series representations of $h(u)$ and $H(u)$ have closed forms. In particular, it follows similar to Bierens (2008) that

Lemma 3.1.^(*) *For an arbitrary density function $h(u)$ on $[0, 1]$ with corresponding distribution function $H(u)$ there exist possibly uncountable many sequences $\delta = \{\delta_m\}_{m=1}^\infty$ satisfying $\sum_{m=1}^\infty \delta_m^2 < \infty$ such that almost everywhere (a.e.) on $(0, 1)$,*

$$h(u) = h(u|\delta) = \frac{\left(1 + \sum_{k=1}^\infty \delta_k \sqrt{2} \cos(k\pi u)\right)^2}{1 + \sum_{m=1}^\infty \delta_m^2}, \quad (3.1)$$

$$\begin{aligned} H(u) &= H(u|\delta) \\ &= u + \frac{1}{1 + \sum_{i=1}^\infty \delta_i^2} \left[2\sqrt{2} \sum_{k=1}^\infty \delta_k \frac{\sin(k\pi u)}{k\pi} + \sum_{k=1}^\infty \delta_k^2 \frac{\sin(2k\pi u)}{2k\pi} \right. \\ &\quad \left. + 2 \sum_{k=2}^\infty \sum_{m=1}^{k-1} \delta_k \delta_m \frac{\sin((k+m)\pi u)}{(k+m)\pi} + 2 \sum_{k=2}^\infty \sum_{m=1}^{k-1} \delta_k \delta_m \frac{\sin((k-m)\pi u)}{(k-m)\pi} \right]. \end{aligned} \quad (3.2)$$

The result for $H(u)$ follows straightforwardly from (3.1) and the well-known sine-cosine formulas. The proof of the "a.e." part is standard and will therefore be given in Bierens (2011).

However, the lack of uniqueness of the δ_k 's is not too serious a problem because for most applications, including the SNP discrete choice model under review, the density $h(u)$ involved satisfies the conditions of the following lemma.

Lemma 3.2. *If a density $h(u)$ on $[0, 1]$ is continuous and positive on $(0, 1)$ then it has a unique series representation (3.1), with*

$$\delta_k = \frac{\int_0^1 \sqrt{2} \cos(k\pi u) \sqrt{h(u)} du}{\int_0^1 \sqrt{h(u)} du}, \quad k \in \mathbb{N}. \quad (3.3)$$

Proof: Appendix A.

Similar to Bierens (2008) the condition $\sum_{k=1}^{\infty} \delta_k^2 < \infty$ can be imposed by restriction the space of densities to the following space $\mathcal{D}(0, 1)$.

Definition 3.1. *Given an a priori chosen positive sequence $\{\bar{\delta}_k\}_{k=1}^{\infty}$ satisfying $\sum_{k=1}^{\infty} \bar{\delta}_k^2 < \infty$, let $\Delta = \mathbf{X}_{m=1}^{\infty}[-\bar{\delta}_m, \bar{\delta}_m]$ and $\mathcal{D}(0, 1) = \{h(u|\delta) : \delta \in \Delta\}$.*

It has been shown in Bierens (2008) that the space Δ endowed with the metric $\|\delta_1 - \delta_2\| = \sqrt{\sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m})^2}$, where $\delta_i = \{\delta_{i,m}\}_{m=1}^{\infty}$ for $i = 1, 2$, is compact, and so is the space $\mathcal{D}(0, 1)$ endowed with the L^1 metric $d(h_1, h_2) = \int_0^1 |h_1(u) - h_2(u)| du$.

Given that the sequence $\{\bar{\delta}_m\}_{m=1}^{\infty}$ is chosen such that $h_0 \in \mathcal{D}(0, 1)$, we can use either the space $\mathcal{D}(0, 1)$ or the space Δ as the non-Euclidean parameter space. Along the lines in Bierens (2008) it can be shown that the Euclidean parameters $\theta_0 = (\alpha_0, \beta_0)'$ together with the distribution function H_0 in (2.3) can be estimated consistently by a penalized nonlinear least squares sieve approach.

However, in this paper I want to focus on the asymptotic normality of the sieve estimator of θ_0 . As will be shown below, the latter requires that the densities (3.1) are differentiable in u as well as twice differentiable in the δ_k 's. Therefore, we need to impose smoothness conditions on (3.1). How to do that will be shown in the next subsection.

3.2. Smoothness

As argued before, under certain conditions and the appropriate choice of the initial guess G we may assume that the true density $h_0(u)$ corresponding to an SNP model is uniformly continuous on $[0, 1]$. C.f. Assumption 2.2. A sufficient condition for this is that the δ_k 's in the representation (3.1) of $h_0(u)$ satisfy $\sum_{k=1}^{\infty} |\delta_k| < \infty$, as is easy to verify. Moreover, if the true density $h_0(u)$ in an SNP model is ℓ times continuously differentiable on $(0, 1)$ we can impose this condition by restricting the δ_k 's in the representation (3.1) of $h_0(u)$ to those for which $\sum_{k=1}^{\infty} k^{\ell} |\delta_k| < \infty$. Again, this condition can be imposed by confining $h_0(u)$ to the following space $\mathcal{D}_{\ell}(0, 1)$.

Definition 3.2. *Given an $\ell \in \mathbb{N}_0$ and an a priori chosen positive sequence $\{\bar{\delta}_{\ell,m}\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty} m^{\ell} \bar{\delta}_{\ell,m} < \infty$, let $\Delta_{\ell} = \mathbf{X}_{m=1}^{\infty}[-\bar{\delta}_{\ell,m}, \bar{\delta}_{\ell,m}]$ and $\mathcal{D}_{\ell}(0, 1) = \{h(u|\delta) : \delta \in \Delta_{\ell}\}$.*

Thus,

Lemma 3.3. *For $\ell = 0$ the densities in $\mathcal{D}_0(0, 1)$ are uniformly continuous on $[0, 1]$, and for $\ell \in \mathbb{N}$, $h \in \mathcal{D}_\ell(0, 1)$, the derivatives $h^{(k)}(u) = d^k h(u)/(du)^k$, $k = 1, 2, \dots, \ell$, are uniformly continuous on $[0, 1]$.*

Moreover, we have:

Lemma 3.4.^(*) *Let $\delta = \{\delta_m\}_{m=1}^\infty \in \Delta_\ell$, $\delta_i = \{\delta_{i,m}\}_{m=1}^\infty \in \Delta_\ell$, $i = 1, 2$. Moreover, denote for $k = 1, \dots, \ell$, $\|\delta_1 - \delta_2\|_k = \sum_{m=1}^\infty m^k |\delta_{1,m} - \delta_{2,m}|$, $h^{(k)}(u|\delta) = d^k h(u|\delta)/(du)^k$, and let $h^{(0)}(u|\delta) = h(u|\delta)$, where the latter is defined by (3.1). Furthermore, let $C_k \in (0, \infty)$ be a generic constant. Then for $k = 0, 1, \dots, \ell$ and $m, m_1, m_2 \in \mathbb{N}$,*

$$\begin{aligned} \sup_{\delta \in \Delta_\ell} \sup_{0 \leq u \leq 1} |h^{(k)}(u|\delta)| &< C_k, \\ \sup_{0 \leq u \leq 1} |h^{(k)}(u|\delta_1) - h^{(k)}(u|\delta_2)| &< C_k \cdot \|\delta_1 - \delta_2\|_k, \\ \sup_{\delta \in \Delta_\ell} \sup_{0 \leq u \leq 1} |\nabla_m h^{(k)}(u|\delta)| &< C_k \cdot m^k, \\ \sup_{0 \leq u \leq 1} |\nabla_m h^{(k)}(u|\delta_1) - \nabla_m h^{(k)}(u|\delta_2)| &< C_k \cdot m^k \cdot \|\delta_1 - \delta_2\|_k, \\ \sup_{\delta \in \Delta_\ell} \sup_{0 \leq u \leq 1} |\nabla_{m_1, m_2} h^{(k)}(u|\delta)| &< C_k \cdot m_1^k m_2^k, \\ \sup_{0 \leq u \leq 1} |\nabla_{m_1, m_2} h^{(k)}(u|\delta_1) - \nabla_{m_1, m_2} h^{(k)}(u|\delta_2)| &< C_k \cdot m_1^k m_2^k \cdot \|\delta_1 - \delta_2\|_k. \end{aligned}$$

This lemma plays a key-role in proving asymptotic normality of the sieve estimator of the Euclidean parameter vector θ_0 in the SNP Logit model.

3.3. Compactness

Note that similar to Lemma A.1 in Bierens (2008),

Lemma 3.5.^(*) *The space Δ_ℓ endowed with the metric $\|\delta_1 - \delta_2\|_\ell = \sum_{m=1}^\infty m^\ell |\delta_{1,m} - \delta_{2,m}|$, where $\delta_i = \{\delta_{i,m}\}_{m=1}^\infty$ for $i = 1, 2$, is compact.*

Moreover, using the same notation as in Lemma 3.4 it follows that

Lemma 3.6. *The space $\mathcal{D}_\ell(0, 1)$ endowed with the Sobolev¹ metric $d_\ell(h_1, h_2) = \max_{0 \leq m \leq \ell} \sup_{0 \leq u \leq 1} |h_1^{(m)}(u) - h_2^{(m)}(u)|$ is compact.*

This result follows from Lemma 3.5 and the fact that for each pair $h_1, h_2 \in \mathcal{D}_\ell(0, 1)$ with corresponding sequences $\delta_1 = \{\delta_{1,m}\}_{m=1}^\infty$ and $\delta_2 = \{\delta_{2,m}\}_{m=1}^\infty$ we have $d_\ell(h_1, h_2) = O(\|\delta_1 - \delta_2\|_\ell)$.

4. Reformulation of the SNP Logit model

4.1. The parameter space

The space Δ_ℓ for some $\ell \in \mathbb{N}_0$, endowed with the metric $\|\delta_1 - \delta_2\|_\ell$ in Lemma 3.5 will now be used as the parameter space for the non-Euclidean parameter(s) in the SNP Logit model. Also the Euclidean parameter vector $\theta_0 = (\alpha_0, \beta'_0)$ is assumed to be confined to a compact parameter space $\Theta \subset \mathbb{R}^p$. In particular, for notational convenience it will be assumed that Θ is a hypercube. Thus,

Assumption 4.1. *The SNP Logit model is parametrized as $\Pr[Y = 1|X] = H(G((1, X')\theta_0)|\delta_0)$, where $\theta_0 = (\alpha_0, \beta'_0)' \in \Theta = \mathbf{X}_{i=1}^p[-\bar{\theta}_i, \bar{\theta}_i]$, $\delta_0 \in \Delta_\ell$ for some $\ell \in \mathbb{N}_0$, and $H(u|\delta)$ is given in (3.2).*

Next, denote for $\theta = (\theta_1, \dots, \theta_p) \in \Theta$ and $\delta = \{\delta_k\}_{k=1}^\infty \in \Delta_\ell$,

$$\begin{aligned} \xi &= (\theta, \delta) = \{\xi_k\}_{k=1}^\infty, \text{ where } \xi_k = \begin{cases} \theta_k & \text{for } k = 1, \dots, p, \\ \delta_{k-p} & \text{for } k \geq p + 1, \end{cases} \\ \xi^0 &= (\theta_0, \delta_0) = \{\xi_{0,k}\}_{k=1}^\infty \end{aligned}$$

and let

$$\Xi_\ell = \mathbf{X}_{k=1}^\infty[-\bar{\xi}_{\ell,k}, \bar{\xi}_{\ell,k}], \text{ where } \bar{\xi}_{\ell,k} = \begin{cases} \bar{\theta}_k & \text{for } k = 1, \dots, p, \\ \bar{\delta}_{\ell,k-p} & \text{for } k \geq p + 1, \end{cases} \quad (4.1)$$

$$\|\xi_1 - \xi_2\|_\ell = \sum_{m=1}^\infty m^\ell |\xi_{1,m} - \xi_{2,m}|, \text{ where } \xi_i = \{\xi_{i,m}\}_{m=1}^\infty \text{ for } i = 1, 2. \quad (4.2)$$

As to the metric on Ξ_ℓ , we may combine the Euclidean metric on Θ with the metric $\|\delta_1 - \delta_2\|_\ell$ on Δ_ℓ , for example the metric $\|\theta_1 - \theta_2\| + \|\delta_1 - \delta_2\|_\ell$. However,

¹See for example Adams and Fournier (2003).

because (4.1) implies $\sum_{m=1}^{\infty} m^{\ell} \bar{\xi}_{\ell,m} < \infty$, we may without loss of generality endow Ξ_{ℓ} with the metric (4.2). This metric is convenient in deriving asymptotic normality results, as we will see below.

Note that similar to Lemma 3.5, Ξ_{ℓ} is compact. Moreover, it is not hard to verify that

Lemma 4.1.^(*) $H(G((1, X')\theta)|\delta)$ is a.s. continuous in $\xi = (\theta, \delta) \in \Xi_{\ell}$.

4.2. The SNP Logit model in penalized least squares form

Denote

$$f_{LS}(Z, \xi) = (Y - H(G((1, X')\theta)|\delta))^2 + \Pi(\delta), \quad (4.3)$$

where $Z = (Y, X)'$, $\xi = (\theta, \delta)$, and $\Pi(\delta)$ is a penalty function to enforce the quantile restrictions (2.8). In particular, I will use

$$\Pi(\delta) = (u_1 - H(u_1|\delta))^4 + (u_2 - H(u_2|\delta))^4 \quad (4.4)$$

for given pair $u_1 \neq u_2$ in $(0, 1)$. The reason for the power 4 is that then the penalty function does not affect the asymptotic normality of the sieve estimator of θ_0 , as will become clear below.

Note that by Lemma 3.4, $\Pi(\delta)$ is uniformly continuous on Δ_{ℓ} , hence by Lemma 4.1, $f_{LS}(Z, \xi)$ is a.s. continuous on Ξ_{ℓ} . Moreover, because obviously $0 \leq f_{LS}(Z, \xi) \leq 33$, it follows that

$$E \left[\sup_{\xi \in \Xi_{\ell}} |f_{LS}(Z, \xi)| \right] < \infty, \quad (4.5)$$

hence by the bounded convergence theorem and the compactness of Ξ_{ℓ} , $E[f_{LS}(Z, \xi)]$ is uniformly continuous on Ξ_{ℓ} . Therefore

$$\xi^0 = \arg \min_{\xi \in \Xi_{\ell}} E[f_{LS}(Z, \xi)] \in \Xi_{\ell},$$

which by Assumption 2.1 and Lemma 3.3 is unique.

In this case the parameter ξ^0 can be estimated strongly consistent by sieve estimation, on the basis of a random sample Z_1, \dots, Z_N from the distribution of $Z = (Y, X)'$, as follows. Denote

$$\Xi_{\ell,n} = (\mathbf{X}_{k=1}^n [-\bar{\xi}_{\ell,k}, \bar{\xi}_{\ell,k}]) \times (\mathbf{X}_{k=n+1}^{\infty} \{0\}),$$

which is a sequence of sieve spaces of Ξ_ℓ , and note that $\Xi_\ell = \overline{\bigcup_{n=1}^{\infty} \Xi_{\ell,n}}$, where the bar denotes the closure. The latter follows from the fact that for any $\xi \in \Xi_\ell$ there exists a sequence $\xi_n \in \Xi_{\ell,n}$ such that $\lim_{n \rightarrow \infty} \|\xi_n - \xi\|_\ell = 0$. Then the following result holds.

Theorem 4.1. *Let n_N be an arbitrary subsequence of N satisfying $\lim_{N \rightarrow \infty} n_N = \infty$, and let $\hat{\xi}_n = \arg \min_{\xi \in \Xi_{\ell,n}} \frac{1}{N} \sum_{j=1}^N f_{LS}(Z_j, \xi)$. Under Assumptions 2.1 and 4.1, $\|\hat{\xi}_{n_N} - \xi^0\|_\ell \xrightarrow{\text{a.s.}} 0$ for any $\ell \in \mathbb{N}_0$.*

This result will be proved in more general terms in the next section. As will appear, (4.5) is a crucial condition for this result.

4.3. The SNP Logit model in penalized maximum likelihood form

Using the same notation as before, the penalized log-likelihood function of the SNP Logit model takes the form

$$f_{ML}(Z, \xi) = Y \cdot \ln(H(G((1, X')\theta)|\delta)) + (1 - Y) \cdot \ln(1 - H(G((1, X')\theta)|\delta)) - \Pi(\delta), \quad (4.6)$$

where $\Pi(\delta)$ is the penalty function (4.4). It is a standard maximum likelihood exercise to verify that $E[f_{ML}(Z, \xi)|X] \leq E[f_{ML}(Z, \xi^0)|X]$ a.s. and that $E[f_{ML}(Z, \xi)|X] = E[f_{ML}(Z, \xi^0)|X]$ a.s. if and only if $\Pi(\delta) = 0$ and $H(G((1, X')\theta)|\delta) = H(G((1, X')\theta_0)|\delta_0)$ a.s. As we have seen in Lemma 2.1, under Assumption 2.1 the latter implies $\theta = \theta_0$ and $\delta = \delta_0$. Thus,

$$\xi^0 = \arg \max_{\xi \in \Xi_\ell} E[f_{ML}(Z, \xi)]$$

is unique.

However, due to the logs it is possible that $E[f_{ML}(Z, \xi)] = -\infty$ for some ξ . For example, suppose that X is univariate and is distributed as G , and $\Pr[Y = 1|X] = G(X)$, so that $\theta_0 = (0, 1)'$ and $h_0(u) = 1$. Next, let $\xi_* = (\theta_0, \delta_*)$, where δ_* is such that $H(u|\delta_*) = u \cdot \exp(1 - u^{-2})$. Because $U = G(X)$ is uniformly $[0, 1]$ distributed it follows that $E[f_{ML}(Z, \xi_*)] = \int_0^1 u \cdot \ln(H(u|\delta_*)) du + \int_0^1 (1 - u) \ln(1 - H(u|\delta_*)) du - \Pi(\delta_*) = -\infty$, where the latter is due to $\int_0^1 u \cdot \ln(H(u|\delta_*)) du = \int_0^1 u \cdot \ln(u) du + \frac{1}{2} - \int_0^1 u^{-1} du = -\infty$.

Nevertheless, it is possible to get around this problem, but at a price: We now need Assumption 2.3, and we have to trade in strong consistency for weak consistency.

Theorem 4.2. *Let n_N be an arbitrary subsequence of N satisfying $\lim_{N \rightarrow \infty} n_N = \infty$, and let $\widehat{\xi}_{n_N} = \arg \max_{\xi \in \Xi_{\ell, n_N}} f_{ML}(Z_j, \xi)$. Under Assumptions 2.1, 2.3 and 4.1, $\text{plim}_{N \rightarrow \infty} \|\widehat{\xi}_{n_N} - \xi^0\|_{\ell} = 0$ for any $\ell \in \mathbb{N}_0$.*

Again, this result will be proved in more general terms in the next section. The specialization of the general conditions involved to the case under review employs the fact that

Lemma 4.2.^(*) *Under the conditions of Theorem 4.2, $E[f_{ML}(Z, \xi)]$ is continuous in ξ^0 .*

5. Consistency of sieve estimators

5.1. General SNP model

More generally,

Assumption 5.1. *Consider an SNP model characterized by a real valued random function $f(Z, \xi)$ on Ξ , where Ξ is a (possibly non-Euclidean) compact metric space with metric $d(\xi_1, \xi_2)$, and Z is a random vector representing the data generating process. The support of Z is contained in a open subset \mathcal{Z} of a Euclidean space. The data set involved is $\{Z_j\}_{j=1}^N$, where the Z_j 's are independent replications of Z , defined on a common probability space $\{\Omega, \mathcal{F}, P\}$. Moreover, for each $z \in \mathcal{Z}$, $f(z, \xi)$ is continuous in $\xi \in \Xi$, and for each $\xi \in \Xi$, $f(z, \xi)$ is a Borel measurable function on \mathcal{Z} .*

Also, suppose that

Assumption 5.2. *In addition to the conditions in Assumption 5.1,*

- (a) *There exists an increasing sequence $\{\Xi_n\}_{n=1}^{\infty}$ of compact subspaces of Ξ that is dense in Ξ , i.e., $\Xi = \overline{\bigcup_{n=1}^{\infty} \Xi_n}$;*
- (b) *Each subspace Ξ_n corresponds one-to-one to a compact subset Σ_n of a Euclidean space;*
- (c) *$\sup_{\xi \in \Xi} f(Z, \xi) \leq 0$ a.s.;*
- (d) *The parameter of interest, $\xi^0 = \arg \max_{\xi \in \Xi} E[f(Z, \xi)]$, is a unique element of Ξ .*

These conditions hold for the SNP Logit case in penalized least squares form, with $f(Z, \xi) = -f_{LS}(Z, \xi)$, as well as for the SNP Logit case in penalized maximum likelihood form, with $f(Z, \xi) = f_{ML}(Z, \xi)$, where $\Xi = \Xi_\ell$, $\Xi_n = \Xi_{\ell, n}$, $\Sigma_n = \mathbf{X}_{k=1}^n[-\bar{\xi}_{\ell, k}, \bar{\xi}_{\ell, k}]$ and $d(\xi_1, \xi_2) = \|\xi_1 - \xi_2\|_\ell$, for any $\ell \in \mathbb{N}_0$.

Condition (c) may be replaced by $\sup_{\xi \in \Xi} f(Z, \xi) < \infty$ a.s. because then Assumptions 5.1 and 5.2 still hold for $f_0(Z, \xi) = f(Z, \xi) - \sup_{\xi_* \in \Xi} f(Z, \xi_*)$.

The roles of parts (a) and (b) of Assumption 5.2 are two-fold. First, these conditions guarantee that, with

$$\widehat{Q}_N(\xi) = \frac{1}{N} \sum_{j=1}^N f(Z_j, \xi), \quad (5.1)$$

the computation of the sieve estimator

$$\widehat{\xi}_n = \arg \max_{\xi \in \Xi_n} \widehat{Q}_N(\xi) \quad (5.2)$$

is feasible. Second, they enable us to generalize Jennrich's (1969) Lemma 2:

Lemma 5.1.^(*) *Under Assumption 5.1 and parts (a) and (b) of Assumption 5.2, $\sup_{\xi \in \Xi} f(z, \xi)$ and $\inf_{\xi \in \Xi} f(z, \xi)$ are Borel measurable functions on \mathcal{Z} . Moreover, let $\bar{\xi}(z) = \arg \max_{\xi \in \Xi} f(z, \xi)$, $\underline{\xi}(z) = \arg \min_{\xi \in \Xi} f(z, \xi)$.² There exist versions of $\bar{\xi}(z)$ and $\underline{\xi}(z)$ such that for any continuous real function Φ on Ξ , $\Phi(\bar{\xi}(z))$ and $\Phi(\underline{\xi}(z))$ are Borel measurable functions on \mathcal{Z} .*

An example of such a continuous function Φ is the metric $d(\xi, \xi^0)$ with $\xi^0 \in \Xi$ fixed and ξ variable, because by the triangular inequality $|d(\xi_1, \xi^0) - d(\xi_2, \xi^0)| \leq d(\xi_1, \xi_2)$. Hence, $d(\bar{\xi}(z), \xi^0)$ is Borel measurable and therefore $d(\bar{\xi}(Z), \xi^0)$ is a well-defined random variable. The same applies to $d(\widehat{\xi}_n, \xi^0)$.

5.2. Strong consistency

Recall that in the case of the SNP Logit model in least squares form we have

Assumption 5.3. $E[\sup_{\xi \in \Xi} |f(Z, \xi)|] < \infty$.

²Note that $\bar{\xi}(z)$ and/or $\underline{\xi}(z)$ may not be unique.

This assumption guarantees, by the dominated convergence theorem, that

$$\overline{Q}(\xi) = E[f(Z, \xi)] \quad (5.3)$$

is continuous on Ξ , so that by the compactness of Ξ , $\xi^0 \in \Xi$.

The core of the strong consistency proof for the sieve estimator of ξ^0 is the following generalization of the uniform strong law of large numbers (USLLN) of Jennrich (1969):

Lemma 5.2. *Under Assumptions 5.1 and 5.3, $\sup_{\xi \in \Xi} \left| \widehat{Q}_N(\xi) - \overline{Q}(\xi) \right| \xrightarrow{\text{a.s.}} 0^3$ as $N \rightarrow \infty$, where $\widehat{Q}_N(\xi)$ and $\overline{Q}(\xi)$ are defined by (5.1) and (5.3), respectively.*

Jennrich (1969) proved this result for the case that Ξ is a compact subset of a Euclidean space. However, inspecting the more detailed proof of Jennrich's USLLN in Bierens (2004, Appendix to Chapter 6), and applying Lemma 5.1, it follows straightforwardly that Jennrich's result carries over to general compact metric spaces.

Using Lemma 5.2 it is easy to prove the following standard strong consistency result for sieve estimators, similar to White and Wooldridge (1991).

Theorem 5.1.^(*) *Under Assumptions 5.1-5.3, $d(\widehat{\xi}_{n_N}, \xi^0) \xrightarrow{\text{a.s.}} 0$, where $\widehat{\xi}_n$ is defined by (5.2) and n_N is an arbitrary subsequence of the sample size N satisfying $\lim_{N \rightarrow \infty} n_N = \infty$.*

Remark. Theorem 4.1 is now a straightforward corollary of Theorem 5.1. However, it is not hard to verify from the conditions of Theorem 5.1 that Theorem 4.1 carries over if we replace Ξ_ℓ by $\Xi = \Theta \times \mathcal{D}(0, 1)$, where $\mathcal{D}(0, 1)$ is defined in Definition 3.1, the sieve spaces $\Xi_{\ell, n}$ by $\Xi_n = \Theta \times \mathcal{D}_n(0, 1)$ where $\mathcal{D}_n(0, 1) = \{h(u) = h(u|\delta) : \delta \in \Delta_n\}$ with $\Delta_n = (\mathbf{X}_{m=1}^n[-\bar{\delta}_m, \bar{\delta}_m]) \times (\mathbf{X}_{m=n+1}^\infty\{0\})$, and the metric $\|\xi_1 - \xi_2\|_\ell$ by the metric $\|\theta_1 - \theta_2\| + \int_0^1 |h_1(u) - h_2(u)| du$, for example.

5.3. Weak consistency under weak conditions

Suppose that instead of Assumption 5.3 the following conditions hold.

³Note that by Lemma 5.1 the supremum involved is a well-defined random variable.

Assumption 5.4.

- (a) There exists an element $\underline{\xi} \in \Xi$, $\underline{\xi} \neq \xi^0$, such that $E[f(Z, \underline{\xi})] > -\infty$;
(b) $E[f(Z, \xi)]$ is continuous in ξ^0 on the space $\Xi = \{\xi \in \Xi : E[f(Z, \xi)] \geq E[f(Z, \underline{\xi})]\}$, i.e., $\lim_{\varepsilon \downarrow 0} \inf_{\xi \in \Xi, d(\xi, \xi^0) < \varepsilon} E[f(Z, \xi)] = E[f(Z, \xi^0)]$.

In the case (4.6) let for example $\underline{\xi} = (0, 0, 0, \dots)$, as then $f(Z, \underline{\xi}) = -\ln(2)$. Moreover, condition (b) follows from the conditions of Lemma 4.2.

The following special case of the uniform weak law of large numbers plays a key-role in proving weak consistency of the sieve estimator involved.

Lemma 5.3.^(*) For $K > 0$, let $\widehat{Q}_{K,N}(\xi) = \frac{1}{N} \sum_{j=1}^N \max(f(Z_j, \xi), -K)$ and $\overline{Q}_K(\xi) = E[\max(f(Z, \xi), -K)]$. Under Assumption 5.1 and part (c) of Assumption 5.2 there exists a sequence K_N converging to infinity with N such that $\text{plim}_{N \rightarrow \infty} \sup_{\xi \in \Xi} |\widehat{Q}_{K_N,N}(\xi) - \overline{Q}_{K_N}(\xi)| = 0$.

This result is part of the proof of Theorem 10 in Bierens (2008). A slightly improved version of the proof of Lemma 5.3 is given in Bierens (2011).

Lemma 5.3 will now be used to prove that

Theorem 5.2. Under Assumptions 5.1, 5.2 and 5.4, $\text{plim}_{N \rightarrow \infty} d(\widehat{\xi}_{n_N}, \xi^0) = 0$, where $\widehat{\xi}_n$ is defined by (5.2) and n_N is an arbitrary subsequence of the sample size N satisfying $\lim_{N \rightarrow \infty} n_N = \infty$.

Proof: Appendix A.

Remark 1. Theorem 5.2 is a substantial improvement of Theorems 10 and 11 in Bierens (2008). Only Lemma 5.3 is taken from the proof of Theorem 10 in Bierens (2008); the rest of the proof of Theorem 5.2 is new. The results in Bierens (2008) were based on the assumption that for all $\xi \in \Xi$, $E[|f(Z, \xi)|] < \infty$. However, this condition may not hold for SNP models in log-likelihood form.

Remark 2. Theorem 5.2 is (somewhat) related to Theorem 5.14 in Van der Vaart (1998), which also allows for $E[f(Z, \xi)] = -\infty$ for some values of ξ . However, the key condition in the latter theorem is that, in our notation, $\widehat{Q}_N(\widehat{\xi}_{n_N}) \geq \widehat{Q}_N(\xi^0) - o_p(1)$, which may not hold for sieve estimators. Moreover, this theorem also assumes compactness of the parameter space.

6. Asymptotic normality

6.1. Introduction

As mentioned in the introduction, the conditions for asymptotic normality of sieve estimators proposed in the literature are based on high-level and difficult to verify assumptions. Therefore, in empirical applications it is usually assumed that the SNP model involved represents merely a flexible functional form. See for example Gabler et al. (1993). This is not unreasonable an assumption, as all econometric and statistical models are approximations of data-generating processes. In our case this assumption amounts to the condition that $\Xi = \Xi_n$ for some unknown n , so that there exists a smallest \underline{n} such that $\Xi = \Xi_{\underline{n}}$. The order \underline{n} can be estimated consistently using an information criterion similar to the well-known Hannan-Quinn (1979) and Schwarz (1978) information criteria for the dimension of time series models. Given an estimator \hat{n}_N of \underline{n} satisfying $\lim_{N \rightarrow \infty} \Pr[\hat{n}_N = \underline{n}] = 1$, one may treat the estimator \hat{n}_N as the true value. The model then becomes fully parametric, and therefore asymptotic normality can be derived in a standard way, provided that the SNP parameters involved are unique. This is the approach followed by Bierens and Carvalho (2007), for example.

In this section I will propose an alternative asymptotic normality proof on the basis of the standard mean value approach for the first-order conditions. The problem of the expanding dimension of the mean value equations will be solved by converting them to random functions. Asymptotic normality can then be derived from the functional central limit theorem, and the parameters of interest can be singled out by using projection residuals.

6.2. The model

Consider again the model in Assumptions 5.1 and 5.2, where now

Assumption 6.1. *In addition to Assumptions 5.1 and 5.2,*

- (a) $\Xi = \mathbf{X}_{k=1}^{\infty}[-\bar{\xi}_k, \bar{\xi}_k]$, with $\bar{\xi}_k$ a given sequence of positive numbers satisfying $\sum_{k=1}^{\infty} k^{\ell} \bar{\xi}_k < \infty$ for some natural number $\ell \geq 1$.
- (b) The space Ξ is endowed with norm $\|\xi\|_{\ell} = \sum_{k=1}^{\infty} k^{\ell} |\xi_k|$, $\xi = (\xi_1, \xi_2, \xi_3, \dots) \in \Xi$, and associated metric $\|\xi_1 - \xi_2\|_{\ell}$, so that in Assumption 5.1, $d(\xi_1, \xi_2) = \|\xi_1 - \xi_2\|_{\ell}$;
- (c) The sieve spaces involved are of the form $\Xi_n = (\mathbf{X}_{k=1}^n[-\bar{\xi}_k, \bar{\xi}_k]) \times (\mathbf{X}_{k=n+1}^{\infty}\{0\})$.
- (d) $f(Z, \xi)$ is a.s. twice continuously differentiable in the components of $\xi = (\xi_1, \xi_2, \xi_3, \dots) \in \Xi$;

(e) $\xi^0 = (\xi_{0,1}, \xi_{0,2}, \xi_{0,3}, \dots) \in \Xi^{\text{Int}} = \mathbf{X}_{k=1}^{\infty}(-\bar{\xi}_k, \bar{\xi}_k)$;

In the SNP Logit cases (4.3) and (4.6) the non-Euclidean parameters correspond to the parameter sequence $\delta = \{\delta_m\}_{m=1}^{\infty}$ in $H(G((1, X')\theta)|\delta)$ and $\Pi(\delta)$. It is not hard to verify that in these cases the second derivatives $\partial^2 f(Z, \xi)/(\partial \xi_{i_1} \partial \xi_{i_2})$ involve $G((1, X')\theta)$, $H(u|\delta)$, $h(u|\delta)$, and derivatives of the type $\partial h(u|\delta)/\partial \delta_k$, $\partial^2 h(u|\delta)/(\partial \delta_k \partial \delta_m)$ and $\partial h(u|\delta)/\partial u$. In view of the latter and Lemma 3.4, the choice $\ell = 1$ suffices. However, to accommodate SNP models for which $h(u|\delta)$ enters the expression for $f(Z, \xi)$ directly the asymptotic normality results below will be derived for a general $\ell \geq 1$.

As before, let $\hat{\xi}_n = (\hat{\xi}_{n,1}, \hat{\xi}_{n,2}, \hat{\xi}_{n,3}, \dots, \hat{\xi}_{n,n}, 0, 0, 0, \dots) = \arg \max_{\xi \in \Xi_{\ell,n}} \frac{1}{N} \sum_{j=1}^N f(Z_j, \xi)$ be the sieve estimator. It will be assumed that the conditions of Theorem 5.1 or 5.2 hold, so that the sieve estimator $\hat{\xi}_n$ is weakly consistent.

Assumption 6.2. *For any subsequence n of the sample size N satisfying $n \rightarrow \infty$ as $N \rightarrow \infty$, $\text{plim}_{N \rightarrow \infty} \|\hat{\xi}_n - \xi^0\|_{\ell} = 0$.*

6.3. First-order conditions and mean value expansion

Denote $f_j(\xi) = f(Z_j, \xi)$ and $\xi_n^0 = (\xi_{0,1}, \xi_{0,2}, \xi_{0,3}, \dots, \xi_{0,n}, 0, 0, 0, \dots)$. It follows from the mean value theorem that there exist a sequence $\lambda_k \in [0, 1]$ of random variables (depending on n and N as well) such that for $k = 1, \dots, n$,

$$\begin{aligned} & \frac{1}{\sqrt{N}} \sum_{j=1}^N \nabla_k f_j(\hat{\xi}_n) + \frac{1}{\sqrt{N}} \sum_{j=1}^N (\nabla_k f_j(\xi^0) - \nabla_k f_j(\xi_n^0)) \\ &= \frac{1}{\sqrt{N}} \sum_{j=1}^N \nabla_k f_j(\xi^0) \\ &+ \sum_{m=1}^n \left(\frac{1}{N} \sum_{j=1}^N \nabla_{k,m} f_j(\xi_n^0 + \lambda_k(\hat{\xi}_n - \xi_n^0)) \right) \sqrt{N} (\hat{\xi}_{n,m} - \xi_{0,m}) \end{aligned} \tag{6.1}$$

In the case that n is fixed, i.e., $\xi_n^0 = \xi^0$, it follows from part (e) of Assumption 6.1 and Assumption 6.2 that $\lim_{N \rightarrow \infty} \Pr[\hat{\xi}_n \in \mathbf{X}_{k=1}^n(-\bar{\xi}_k, \bar{\xi}_k)] = 1$, so that $\lim_{N \rightarrow \infty} \Pr[\sum_{j=1}^N \nabla_k f_j(\hat{\xi}_n) = 0 \text{ for } k = 1, \dots, n] = 1$. However, this may not be true for $n \rightarrow \infty$. Nevertheless,

Lemma 6.1.^(*) *Under Assumptions 6.1 and 6.2 there exists a sub-sequence K_n of n , i.e., $K_n \leq n$, $\lim_{n \rightarrow \infty} K_n = \infty$, such that $\lim_{N \rightarrow \infty} \Pr[\sum_{j=1}^N \nabla_k f_j(\widehat{\xi}_n) = 0 \text{ for } k = 1, \dots, K_n] = 1$.*

Of course, the subsequence K_n is unknown, but that does not matter for the asymptotic normality results to be derived below.

In order to convert (6.1) to a single equation in random function form, denote

$$\widehat{W}_n(u) = \sum_{k=1}^{K_n} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N \nabla_k f_j(\widehat{\xi}_n) \right) \eta_k(u) \quad (6.2)$$

$$\widehat{V}_n(u) = \sum_{k=1}^{K_n} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N (\nabla_k f_j(\xi^0) - \nabla_k f_j(\xi_n^0)) \right) \eta_k(u) \quad (6.3)$$

$$\widehat{Z}_n(u) = \sum_{k=1}^{K_n} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N \nabla_k f_j(\xi^0) \right) \eta_k(u) \quad (6.4)$$

$$\widehat{b}_{m,n}(u) = - \sum_{k=1}^{K_n} \left(\frac{1}{N} \sum_{j=1}^N \nabla_{k,m} f_j(\xi_n^0 + \lambda_k(\widehat{\xi}_n - \xi_n^0)) \right) \eta_k(u) \quad (6.5)$$

where K_n is the subsequence in Lemma 6.1 and the $\eta_k(u)$'s are orthogonal weight functions on $[0, 1]$, i.e., $\int_0^1 \eta_k(u) \eta_m(u) du = 0$ for $k \neq m$. As we will see, the type of $\eta_k(u)$ is not important, but we need to require that $\sigma_k = \int_0^1 \eta_k(u)^2 du$ converges fast enough to zero as $k \rightarrow \infty$. Therefore, I will choose

$$\eta_k(u) = 2^{-k} \sqrt{2} \cos(k\pi u). \quad (6.6)$$

We can now write the system of equations (6.1) as

$$\sum_{m=1}^n \widehat{b}_{m,n}(u) \sqrt{N} (\widehat{\xi}_{n,m} - \xi_{0,m}) = \widehat{Z}_n(u) - \widehat{W}_n(u) - \widehat{V}_n(u) \quad (6.7)$$

First, note that by (6.2) and Lemma 6.1,

$$\sup_{0 \leq u \leq 1} |\widehat{W}_n(u)| = o_p(1). \quad (6.8)$$

Next, observe from (6.6) and (6.3) that $E[\sup_{0 \leq u \leq 1} |\widehat{V}_n(u)|] \leq \sqrt{2N} \sum_{k=1}^n 2^{-k} E[|\nabla_k f_1(\xi^0) - \nabla_k f_1(\xi_n^0)|]$, which converges to zero if

Assumption 6.3. *There exists a nonnegative integer $\ell_0 \leq \ell$ such that the following local Lipschitz conditions hold for all $k \in \mathbb{N}$: $E[|\nabla_k f_1(\xi^0) - \nabla_k f_1(\xi_n^0)|] \leq M_k \cdot \|\xi^0 - \xi_n^0\|_{\ell_0}$, where $\sum_{k=1}^{\infty} 2^{-k} M_k < \infty$, and the sieve order $n = n_N$ is chosen such that $\lim_{N \rightarrow \infty} \sqrt{N} \sum_{m=n_N+1}^{\infty} m^{\ell_0} \bar{\xi}_m = 0$.*

Thus under Assumption 6.3,

$$\sup_{0 \leq u \leq 1} |\widehat{V}_n(u)| = o_p(1). \quad (6.9)$$

Equation (6.7) now reads $\sum_{m=1}^n \widehat{b}_{m,n}(u) \sqrt{N} (\widehat{\xi}_{n,m} - \xi_{0,m}) = \widehat{Z}_n(u) + o_p(1)$, where the $o_p(1)$ term is uniform in $u \in [0, 1]$.

6.4. Weak convergence

The next step is to set forth conditions such that \widehat{Z}_n converges weakly to a zero mean Gaussian process Z , as follows. Suppose that

Assumption 6.4. *For all $k \in \mathbb{N}$, $E[\nabla_k f_1(\xi^0)] = 0$.*

More primitive conditions for this assumption can be derived on the basis of the dominated convergence theorem such that $E[\nabla_k f_1(\xi^0)] = \nabla_k E[f_1(\xi^0)] = 0$, where the latter follows from the first-order conditions of a maximum of $E[f_1(\xi)]$ in ξ^0 .

Moreover, suppose that

Assumption 6.5. $\sum_{k=1}^{\infty} k \cdot 2^{-k} E[(\nabla_k f_1(\xi^0))^2] < \infty$.

Then using Theorem 8.2 in Billingsley (1968) it can be shown that

Lemma 6.2.^(*) *Under Assumptions 6.4 and 6.5, $\widehat{Z}_n \Rightarrow Z$ on $[0, 1]$, where Z is a zero-mean Gaussian process with covariance function*

$$\begin{aligned} \Gamma(u_1, u_2) &= E[Z(u_1)Z(u_2)] \\ &= \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} E[(\nabla_k f_1(\xi^0))(\nabla_m f_1(\xi^0))] \eta_k(u_1) \eta_m(u_2). \end{aligned} \quad (6.10)$$

Consequently, Assumptions 6.1-6.5 imply that

$$\sum_{m=1}^n \widehat{b}_{m,n}(u) \sqrt{N}(\widehat{\xi}_{n,m} - \xi_{0,m}) \Rightarrow Z(u). \quad (6.11)$$

6.5. Extracting the parameters of interest via projection residuals

Recall that in the case of the SNP discrete choice model the parameters $\xi_{0,1}, \dots, \xi_{0,p}$ correspond to the components of the Euclidean parameter vector θ_0 . Therefore, in order to determine the limiting distribution of $\sqrt{N}(\widehat{\xi}_{n,1} - \xi_{0,1}, \dots, \widehat{\xi}_{n,p} - \xi_{0,p})'$, we need to eliminate $\sum_{m=p+1}^n \widehat{b}_{m,n}(u) \sqrt{N}(\widehat{\xi}_{n,m} - \xi_{0,m})$ from (6.11). A possible way to do that is to project each $\widehat{b}_{m,n}(u)$ for $m \leq p$ on the space spanned by $\widehat{b}_{p+1,n}(u), \dots, \widehat{b}_{n,n}(u)$, and use the residuals $\widehat{a}_{m,n}(u)$ involved to wipe out the functions $\widehat{b}_{p+1,n}(u), \dots, \widehat{b}_{n,n}(u)$, as follows. Denote

$$\widehat{a}_n(u) = (\widehat{a}_{1,n}(u), \dots, \widehat{a}_{p,n}(u))',$$

and note that by the standard properties of projection residuals, $\int_0^1 \widehat{a}_n(u) (\widehat{b}_{p+1,n}(u), \dots, \widehat{b}_{n,n}(u)) du = O_{p,n-p}{}^4$ and $\int_0^1 \widehat{a}_n(u) (\widehat{b}_{1,n}(u), \dots, \widehat{b}_{p,n}(u)) du = \int_0^1 \widehat{a}_n(u) \widehat{a}_n(u)' du$. Hence by (6.7),

$$\int_0^1 \widehat{a}_n(u) \widehat{a}_n(u)' du \sqrt{N} \begin{pmatrix} \widehat{\xi}_{n,1} - \xi_{0,1} \\ \vdots \\ \widehat{\xi}_{n,p} - \xi_{0,p} \end{pmatrix} = \int_0^1 \widehat{a}_n(u) \left(\widehat{Z}_n(u) - \widehat{W}_n(u) - \widehat{V}_n(u) \right) du.$$

Now suppose that there exists a non-random \mathbb{R}^p -valued function $a(u)$ on $[0, 1]$ such that

$$\text{plim}_{N \rightarrow \infty} \int_0^1 (\widehat{a}_n(u) - a(u))' (\widehat{a}_n(u) - a(u)) du = 0 \quad (6.12)$$

and

$$\int_0^1 a(u)' a(u) du < \infty. \quad (6.13)$$

Then it follows that

$$\text{plim}_{N \rightarrow \infty} \int_0^1 \widehat{a}_n(u) \widehat{a}_n(u)' du = \int_0^1 a(u) a(u)' du, \quad (6.14)$$

⁴Here and in the sequel $O_{k,m}$ denotes the $k \times m$ zero matrix.

as is easy to verify, whereas by (6.8), (6.9), (6.12) and Lemma 6.2,

$$\int_0^1 \widehat{a}_n(u)(\widehat{Z}_n(u) - \widehat{W}_n(u) - \widehat{V}_n(u))du \xrightarrow{d} \int_0^1 a(u)Z(u)du.$$

Hence,

$$\sqrt{N}(\widehat{\xi}_{n,1} - \xi_{0,1}, \dots, \widehat{\xi}_{n,p} - \xi_{0,p})' \xrightarrow{d} \left(\int_0^1 a(u)a(u)'du \right)^{-1} \int_0^1 a(u)Z(u)du,$$

provided that

$$\det \left(\int_0^1 a(u)a(u)'du \right) > 0. \quad (6.15)$$

6.6. Convergence of the projection residuals

The next step is to determine the probability limit a of \widehat{a}_n , by specializing the conditions of Theorem B.1 in Appendix B to the present case. According to the latter theorem, for proving (6.12) we need to show that there exist nonrandom functions $b_m(u)$ such that for $m = 1, \dots, p$,

$$\|\widehat{b}_{m,n} - b_m\| = \sqrt{\int_0^1 (\widehat{b}_{m,n}(u) - b_m(u))^2 du} = o_p(1), \quad (6.16)$$

and that there exists a sequence ρ_m of positive numbers such that

$$\sum_{m=p+1}^n \rho_m \|\widehat{b}_{m,n} - b_m\| = o_p(1) \quad (6.17)$$

and

$$\liminf_{n \rightarrow \infty} \left\| \sum_{m=p+1}^n \rho_m b_m \right\| > 0. \quad (6.18)$$

Then (6.12) holds, with $a(u)$ the vector of residuals of the projections of $b_1(u)$, $b_2(u)$, \dots , $b_p(u)$ on $\text{span}(\{b_m(u)\}_{m=p+1}^\infty)$.

In view of (6.5) and (6.6), obvious candidates for the functions $b_m(u)$ are

$$b_m(u) = - \sum_{k=1}^{\infty} E[\nabla_{k,m} f_1(\xi^0)] 2^{-k} \sqrt{2} \cos(k\pi u) \quad (6.19)$$

Sufficient conditions for (6.16), (6.17) and (6.18) are that:

Assumption 6.6. For some $\tau \geq 0$,

- (a) $\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} (k.m)^{-2-\tau} E[|\nabla_{k,m} f_1(\xi^0)|] < \infty$, and
- (b) $\lim_{\varepsilon \downarrow 0} \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} (k.m)^{-2-\tau} E[\sup_{\|\xi - \xi^0\|_{\ell} \leq \varepsilon} |\nabla_{k,m} f_1(\xi) - \nabla_{k,m} f_1(\xi^0)|] = 0$.
- (c) For at least one pair $k, m \in \mathbb{N}$, $E[\nabla_{k,p+m} f_1(\xi^0)] \neq 0$.

Note that Assumption 6.6(a) implies

$$\begin{aligned} \int_0^1 b_m(u)^2 du &= \sum_{k=1}^{\infty} 2^{-2k} (E[\nabla_{k,m} f_1(\xi^0)])^2 \\ &\leq \left(\sum_{k=1}^{\infty} k^{4+2\tau} 2^{-2k} \right) \cdot \left(\sum_{k=1}^{\infty} k^{-2-\tau} E[|\nabla_{k,m} f_1(\xi^0)|] \right)^2 < \infty, \end{aligned}$$

so that (6.13) holds. Moreover,

Lemma 6.4. Under Assumptions 6.2, 6.3 and 6.6 the conditions (6.16), (6.17) and (6.18) hold for the b_m 's defined by (6.19), with $\rho_m = t^m/m!$ for some $t \in (0, 1)$

Proof. Appendix A.

6.7. Main results

Recall that $a(u)$ is the residual of the projection of $b(u) = (b_1(u), \dots, b_p(u))'$ on the Hilbert space $\mathcal{S}_{p+1}^{\infty} = \text{span}(\{b_{p+k}(u)\}_{k=1}^{\infty})$ spanned by the sequence $\{b_k(u)\}_{k=p+1}^{\infty}$. Obviously, (6.15) holds if and only if

$$\text{for all } \xi = (\xi_1, \dots, \xi_p)' \in \mathbb{R}^p \text{ with } \xi' \xi > 0, \sum_{m=1}^p \xi_m b_m(u) \notin \mathcal{S}_{p+1}^{\infty}. \quad (6.20)$$

This condition is equivalent to the condition that, with $\mathcal{S}_1^p = \text{span}(\{b_k(u)\}_{k=1}^p)$,

$$\mathcal{S}_1^p \cap \mathcal{S}_{p+1}^{\infty} = \{0\}, \quad (6.21)$$

the latter being the singleton of the zero function.

Now denote $\mathcal{S}_{p+1}^{p+n} = \text{span}(\{b_{p+k}(u)\}_{k=1}^n)$ and recall that

$$\mathcal{S}_{p+1}^{\infty} = \overline{\cup_{n=1}^{\infty} \mathcal{S}_{p+1}^{p+n}} = \left(\cup_{n=1}^{\infty} \mathcal{S}_{p+1}^{p+n} \right) \cup \mathcal{C}_{\infty}$$

where the bar denotes the closure, and $\mathcal{C}_\infty = \mathcal{S}_{p+1}^\infty \setminus (\cup_{n=1}^\infty \mathcal{S}_{p+1}^{p+n})$. Note that \mathcal{C}_∞ does not contain the zero function 0 because $0 \in \mathcal{S}_{p+1}^{p+n}$ for all $n \in \mathbb{N}$. Then (6.21) is true if and only if

$$\text{for all } n \in \mathbb{N}, \mathcal{S}_1^p \cap \mathcal{S}_{p+1}^{p+n} = \{0\} \quad (6.22)$$

and

$$\mathcal{S}_1^p \cap \mathcal{C}_\infty = \emptyset. \quad (6.23)$$

It is easy to verify that condition (6.22) is implied by the following more transparent and verifiable condition:

Assumption 6.7. Denote $\beta_{k,m} = E[\nabla_{k,m} f_1(\xi^0)]$ and

$$B_{k,m} = \begin{pmatrix} \beta_{1,1} & \cdots & \beta_{1,m} \\ \vdots & \ddots & \vdots \\ \beta_{k,1} & \cdots & \beta_{k,m} \end{pmatrix}.$$

For each $n > p$ there exists a $k \geq n$ such that $\text{rank}(B_{k,n}) = n$.

Note that for parametric models with n -dimensional parameter vector ξ^0 , Assumption 6.7 reduces to the standard assumption that the matrix $B_{n,n}$ is nonsingular.

Lemma 6.5. Let $a_n(u)$ be the residual of the projection of $b(u) = (b_1(u), \dots, b_p(u))'$ on $\text{span}(\{b_{p+k}(u)\}_{k=1}^n)$. Under Assumption 6.7, $\int_0^1 a_n(u)a_n(u)' du$ is non-singular for all $n \in \mathbb{N}$.

Proof. Appendix.

However, condition (6.23) is too difficult, if not impossible, to break down in more primitive and verifiable conditions. Therefore condition (6.23) has to be assumed, either directly or indirectly as:

Assumption 6.8. $\lim_{n \rightarrow \infty} \int_0^1 a_n(u)a_n(u)' du$ is non-singular as well.

Note that Assumption 6.7 together with condition (6.23) imply (6.20), which by

$$\int_0^1 a(u)a(u)' du = \lim_{n \rightarrow \infty} \int_0^1 a_n(u)a_n(u)' du$$

implies Assumption 6.8.

Now we have all the necessary conditions for our main result:

Theorem 6.1. *Let $b_m(u)$ be defined by (6.19), and let $a(u)$ be the vector of residuals of the projection of $b(u) = (b_1(u), \dots, b_p(u))'$ on $\text{span}(\{b_{p+m}(u)\}_{m=1}^\infty)$. Denote $\widehat{\theta}_n = (\widehat{\xi}_{n,1}, \dots, \widehat{\xi}_{n,p})'$ and $\theta_0 = (\xi_{0,1}, \dots, \xi_{0,p})'$. Under Assumptions 6.1-6.8,*

$$\sqrt{N}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \left(\int_0^1 a(u)a(u)'du \right)^{-1} \int_0^1 a(u)Z(u)du \sim \mathcal{N}_p[0, \Sigma],$$

where

$$\begin{aligned} \Sigma &= \left(\int_0^1 a(u)a(u)'du \right)^{-1} \left(\int_0^1 \int_0^1 a(u_1)\Gamma(u_1, u_2)a(u_2)'du_1du_2 \right) \\ &\quad \times \left(\int_0^1 a(u)a(u)'du \right)^{-1} \end{aligned}$$

with $\Gamma(u_1, u_2)$ defined by (6.10).

Because the way asymptotic normality is proved should not matter for the asymptotic normality result, the asymptotic variance matrix Σ must be invariant for the choice of the weight functions $\eta_k(u)$ defined by (6.6). On the other hand, different specifications of $\eta_k(u)$ yield different functions $\widehat{b}_{m,n}(u)$ and therefore different residual vectors $\widehat{a}_n(u)$. Thus, at first sight the result of Theorem 6.1 seems paradoxical.

This paradox will be solved by constructing a consistent estimator of Σ , as follows.

6.8. A consistent variance estimator

To estimate Σ consistently, we need a consistent estimator of the covariance function $\Gamma(u_1, u_2)$. The expression for the latter in (6.10) suggests the estimator

$$\widehat{\Gamma}_n(u_1, u_2) = \sum_{k=1}^n \sum_{m=1}^n \frac{1}{N} \sum_{j=1}^N (\nabla_k f_j(\widehat{\xi}_n)) (\nabla_m f_j(\widehat{\xi}_n)) \eta_k(u_1) \eta_m(u_2).$$

To prove the uniform consistency of $\widehat{\Gamma}_n(u_1, u_2)$ we need the condition that

Assumption 6.9. $\lim_{\varepsilon \downarrow 0} \sum_{k=1}^{\infty} 2^{-k} E[\sup_{\|\xi - \xi^0\|_{\ell} \leq \varepsilon} (\nabla_k f_1(\xi) - \nabla_k f_1(\xi^0))^2] = 0.$

Then it is not hard to verify that

Lemma 6.6.^(*) *Under Assumptions 6.2, 6.5 and 6.9, $\text{plim}_{N \rightarrow \infty} \sup_{(u_1, u_2) \in [0,1] \times [0,1]} |\widehat{\Gamma}_n(u_1, u_2) - \Gamma(u_1, u_2)| = 0.$*

Next, modify (6.5) to

$$\widehat{b}_{m,n}(u) = - \sum_{k=1}^n \left(\frac{1}{N} \sum_{j=1}^N \nabla_{k,m} f_j(\widehat{\xi}_n) \right) \eta_k(u), \quad (6.24)$$

and let now $\widehat{a}_n(u)$ be based on (6.24). Then obviously Lemma 6.4 carries over, so that (6.14) carries over, hence under the conditions of Theorem 6.1 and Lemma 6.6,

$$\begin{aligned} \widehat{\Sigma}_n &= \left(\int_0^1 \widehat{a}_n(u) \widehat{a}_n(u)' du \right)^{-1} \left(\int_0^1 \widehat{a}_n(u_1) \widehat{\Gamma}_n(u_1, u_2) \widehat{a}_n(u_2)' du_1 du_2 \right) \\ &\quad \times \left(\int_0^1 \widehat{a}_n(u) \widehat{a}_n(u)' du \right)^{-1} \end{aligned} \quad (6.25)$$

is a consistent estimator of Σ .

Theorem 6.2.^(*) *Let the conditions of Theorem 6.1 hold and let $n > p$. Denote $\widehat{\beta}_{k,m} = \frac{1}{N} \sum_{j=1}^N \nabla_{k,m} f_j(\widehat{\xi}_n)$, $\widehat{\gamma}_{k,m} = \frac{1}{N} \sum_{j=1}^N (\nabla_k f_j(\widehat{\xi}_n)) (\nabla_m f_j(\widehat{\xi}_n))$, and consider the matrices $\widehat{B}_{1,n} = (\widehat{\beta}_{k,m}; k = 1, \dots, n, m = 1, \dots, p)$, $\widehat{B}_{2,n} = (\widehat{\beta}_{k,m}; k = 1, \dots, n, m = p + 1, \dots, n)$, $\widehat{B}_n = (\widehat{B}_{1,n}, \widehat{B}_{2,n})$, and $\widehat{C}_n = (\widehat{\gamma}_{k,m}; k, m = 1, \dots, n)$. Suppose that $\text{rank}(\widehat{B}_n) = n$. Then the matrix $\widehat{\Sigma}_n$ in (6.25) takes the form*

$$\widehat{\Sigma}_n = (\widehat{B}'_{1,n} \widehat{M}_n \widehat{B}_{1,n})^{-1} \widehat{B}'_{1,n} \widehat{M}_n \widehat{C}_n \widehat{M}_n \widehat{B}_{1,n} (\widehat{B}'_{1,n} \widehat{M}_n \widehat{B}_{1,n})^{-1},$$

where $\widehat{M}_n = I_n - \widehat{B}_{2,n} (\widehat{B}'_{2,n} \widehat{B}_{2,n})^{-1} \widehat{B}'_{2,n}$.

The proof of Theorem 6.2 involves tedious but quite standard linear algebra exercises, and will therefore be given in Bierens (2011).

Thus, the estimator $\widehat{\Sigma}_n$ appears to be same as the standard variance estimator in the fixed n case, and therefore Σ is invariant for the choice of the weight

functions $\eta_k(u)$ as long as asymptotic normality is preserved. This result confirms similar conclusions by Newey (1994), Ai and Chen (2007), Ackerberg et al. (2010) and Ichimura and Lee (2010).

7. Verifying the asymptotic normality conditions for the SNP Logit model

7.1. Does the quantile penalty function matter for asymptotic normality?

It follows from (4.4) that $\nabla_k \Pi(\delta_0) = 0$, $\nabla_{k,m} \Pi(\delta_0) = 0$, and by Lemma 3.4 that there exists a constant C such that

$$\begin{aligned} \sup_{k \in \mathbb{N}} |\nabla_k \Pi(\delta)| &= \sup_{k \in \mathbb{N}} |\nabla_k \Pi(\delta) - \nabla_k \Pi(\delta_0)| \leq C \cdot \|\delta - \delta_0\|_0, \\ \sup_{k,m \in \mathbb{N}} |\nabla_{k,m} \Pi(\delta)| &= \sup_{k,m \in \mathbb{N}} |\nabla_{k,m} \Pi(\delta) - \nabla_{k,m} \Pi(\delta_0)| = C \cdot \|\delta - \delta_0\|_0. \end{aligned}$$

It is now easy to verify from Section 6 that the penalty function $\Pi(\delta)$ has neither an effect on the functions $b_m(u)$ defined in (6.19) nor on the Gaussian process Z in Lemma 6.2, and therefore has no effect on the asymptotic normality of $\hat{\theta}_n$.

That would have been different if we had chosen $\Pi(\delta) = (H(u_1|\delta) - u_1)^2 + (H(u_2|\delta) - u_2)^2$ for example, because in that case $\nabla_{k,m} \Pi(\delta_0) > 0$, so that the functions $b_m(u)$ would depend on $\nabla_{k,m} \Pi(\delta_0)$.

7.2. The SNP Logit model in least squares form

Because we may ignore the penalty function, the function $f(Z, \xi)$ in Assumption 6.1 for the SNP Logit model in least squares form is now

$$f(Z, \xi) = f(Z, (\theta, \delta)) = -(Y - H(G((1, X')\theta)|\delta))^2, \quad Z = (Y, X)'$$

where $(\theta, \delta) \in \Theta \times \Delta_\ell$, with ℓ to be determined.

Using Lemma 3.4 and the mean value theorem it can be shown that

Lemma 7.1.^(*) *Under the conditions of Theorem 4.1 we have:*

- (a) $\sup_{k \in \mathbb{N}} |\nabla_k f(Z, \xi) - \nabla_k f(Z, \xi^0)| < C \cdot (1 + \|X\|)^2 \|\xi - \xi^0\|_0$ for some constant C ;
- (b) $E[\nabla_k f(Z, \xi^0)] = 0$ for all $k \in \mathbb{N}$;

- (c) $\sup_{k \in \mathbb{N}} E[(\nabla_k f(Z, \xi^0))^2] < \infty$;
(d) $\lim_{\varepsilon \downarrow 0} \sup_{k \in \mathbb{N}} E[\sup_{\|\xi - \xi^0\|_0 \leq \varepsilon} (\nabla_k f_1(\xi) - \nabla_k f_1(\xi^0))^2] = 0$.

Consequently, Assumptions 6.3, 6.4, 6.5 and 6.9 hold, with $\ell_0 = 0$ in Assumption 6.3.

The next condition is needed in part (c) of Lemma 7.2 below.

Assumption 7.1. *The functions*

$$u(1-u) \frac{\partial}{\partial u} \left\{ \frac{\nabla_k H(u|\delta_0)}{u(1-u) \cdot h(u|\delta_0)} \right\}, \quad k \in \mathbb{N},$$

are not constant on $(0, 1)$.

The plausibility of this assumption can be verified from the expression for $h_0(u) = h(u|\delta_0)$ in (3.1).

It can now be shown that

Lemma 7.2.^(*) *Under the conditions of Theorem 4.1 we have:*

- (a) $\lim_{\varepsilon \downarrow 0} \sup_{\|\xi - \xi^0\|_1 \leq \varepsilon} \sup_{k, m \in \mathbb{N}} E[|\nabla_{k,m} f(Z, \xi) - \nabla_{k,m} f(Z, \xi^0)|] = 0$;
(b) $\sup_{k, m \in \mathbb{N}} |\nabla_{k,m} f(Z, \xi^0)| < C \cdot (1 + \|X\|)^2$ for some constant C ;
(c) *The matrix $B_{p,p}$ in Assumption 6.7 takes the form*

$$B_{p,p} = -2 \cdot E \left[(h(G((1, X')\theta_0)|\delta_0) G'((1, X')\theta_0))^2 \begin{pmatrix} 1 & X' \\ X & XX' \end{pmatrix} \right]$$

and is nonsingular.⁵ Moreover, under Assumption 7.1 the matrices $B_{p+k, p+k}$, $k \in \mathbb{N}$, are nonsingular;

- (d) $E[\nabla_{k+p, k+p} f(Z, \xi^0)] \neq 0$ for all $k \in \mathbb{N}$.

It follows now from Lemmas 7.1 and 7.2 that

Theorem 7.1. *Under the conditions of Theorem 4.1 and Assumptions 6.8 and 7.1 all the conditions of Theorems 6.1 and 6.2 are satisfied for the SNP Logit model in least squares form, with $\ell = 1$ and $\ell_0 = 0$.*

⁵Due to Assumption 2.1.

7.3. The SNP Logit model in log-likelihood form

The function $f(Z, \xi)$ for the SNP Logit model in log-likelihood form without penalty takes the form

$$\begin{aligned} f(Z, \xi) &= f(Z, (\theta, \delta)) \\ &= Y \ln(H(G((1, X')\theta)|\delta)) + (1 - Y) \ln(1 - H(G((1, X')\theta)|\delta)), \end{aligned}$$

where again $Z = (Y, X)'$ and $\xi = (\theta, \delta) \in \Theta \times \Delta_\ell$ for some $\ell \geq 1$.

Similar to Lemma 7.1 it can be shown (after some tedious derivations) that

Lemma 7.3.^(*) *Under the conditions of Theorem 4.2 we have:*

- (a) *There exist constants C and d such that $\sup_{k \in \mathbb{N}} |\nabla_k f(Z, \xi) - \nabla_k f(Z, \xi^0)| < C \cdot (1 + \|X\|)^2 \|\xi - \xi^0\|_0$ if $\|\xi - \xi^0\|_0 < d$;*
- (b) *$E[\nabla_k f(Z, \xi^0)] = 0$ for all $k \in \mathbb{N}$;*
- (c) *$\sup_{k \in \mathbb{N}} E[(\nabla_k f(Z, \xi^0))^2] < \infty$;*
- (d) *$\lim_{\varepsilon \downarrow 0} \sup_{k \in \mathbb{N}} E[\sup_{\|\xi - \xi^0\|_0 \leq \varepsilon} (\nabla_k f_1(\xi) - \nabla_k f_1(\xi^0))^2] = 0$.*

Thus again, Assumptions 6.3, 6.4, 6.5 and 6.8 hold, with $\ell_0 = 0$ in Assumption 6.3.

Next, suppose now that instead of Assumption 7.1,

Assumption 7.2. *The functions*

$$u(1-u) \frac{\partial}{\partial u} \left\{ \left(\frac{H(u|\delta_0)(1-H(u|\delta_0))}{u(1-u)} \right)^2 \frac{\nabla_k H(u|\delta_0)}{h(u|\delta_0)} \right\}, \quad k \in \mathbb{N},$$

are not constant on $(0, 1)$.

Then after tedious derivations it can be shown that

Lemma 7.4.^(*) *Under the conditions of Theorem 4.2 we have:*

- (a) *$\lim_{\varepsilon \downarrow 0} \sup_{\|\xi - \xi^0\|_1 \leq \varepsilon} \sup_{k, m \in \mathbb{N}} E[\|\nabla_{k,m} f(Z, \xi) - \nabla_{k,m} f(Z, \xi^0)\|] = 0$;*
- (b) *$\sup_{k, m \in \mathbb{N}} |\nabla_{k,m} f(Z, \xi^0)| < C \cdot (1 + \|X\|)^2$ for some constant C ;*
- (c) *The matrix $B_{p,p}$ in Assumption 6.7 takes the form*

$$B_{p,p} = -E \left[\frac{(h(G((1, X')\theta_0)|\delta_0)G'((1, X')\theta_0))^2}{H(G((1, X')\theta_0)|\delta_0)(1-H(G((1, X')\theta_0)|\delta_0))} \begin{pmatrix} 1 & X' \\ X & XX' \end{pmatrix} \right],$$

which nonsingular. Moreover, under Assumption 7.2 the matrices $B_{p+k,p+k}$, $k \in \mathbb{N}$, are nonsingular.

(d) $E[\nabla_{k+p,k+p}f(Z, \xi^0)] \neq 0$ for all $k \in \mathbb{N}$.

Thus it follows from Lemmas 7.3 and 7.4 that

Theorem 7.2. *Under the conditions of Theorem 4.2 and Assumptions 6.8 and 7.2 all the conditions of Theorems 6.1 and 6.2 are satisfied for the SNP Logit model in log-likelihood form, with $\ell = 1$ and $\ell_0 = 0$.*

8. Conclusions

In this paper I have shown that consistency of sieve estimators requires only a few mild conditions, without relying directly on a uniform law of large numbers [c.f. Lemma 5.2], and that the sieve estimators of the Euclidean parameters are asymptotically normally distributed similar to the standard finite dimensional M-estimation approach. The latter is my main contribution to the sieve estimation literature. Although asymptotic normality of sieve estimators has already been established in the literature, albeit under high-level conditions, the novelty in this paper is the alternative way asymptotic normality is proved and the weak conditions involved. Of course, the trade off is that the approach in this paper is confined to smooth SNP models for which the infinite-dimensional parameter is confined to a pre-specified compact metric space, which is less general than the standard conditions in the sieve estimation literature. See Chen (2007) for the latter.

The results in this paper are motivated and illustrated by an SNP discrete choice model. However, my results are applicable to most SNP models based on series expansions of unknown functions. For example, consider the monotone index regression model $E[Y|X] = f(\alpha + \beta'X)$, where $f(x)$ is a strictly monotonic increasing continuous function on \mathbb{R} . Then we can write $f(x) = G^{-1}(H(G(x)))$, where G is an a priori chosen distribution function on \mathbb{R} with inverse G^{-1} and $H(u)$ is a distribution function on the unit interval. This model is identified under the same conditions as the SNP discrete choice model, and can be estimated by penalized sieve least squares.

Similarly, my approach is applicable to copula models with SNP marginal distributions. See Chen et al. (2006). However, translating the conditions in the latter paper to mine is beyond the scope and size limitation of the current paper.

As mentioned in the introduction, another example of an SNP model is the mixed proportional hazard (MPH) model. In Bierens (2008) I have shown that the interval censored MPH model is identified under similar conditions as for the SNP discrete choice model and that under these conditions the sieve estimators of the Euclidean parameters and the unobserved heterogeneity distribution are consistent. Therefore, it seems that the asymptotic normality results in the current paper are applicable as well. However, an issue with the uncensored single spell MPH model is that for particular specifications of the baseline hazard its efficiency bound is singular, which implies that any consistent estimator of the Euclidean parameter vector in the MPH model involved converges at a slower rate than the square root of the sample size. See Newey (1990) for a general review of efficiency bounds, and Hahn (1994) and Ridder and Woutersen (2003) for the efficiency bound of the MPH model. On the other hand, Hahn (1994) has also shown that in general the multiple spell MPH model does not suffer from this problem, which is confirmed by the estimation results of Bierens and Carvalho (2007). Thus, Assumptions 6.7 and/or 6.8 may not hold for the single spell MPH model, but investigating this issue further is also beyond the scope and size limitation of the current paper.

Although the results in this paper are based on the i.i.d. assumption, they can be extended straightforwardly to SNP time series models. All we need is to generalize the uniform strong law of large numbers in Lemma 5.2 to the time series case, for example by assuming ergodicity, and to replace the reference to the standard central limit theorem in the proof of Lemma 6.2 in Bierens (2011) by the martingale difference central limit theorem of McLeish (1974).⁶

9. Appendix A: Proofs

9.1. Proof of Lemma 3.2

The density $h(u)$ in (3.1) can be written as $h(u) = \eta(u)^2 / \int_0^1 \eta(v)^2 dv$, where

$$\eta(u) = 1 + \sum_{m=1}^{\infty} \delta_m \sqrt{2} \cos(m\pi u) \text{ a.e. on } (0, 1). \quad (9.1)$$

⁶Because for each k the derivatives $\nabla_k f_j(\xi^0)$ in (6.1) are then martingale differences, with j the time index.

Moreover, it follows from the argument in Bierens (2008) that in general,

$$\delta_m = \frac{\int_0^1 (I(u \in B) - I(u \in [0, 1] \setminus B)) \sqrt{2} \cos(m\pi u) \sqrt{h(u)} du}{\int_0^1 (I(u \in B) - I(u \in [0, 1] \setminus B)) \sqrt{h(u)} du}, \quad (9.2)$$

$$\frac{1}{\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2}} = \int_0^1 (I(u \in B) - I(u \in [0, 1] \setminus B)) \sqrt{h(u)} du.$$

for some Borel set B satisfying $\int_0^1 (I(u \in B) - I(u \in [0, 1] \setminus B)) \sqrt{h(u)} du > 0$, hence

$$\eta(u) = (I(u \in B) - I(u \in [0, 1] \setminus B)) \sqrt{h(u)} \sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2} \quad (9.3)$$

Next, let

$$h_n(u) = \frac{(1 + \sum_{m=1}^n \delta_m \sqrt{2} \cos(m\pi u))^2}{1 + \sum_{m=1}^n \delta_m^2} = \eta_n(u)^2 / \int_0^1 \eta_n(v)^2 dv,$$

where

$$\begin{aligned} \eta_n(u) &= 1 + \sum_{m=1}^n \delta_m \sqrt{2} \cos(m\pi u) \\ &= (I(u \in B) - I(u \in [0, 1] \setminus B)) \sqrt{h_n(u)} \sqrt{1 + \sum_{m=1}^n \delta_m^2}. \end{aligned} \quad (9.4)$$

and note that by (3.1),

$$h(u) = \lim_{n \rightarrow \infty} h_n(u) \text{ a.e. on } [0, 1]. \quad (9.5)$$

Now let $S \subset [0, 1]$ be the set with Lebesgue measure zero on which (9.5) fails to hold. Then for any $u_0 \in (0, 1) \setminus S$, $\lim_{n \rightarrow \infty} h_n(u_0) = h(u_0) > 0$, hence for a sufficiently large n , $h_n(u_0) > 0$. Because obviously $h_n(u)$ and $\eta_n(u)$ are continuous on $(0, 1)$, for such an n there exists a small $\varepsilon_n(u_0) > 0$ such that $h_n(u) > 0$ for all $u \in (u_0 - \varepsilon_n(u_0), u_0 + \varepsilon_n(u_0)) \cap (0, 1)$, and therefore

$$I(u \in B) - I(u \in [0, 1] \setminus B) = \frac{\eta_n(u)}{\sqrt{h_n(u)} \sqrt{1 + \sum_{m=1}^n \delta_m^2}} \quad (9.6)$$

is continuous on $(u_0 - \varepsilon_n(u_0), u_0 + \varepsilon_n(u_0)) \cap (0, 1)$. Substituting (9.6) in (9.3) it follows now that $\eta(u)$ is continuous on $(u_0 - \varepsilon_n(u_0), u_0 + \varepsilon_n(u_0)) \cap (0, 1)$, hence by the arbitrariness of $u_0 \in (0, 1)/S$, $\eta(u)$ is continuous on $(0, 1)$.

Next, suppose that $\eta(u)$ takes positive and negative values on $(0, 1)$. Then by the continuity of $\eta(u)$ on $(0, 1)$ there exists a $u_0 \in (0, 1)$ for which $\eta(u_0) = 0$ and thus $h(u_0) = 0$, which however is excluded by the condition that $h(u) > 0$ on $(0, 1)$. Therefore, either $\eta(u) > 0$ for all $u \in (0, 1)$ or $\eta(u) < 0$ for all $u \in (0, 1)$. However, the latter is excluded because by (9.1), $\int_0^1 \eta(u) du = 1$. Thus, $\eta(u) > 0$ on $(0, 1)$, so that by (9.3), $I(u \in B) - I(u \in [0, 1] \setminus B) = 1$ on $(0, 1)$, hence by (9.2),

$$\delta_m = \frac{\int_0^1 \sqrt{2} \cos(m\pi u) \sqrt{h(u)} du}{\int_0^1 \sqrt{h(u)} du}.$$

9.2. Proof of Theorem 5.2

First, observe that $(\bigcup_{n=1}^{\infty} \Xi_n) \cap \Xi = \bigcup_{n=1}^{\infty} (\Xi_n \cap \Xi) \subset \overline{\bigcup_{n=1}^{\infty} (\Xi_n \cap \Xi)}$, hence

$$\Xi \cap \Xi = (\overline{\bigcup_{n=0}^{\infty} \Xi_n}) \cap \Xi \subset \overline{\bigcup_{n=0}^{\infty} (\Xi_n \cap \Xi)}.$$

Because $\xi^0 \in \Xi \cap \Xi$ it follows therefore that, for sufficient large n , say $n \geq \underline{n}$, each space $\Xi_n \cap \Xi$ contains an element ξ_n such that $\lim_{n \rightarrow \infty} d(\xi_n, \xi^0) = 0$, hence by the continuity of the function $\overline{Q}(\xi) = E[f(Z, \xi)]$ in ξ^0 [c.f. condition (b) in Assumption 5.4], $\lim_{n \rightarrow \infty} \overline{Q}(\xi_n) = \overline{Q}(\xi^0)$. Thus, for an arbitrary $\varepsilon > 0$ there exists an $n(\varepsilon)$ such that $\overline{Q}(\xi_{n(\varepsilon)}) > \overline{Q}(\xi^0) - \varepsilon/2$.

Recall that $\widehat{Q}_N(\widehat{\xi}_{n_N}) = \sup_{\xi \in \Xi_{n_N}} \widehat{Q}_N(\xi) \geq \widehat{Q}_N(\xi_{n(\varepsilon)})$ if $n_N \geq n(\varepsilon)$, where $\widehat{Q}_N(\xi)$ is defined by (5.1). Moreover, it follows from condition (c) of Assumption 5.2 and condition (a) of Assumption 5.4 that $E[|f(Z, \xi_{n(\varepsilon)})|] = -E[f(Z, \xi_{n(\varepsilon)})] \leq -E[f(Z, \xi)] < \infty$, so that by Kolmogorov's strong law of large numbers, $\widehat{Q}_N(\xi_{n(\varepsilon)}) \xrightarrow{\text{a.s.}} \overline{Q}(\xi_{n(\varepsilon)})$. Thus for $N \rightarrow \infty$, $\widehat{Q}_N(\widehat{\xi}_{n_N}) \geq \widehat{Q}_N(\xi_{n(\varepsilon)}) \xrightarrow{\text{a.s.}} \overline{Q}(\xi_{n(\varepsilon)}) > \overline{Q}(\xi^0) - \varepsilon/2$. Furthermore, it follows from Lemma 5.3 that $\text{plim}_{N \rightarrow \infty} (\widehat{Q}_{K_N, N}(\widehat{\xi}_{n_N}) - \overline{Q}_{K_N}(\widehat{\xi}_{n_N})) = 0$, whereas by Jensen's inequality and the convexity of the function $\max(x, -K)$, $\widehat{Q}_{K_N, N}(\widehat{\xi}_{n_N}) \geq \max(\widehat{Q}_N(\widehat{\xi}_{n_N}), -K_N) \geq \widehat{Q}_N(\widehat{\xi}_{n_N})$. Hence for $N \rightarrow \infty$,

$$\begin{aligned} \overline{Q}_{K_N}(\widehat{\xi}_{n_N}) &= \overline{Q}_{K_N}(\widehat{\xi}_{n_N}) - \widehat{Q}_{K_N, N}(\widehat{\xi}_{n_N}) + \widehat{Q}_{K_N, N}(\widehat{\xi}_{n_N}) \\ &= \widehat{Q}_{K_N, N}(\widehat{\xi}_{n_N}) + o_p(1) \geq \widehat{Q}_N(\widehat{\xi}_{n_N}) + o_p(1) \\ &\geq \overline{Q}(\xi_{n(\varepsilon)}) + o_p(1) > \overline{Q}(\xi^0) - \varepsilon/2 + o_p(1). \end{aligned}$$

Thus,

$$\lim_{N \rightarrow \infty} \Pr \left[\overline{Q}_{K_N}(\widehat{\xi}_{n_N}) > \overline{Q}(\xi^0) - \varepsilon \right] = 1. \quad (9.7)$$

Recall that convergence in probability is equivalent to a.s. convergence along a further subsequence N_k , say, of an arbitrary subsequence of N . Thus, it follows from (9.7) that there exists a set $A \in \mathcal{F}$ with $P(A) = 1$ such that for all $\omega \in A$,

$$\lim_{k \rightarrow \infty} I \left(\overline{Q}_{K_{N_k}} \left(\widehat{\xi}_{n_{N_k}}(\omega) \right) > \overline{Q}(\xi^0) - \varepsilon \right) = 1 \quad (9.8)$$

Now let $\xi_*(\omega) \in \Xi$ be a limit point of the sequence $\widehat{\xi}_{n_{N_k}}(\omega)$. Then there exists a further subsequence of N_k , say $N_{k(m)}$, possibly depending on ω , such that $\lim_{m \rightarrow \infty} d(\widehat{\xi}_{n_{N_{k(m)}}}(\omega), \xi_*(\omega)) = 0$, hence for any fixed $K > 0$, $\lim_{m \rightarrow \infty} \overline{Q}_K(\widehat{\xi}_{n_{N_{k(m)}}}(\omega)) = \overline{Q}_K(\xi_*(\omega))$, due to the continuity of $\overline{Q}_K(\xi)$. Because for $K_{N_k} \geq K$, $\overline{Q}_{K_{N_k}}(\xi) \leq \overline{Q}_K(\xi)$, it follows now from (9.8) that for arbitrary $K > 0$, $\overline{Q}_K(\xi_*(\omega)) > \overline{Q}(\xi^0) - \varepsilon$. The latter implies that $\overline{Q}(\xi_*(\omega)) > \overline{Q}(\xi^0) - \varepsilon$, because $-\max(f(z, \xi_*(\omega)), -K)$ is non-negative and non-decreasing in K , so that by monotone convergence, $\lim_{K \rightarrow \infty} \overline{Q}_K(\xi_*(\omega)) = \overline{Q}(\xi_*(\omega))$. Because $\varepsilon > 0$ was arbitrary it follows now that $\overline{Q}(\xi_*(\omega)) = \overline{Q}(\xi^0)$, which by the uniqueness of ξ^0 implies that $\xi_*(\omega) = \xi^0$. Thus, all the limit points of $\widehat{\xi}_{n_{N_k}}(\omega)$ are equal to ξ^0 , hence $\lim_{k \rightarrow \infty} d(\widehat{\xi}_{n_{N_k}}, \xi^0) = 0$ a.s., which implies $\text{plim}_{N \rightarrow \infty} d(\widehat{\xi}_{n_N}, \xi^0) = 0$.

9.3. Proof of Lemma 6.4

Note that if we choose ρ_m such that $\sum_{m=1}^{\infty} \rho_m < \infty$ then (6.16) and (6.17) hold if

$$\sum_{m=1}^n \rho_m \|\widehat{b}_{m,n} - b_m\|^2 = o_p(1), \quad (9.9)$$

which is easier to handle.

Suppose that the sequence ρ_m can be chosen such that $\sum_{m=1}^{\infty} m^{4+2\tau} \rho_m < \infty$. Then it follows from (6.24) and (6.19) that

$$\begin{aligned} & \sum_{m=1}^n \rho_m \|\widehat{b}_{m,n} - b_m\|^2 \\ &= \sum_{k=1}^n 2^{-2k} \sum_{m=1}^n \rho_m \left(\frac{1}{N} \sum_{j=1}^N \nabla_{k,m} f_j(\xi_n^0 + \lambda_k(\widehat{\xi}_n - \xi_n^0)) - E[\nabla_{k,m} f_1(\xi^0)] \right)^2 \end{aligned}$$

$$\begin{aligned}
& + \sum_{k=n+1}^n 2^{-2k} \sum_{m=1}^n \rho_m (E[\nabla_{k,m} f_1(\xi^0)])^2 \\
& \leq 2 \left(\sum_{k=1}^{\infty} k^{4+2\tau} 2^{-2k} \right) \left(\sum_{m=1}^{\infty} m^{4+2\tau} \rho_m \right) \\
& \times \left(\frac{1}{N} \sum_{j=1}^N \left(\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} k^{-2-\tau} m^{-2-\tau} \left| \nabla_{k,m} f_j(\xi_n^0 + \lambda_k(\widehat{\xi}_n - \xi_n^0)) - \nabla_{k,m} f_j(\xi^0) \right| \right) \right)^2 \\
& + 2 \left(\sum_{k=1}^{\infty} k^{4+2\tau} 2^{-2k} \right) \left(\sum_{m=1}^{\infty} m^{4+2\tau} \rho_m \right) \\
& \times \left(\frac{1}{N} \sum_{j=1}^N \left(\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} k^{-2-\tau} m^{-2-\tau} (\nabla_{k,m} f_j(\xi^0) - E[\nabla_{k,m} f_1(\xi^0)]) \right) \right)^2 \\
& + \left(\sum_{k=n+1}^{\infty} k^{4+2\tau} 2^{-2k} \right) \left(\sum_{m=1}^{\infty} m^{4+2\tau} \rho_m \right) \left(\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} k^{-2-\tau} m^{-2-\tau} E[|\nabla_{k,m} f_1(\xi^0)|] \right)^2
\end{aligned}$$

Because by Assumption 6.6(a) and Kolmogorov's strong law of large number,

$$\frac{1}{N} \sum_{j=1}^N \left(\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} k^{-2-\tau} m^{-2-\tau} (\nabla_{k,m} f_j(\xi^0) - E[\nabla_{k,m} f_1(\xi^0)]) \right) \xrightarrow{\text{a.s.}} 0$$

and $\lim_{n \rightarrow \infty} \sum_{k=n+1}^{\infty} k^{4+2\tau} 2^{-2k} = 0$, $\sum_{m=1}^{\infty} m^{4+2\tau} \rho_m < \infty$, it follows that

$$\sum_{m=1}^n \rho_m \|\widehat{b}_{m,n} - b_m\|^2 \leq 2 \left(\sum_{k=1}^{\infty} k^{4+2\tau} 2^{-2k} \right) \left(\sum_{m=1}^{\infty} m^{4+2\tau} \rho_m \right) S_N^2 + o_p(1)$$

where

$$\begin{aligned}
S_N & = \frac{1}{N} \sum_{j=1}^N \left(\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} k^{-2-\tau} m^{-2-\tau} \left| \nabla_{k,m} f_j(\xi_n^0 + \lambda_k(\widehat{\xi}_n - \xi_n^0)) - \nabla_{k,m} f_j(\xi^0) \right| \right) \\
& = S_N \cdot I \left(\|\widehat{\xi}_n - \xi^0\|_{\ell} \leq \varepsilon \right) + S_N \cdot I \left(\|\widehat{\xi}_n - \xi^0\|_{\ell} > \varepsilon \right) \\
& \leq \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} k^{-2-\tau} m^{-2-\tau} \frac{1}{N} \sum_{j=1}^N \sup_{\|\xi - \xi^0\|_{\ell} \leq \varepsilon} |\nabla_{k,m} f_j(\xi) - \nabla_{k,m} f_j(\xi^0)| + o_p(1)
\end{aligned}$$

It follows now from Assumption 6.6(b) and Chebyshev's inequality for first moments that $S_N = o_p(1)$, which proves (9.9).

Now observe that $\sum_{m=1}^{\infty} m^{4+2\tau} t^m / m! < \infty$ for arbitrary $t \in (0, 1)$, so that (9.9) holds for $\rho_m = t^m / m!$. Thus for arbitrary $t \in (0, 1)$, $\sum_{m=1}^n \frac{t^m}{m!} \|\widehat{b}_{m,n} - b_m\|^2 = o_p(1)$. The purpose of t will become clear below.

As to condition (6.18), observe from (6.6) and (6.19) that

$$\liminf_{n \rightarrow \infty} \left\| \sum_{m=p+1}^n \frac{t^m}{m!} b_m \right\|^2 = \sum_{k=1}^{\infty} 2^{-2k} \left(\sum_{m=p+1}^{\infty} \frac{t^m}{m!} E[\nabla_{k,m} f_1(\xi^0)] \right)^2$$

hence (6.18) holds if for some $t \in (0, 1)$ there exists at least one $k \in \mathbb{N}$ such that $\sum_{m=p+1}^{\infty} (t^m / m!) E[\nabla_{k,m} f_1(\xi^0)] \neq 0$. This is the case under Assumption 6.6(c). To see this, suppose that for the k in Assumption 6.6(c) and all $t \in (0, 1)$, $\sum_{s=p+1}^{\infty} (t^s / s!) E[\nabla_{k,s} f_1(\xi^0)] = 0$. Then for all $t \in (0, 1)$,

$$0 = \frac{d^{p+m}}{(dt)^{p+m}} \sum_{s=p+1}^{\infty} \frac{t^s}{s!} E[\nabla_{k,s} f_1(\xi^0)] = \sum_{s=0}^{\infty} \frac{t^s}{s!} E[\nabla_{k,s+p+m} f_1(\xi^0)],$$

hence by the continuity of the latter, letting $t \downarrow 0$ it follows that $E[\nabla_{k,p+m} f_1(\xi^0)] = 0$, which contradicts Assumption 6.6(c).

9.4. Proof of Lemma 6.5

For $n > p$ and $k \geq n$, partition the matrix $B_{k,n}$ in Assumption 6.8 as $B_{k,n} = (B_{1,k,n}, B_{2,k,n})$ where $B_{1,k,n}$ is the matrix of the first p columns of $B_{k,n}$. Denote

$$\varphi_k(u) = (\eta_1(u), \dots, \eta_k(u))', \quad \Phi_k = \int_0^1 \varphi_k(u) \varphi_k(u)' du,$$

and recall that $\eta_k(u) = 2^{-k} \sqrt{2} \cos(k\pi u)$, so that $\Phi_k = \text{diag}(2^{-2}, 2^{-4}, \dots, 2^{-2k})$. Moreover, denote

$$b_{k,m}(u) = -b'_{k,m} \varphi_k(u), \quad m = 1, 2, \dots, n,$$

where $b_{k,m}$ is column m of $B_{k,n}$, and observe from (6.19) that

$$b_m(u) - b_{k,m}(u) = \sum_{m=k+1}^{\infty} 2^{-m} E[\nabla_{m,k} f_1(\xi^0)] \sqrt{2} \cos(m\pi u)$$

so that by Assumption 6.6(a),

$$\lim_{k \rightarrow \infty} \sup_{0 \leq u \leq 1} |b_m(u) - b_{k,m}(u)| = 0.$$

Next denote

$$b_k^{(1)}(u) = (b_{k,1}(u), \dots, b_{k,p}(u))', \quad b_k^{(2)}(u) = (b_{k,p+1}(u), \dots, b_{k,n}(u))'.$$

Then

$$b_k^{(1)}(u) = -B'_{1,k,n} \varphi_k(u), \quad b_k^{(2)}(u) = -B'_{2,k,n} \varphi_k(u)$$

The residual $a_{k,n}(u)$ of the projection of $b_n^{(1)}(u)$ on $\text{span}(b_n^{(2)}(u))$ takes the form

$$a_{k,n}(u) = b_k^{(1)}(u) - \Omega_n b_k^{(2)}(u) = -B'_{1,k,n} \varphi_k(u) + \Omega_n B'_{2,k,n} \varphi_k(u)$$

where

$$\Omega_n = B'_{1,k,n} \Phi_k B_{2,k,n} (B'_{2,k,n} \Phi_k B_{2,k,n})^{-1} \quad (9.10)$$

Hence

$$\begin{aligned} & \int_0^1 a_{k,n}(u) a_{k,n}(u)' du \\ &= B'_{1,k,n} \Phi_k B_{1,k,n} - B'_{1,k,n} \Phi_k B_{2,k,n} (B'_{2,k,n} \Phi_k B_{2,k,n})^{-1} B'_{2,k,n} \Phi_k B_{1,k,n} \end{aligned} \quad (9.11)$$

which is non-singular because

$$B'_{k,n} \Phi_k B_{k,n} = \begin{pmatrix} B'_{1,k,n} \Phi_k B_{1,k,n} & B'_{1,k,n} \Phi_k B_{2,k,n} \\ B'_{2,k,n} \Phi_k B_{1,k,n} & B'_{2,k,n} \Phi_k B_{2,k,n} \end{pmatrix} \quad (9.12)$$

is nonsingular, and the upper-left block of the inverse of (9.12) is just the inverse of the matrix (9.11).

Finally, note that

$$\lim_{k \rightarrow \infty} B'_{k,n} \Phi_k B_{k,n} = \begin{pmatrix} \lim_{k \rightarrow \infty} B'_{1,k,n} \Phi_k B_{1,k,n} & \lim_{k \rightarrow \infty} B'_{1,k,n} \Phi_k B_{2,k,n} \\ \lim_{k \rightarrow \infty} B'_{2,k,n} \Phi_k B_{1,k,n} & \lim_{k \rightarrow \infty} B'_{2,k,n} \Phi_k B_{2,k,n} \end{pmatrix}$$

is nonsingular, and therefore so is

$$\begin{aligned} & \int_0^1 a_n(u) a_n(u)' du = \lim_{k \rightarrow \infty} \int_0^1 a_{k,n}(u) a_{k,n}(u)' du \\ &= \lim_{k \rightarrow \infty} B'_{1,k,n} \Phi_k B_{1,k,n} \\ &- \left(\lim_{k \rightarrow \infty} B'_{1,k,n} \Phi_k B_{2,k,n} \right) \left(\lim_{k \rightarrow \infty} B'_{2,k,n} \Phi_k B_{2,k,n} \right)^{-1} \left(\lim_{k \rightarrow \infty} B'_{2,k,n} \Phi_k B_{1,k,n} \right). \end{aligned}$$

10. Appendix B: Projections on a Hilbert space spanned by random elements

Theorem B.1. *Let Y_N and $X_{1,N}, X_{2,N}, \dots, X_{n,N}$ be random elements of a Hilbert space \mathcal{H} on the basis on a sample of size N , where n is a subsequence of N . Let $\widehat{Y}_{n,N}$ be the projection of Y_N on $\text{span}(\{X_{m,N}\}_{m=1}^n)$, with residual $U_{n,N} = Y_N - \widehat{Y}_{n,N}$. Suppose that the following conditions hold.*

(a) *There exists a non-random element y of \mathcal{H} such that*

$$\text{plim}_{N \rightarrow \infty} \|Y_N - y\| = 0. \quad (10.1)$$

(b) *There exist a sequence $\{x_m\}_{m=1}^\infty$ of non-random elements of \mathcal{H} and a sequence $\{\rho_m\}_{m=1}^\infty$ of positive numbers such that*

$$\text{plim}_{N \rightarrow \infty} \sum_{m=1}^n \rho_m \|X_{m,N} - x_m\| = 0 \quad (10.2)$$

and

$$\liminf_{n \rightarrow \infty} \left\| \sum_{m=1}^n \rho_m x_m \right\| > 0. \quad (10.3)$$

Then $\text{plim}_{N \rightarrow \infty} \|\widehat{Y}_{n,N} - \widehat{y}\| = 0$ and $\text{plim}_{N \rightarrow \infty} \|U_{n,N} - u\| = 0$, where \widehat{y} is the projection of y on $\text{span}(\{x_m\}_{m=1}^\infty)$ and $u = y - \widehat{y}$ is the residual involved.

Proof: Note that $\|\widehat{Y}_{n,N} - \widehat{y}\| = \|(Y_N - U_{n,N}) - (y - u)\| = \|(u - U_{n,N}) - (Y_N - y)\| \leq \|U_{n,N} - u\| + \|Y_N - y\|$, hence by condition (10.1), $\|\widehat{Y}_{n,N} - \widehat{y}\| \leq \|U_{n,N} - u\| + o_p(1)$. Therefore it suffices to prove $\|U_{n,N} - u\| = o_p(1)$, as follows.

Let $\widetilde{Y}_{n,N}$ be the projection of y on $\text{span}(\{X_{m,N}\}_{m=1}^n)$, with residual $\widetilde{U}_{n,N} = y - \widetilde{Y}_{n,N}$, and let \widehat{y}_n be the projection of y on $\text{span}(\{x_m\}_{m=1}^n)$, with residual $u_n = y - \widehat{y}_n$. Then by the triangular inequality, $\|U_{n,N} - u\| \leq \|U_{n,N} - \widetilde{U}_{n,N}\| + \|u_n - \widetilde{U}_{n,N}\|$. It will be shown that

$$\|U_{n,N} - \widetilde{U}_{n,N}\| = o_p(1) \quad (10.4)$$

and

$$\|\widehat{y}_n - \widetilde{U}_{n,N}\| = \|u_n - \widetilde{U}_{n,N}\| = o_p(1). \quad (10.5)$$

Because $\lim_{n \rightarrow \infty} \|\widehat{y}_n - \widehat{y}\| = 0$ and thus $\lim_{n \rightarrow \infty} \|u_n - u\| = 0$, the result of the theorem under review then follows from (10.4) and (10.5).

Proof of (10.4)

Denote the angle between two elements x and y of \mathcal{H} by $\varphi(x, y)$. Recall that $\cos(\varphi(x, y)) = \langle x, y \rangle / (\|x\| \cdot \|y\|)$, which implies that

$$\begin{aligned}\sin^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) &= \|U_{n,N}\|^2 / \|Y_N\|^2, \quad \sin^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) = \|\widetilde{U}_{n,N}\|^2 / \|y\|^2, \\ \cos\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) &= \|\widehat{Y}_{n,N}\| / \|Y_N\|, \quad \cos\left(\varphi(y, \widetilde{Y}_{n,N})\right) = \|\widetilde{Y}_{n,N}\| / \|y\|.\end{aligned}$$

Using these formulas we can write

$$\begin{aligned}\|U_{n,N} - \widetilde{U}_{n,N}\|^2 &= \|U_{n,N}\|^2 + \|\widetilde{U}_{n,N}\|^2 - 2\langle U_{n,N}, \widetilde{U}_{n,N} \rangle \\ &= \|Y_N\|^2 \sin^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) + \|y\|^2 \sin^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) - 2\langle U_{n,N}, \widetilde{U}_{n,N} \rangle \\ &= \|Y_N\|^2 + \|y\|^2 - \|Y_N\|^2 \cos^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) - \|y\|^2 \cos^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) \\ &\quad - 2\langle U_{n,N}, \widetilde{U}_{n,N} \rangle \\ &= \|Y_N - y\|^2 - \|Y_N\|^2 \cos^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) - \|y\|^2 \cos^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) \\ &\quad + 2\langle Y_N, y \rangle - 2\langle U_{n,N}, \widetilde{U}_{n,N} \rangle\end{aligned}$$

and

$$\begin{aligned}\langle U_{n,N}, \widetilde{U}_{n,N} \rangle &= \langle U_{n,N}, \widetilde{U}_{n,N} + \widetilde{Y}_{n,N} \rangle = \langle U_{n,N}, y \rangle \\ &= \langle U_{n,N} + \widehat{Y}_{n,N}, y \rangle - \langle \widehat{Y}_{n,N}, y \rangle = \langle Y_N, y \rangle - \langle \widehat{Y}_{n,N}, y \rangle \\ &= \langle Y_N, y \rangle - \cos\left(\varphi(y, \widehat{Y}_{n,N})\right) \|\widehat{Y}_{n,N}\| \cdot \|y\| \\ &= \langle Y_N, y \rangle - \cos\left(\varphi(y, \widehat{Y}_{n,N})\right) \cos\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) \|y\| \cdot \|Y_N\| \\ &\geq \langle Y_N, y \rangle - \cos\left(\varphi(y, \widetilde{Y}_{n,N})\right) \cos\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) \|y\| \cdot \|Y_N\|\end{aligned}$$

where the inequality follows from $\cos\left(\varphi(y, \widehat{Y}_{n,N})\right) \leq \cos\left(\varphi(y, \widetilde{Y}_{n,N})\right)$. Thus

$$\begin{aligned}\|U_{n,N} - \widetilde{U}_{n,N}\|^2 &\leq \|Y_N - y\|^2 \\ &\quad - \left(\|Y_N\| \cos\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) - \|y\| \cos\left(\varphi(y, \widetilde{Y}_{n,N})\right) \right)^2 \\ &\leq \|Y_N - y\|^2 = o_p(1)\end{aligned}$$

where the $o_p(1)$ term is due to condition (10.1). This proves (10.4).

Proof of (10.5)

Let $r_1 = x_1$ and for $m \geq 2$, let r_m be the residual of the projection of x_m on $\text{span}(x_1, \dots, x_{m-1})$. Denote $e_m = \|r_m\|^{-1}r_m$ if $\|r_m\| > 0$ and $e_m = 0$ if $\|r_m\| = 0$. Similarly, let $R_{1,N} = X_{1,N}$ and for $m = 2, \dots, n$, let $R_{m,N}$ be the residual of the projection of $X_{m,N}$ on $\text{span}(X_{1,N}, \dots, X_{m-1,N})$. Note that by condition (10.2),

$$\|R_{m,N} - r_m\| = o_p(1). \quad (10.6)$$

Denote $\hat{e}_{m,N} = \|R_{m,N}\|^{-1}R_{m,N}$ if $\|R_{m,N}\| > 0$, and $\hat{e}_{m,N} = 0$ if $\|R_{m,N}\| = 0$. Then we can write $\hat{y}_n = \sum_{m=1}^n \alpha_m e_m$, where $\alpha_m = \langle y, e_m \rangle$ and $\sum_{m=1}^{\infty} \alpha_m^2 < \infty$, and $\tilde{Y}_{n,N} = \sum_{m=1}^n \hat{\alpha}_{m,N} \hat{e}_{m,N}$, where $\hat{\alpha}_{m,N} = \langle y, \hat{e}_{m,N} \rangle$.

It follows from the trivial equalities $\|\hat{y}_n - \tilde{Y}_{n,N}\|^2 = \|\tilde{Y}_{n,N}\|^2 + \|\hat{y}_n\|^2 - 2\langle \hat{y}_n, \tilde{Y}_{n,N} \rangle$ and $\langle \hat{y}_n, \tilde{Y}_{n,N} \rangle = \langle \hat{y}_n, y - \tilde{U}_{n,N} \rangle = \|\hat{y}_n\|^2 - \langle \hat{y}_n, \tilde{U}_{n,N} \rangle$ that $\|\hat{y}_n - \tilde{Y}_{n,N}\|^2 = \|\tilde{Y}_{n,N}\|^2 - \|\hat{y}_n\|^2 + 2\langle \hat{y}_n, \tilde{U}_{n,N} \rangle$. Moreover, using the Cauchy-Schwarz inequality and the fact that $\|\tilde{U}_{n,N}\| \leq \|y\|$, it follows that

$$\begin{aligned} \left| \langle \hat{y}_n, \tilde{U}_{n,N} \rangle \right| &= \left| \left\langle \sum_{m=1}^n \alpha_m e_m, \tilde{U}_{n,N} \right\rangle \right| = \left| \left\langle \sum_{m=1}^n \alpha_m I(\|e_m\| > 0) e_m, \tilde{U}_{n,N} \right\rangle \right| \\ &= \left| \left\langle \sum_{m=1}^n \alpha_m I(\|e_m\| > 0) (e_m - \hat{e}_{m,N}), \tilde{U}_{n,N} \right\rangle \right| \\ &\leq \|\tilde{U}_{n,N}\| \cdot \left\| \sum_{m=1}^n \alpha_m I(\|e_m\| > 0) (e_m - \hat{e}_{m,N}) \right\| \\ &\leq \|y\| \cdot \left\| \sum_{m=1}^n \alpha_m I(\|e_m\| > 0) (e_m - \hat{e}_{m,N}) \right\| \\ &\leq \|y\| \cdot \left\| \sum_{m=1}^k \alpha_m I(\|e_m\| > 0) (e_m - \hat{e}_{m,N}) \right\| + 2\|y\| \sqrt{\sum_{m=k+1}^{\infty} \alpha_m^2} \end{aligned}$$

Given an arbitrary $\varepsilon > 0$ we can choose k so large that $2\|y\| \sqrt{\sum_{m=k+1}^{\infty} \alpha_m^2} < \varepsilon$. Moreover, note that by (10.6), $(e_m - \hat{e}_{m,N})I(\|e_m\| > 0) = (\|r_m\|^{-1}r_m - \|R_{m,N}\|^{-1}R_{m,N})I(\|r_m\| > 0) = o_p(1)$, hence for $m \leq k$, $\|\sum_{m=1}^k \alpha_m I(\|e_m\| > 0) (e_m - \hat{e}_{m,N})\| = o_p(1)$. Consequently, $\langle \hat{y}_n, \tilde{U}_{n,N} \rangle = o_p(1)$ and thus

$$\|\hat{y}_n - \tilde{Y}_{n,N}\|^2 = \|\tilde{Y}_{n,N}\|^2 - \|\hat{y}_n\|^2 + o_p(1). \quad (10.7)$$

The next step is to show that

$$\|\tilde{Y}_{n,N}\| \leq \|\hat{y}_n\| + o_p(1), \quad (10.8)$$

as follows. Note that

$$\begin{aligned} \|\tilde{U}_{n,N}\|^2 &= \inf_{\beta_1, \dots, \beta_n} \left\| y - \sum_{m=1}^n \beta_m X_{m,N} \right\|^2 \\ &= \inf_{(\xi_1, \dots, \xi_n)' \in \mathcal{X}_{m=1}^n[-\rho_m, \rho_m]} \inf_{\lambda} \left\| y - \lambda \sum_{m=1}^n \xi_m X_{m,N} \right\|^2 \\ &= \inf_{(\xi_1, \dots, \xi_n)' \in \mathcal{X}_{m=1}^n[-\rho_m, \rho_m]} \left\{ \|y\|^2 - \left(\frac{\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle}{\left\| \sum_{m=1}^n \xi_m X_{m,N} \right\|} \right)^2 \right\} \\ &= \|y\|^2 - \sup_{(\xi_1, \dots, \xi_n)' \in \mathcal{X}_{m=1}^n[-\rho_m, \rho_m]} \frac{(\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle)^2}{\left\| \sum_{m=1}^n \xi_m X_{m,N} \right\|^2} \end{aligned}$$

and $\|y\|^2 = \|\tilde{Y}_{n,N} + \tilde{U}_{n,N}\|^2 = \|\tilde{Y}_{n,N}\|^2 + \|\tilde{U}_{n,N}\|^2$, so that

$$\|\tilde{Y}_{n,N}\|^2 = \sup_{(\xi_1, \dots, \xi_n)' \in \mathcal{X}_{m=1}^n[-\rho_m, \rho_m]} \frac{(\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle)^2}{\left\| \sum_{m=1}^n \xi_m X_{m,N} \right\|^2} \quad (10.9)$$

Because without loss of generality we may assume that for the optimal ξ_m 's, $\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle \geq 0$, it follows from (10.9) that

$$\|\tilde{Y}_{n,N}\| = \sup_{(\xi_1, \dots, \xi_n)' \in \mathcal{X}_{m=1}^n[-\rho_m, \rho_m]} \frac{\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle}{\left\| \sum_{m=1}^n \xi_m X_{m,N} \right\|}. \quad (10.10)$$

and similarly,

$$\|\hat{y}_n\| = \sup_{(\xi_1, \dots, \xi_n)' \in \mathcal{X}_{m=1}^n[-\rho_m, \rho_m]} \frac{\langle y, \sum_{m=1}^n \xi_m x_m \rangle}{\left\| \sum_{m=1}^n \xi_m x_m \right\|} \quad (10.11)$$

Note that by condition (10.3) at least one x_m is non-zero, so that (10.10) and (10.11) are well-defined for sufficiently large n .

The ratios in (10.10) and (10.11) are scale-invariant. Therefore, without loss of generality we may impose the normalization

$$\left\| \sum_{m=1}^n \xi_m x_m \right\| = M_n = \frac{1}{2} \left\| \sum_{m=1}^n \rho_m x_m \right\|, \quad (10.12)$$

for example. Note that (10.12) is compatible with $(\xi_1, \dots, \xi_n)' \in \mathbf{X}_{m=1}^n[-\rho_m, \rho_m]$. Thus, denoting

$$\Xi_n = \left\{ (\xi_1, \dots, \xi_n)' \in \mathbf{X}_{m=1}^n[-\rho_m, \rho_m] : \left\| \sum_{m=1}^n \xi_m x_m \right\| = M_n \right\},$$

the expressions (10.10) and (10.11) are equivalent to

$$\|\tilde{Y}_{n,N}\| = \sup_{(\xi_1, \dots, \xi_n)' \in \Xi_n} \frac{\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle}{\left\| \sum_{m=1}^n \xi_m X_{m,N} \right\|} \quad (10.13)$$

and

$$\|\hat{y}_n\| = \sup_{(\xi_1, \dots, \xi_n)' \in \Xi_n} \frac{\langle y, \sum_{m=1}^n \xi_m x_m \rangle}{\left\| \sum_{m=1}^n \xi_m x_m \right\|}, \quad (10.14)$$

respectively. Moreover, note that for $(\xi_1, \dots, \xi_n)' \in \Xi_n$,

$$\begin{aligned} \left\| \sum_{m=1}^n \xi_m X_{m,N} \right\|^2 &= \left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) + \sum_{m=1}^n \xi_m x_m \right\|^2 \\ &= \left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \right\|^2 + \left\| \sum_{m=1}^n \xi_m x_m \right\|^2 \\ &\quad + 2 \left\langle \sum_{m=1}^n \xi_m (X_{m,N} - x_m), \sum_{m=1}^n \xi_m x_m \right\rangle \\ &\geq \left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \right\|^2 + \left\| \sum_{m=1}^n \xi_m x_m \right\|^2 \\ &\quad - 2 \left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \right\| \cdot \left\| \sum_{m=1}^n \xi_m x_m \right\| \\ &= \left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \right\|^2 + M_n^2 - 2M_n \left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \right\| \\ &= \left(\left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \right\| - M_n \right)^2, \end{aligned}$$

hence

$$\frac{\left\| \sum_{m=1}^n \xi_m X_{m,N} \right\|^2}{M_n^2} \geq \left(\frac{\left\| \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \right\|}{M_n} - 1 \right)^2$$

Because $\|\sum_{m=1}^n \xi_m(X_{m,N} - x_m)\| \leq \sum_{m=1}^n \rho_m \|X_{m,N} - x_m\| = o_p(1)$, where the latter follows from (10.2), and by (10.3), $\liminf_{n \rightarrow \infty} M_n > 0$, it follows now that with probability converging to 1,

$$\frac{\|\sum_{m=1}^n \xi_m X_{m,N}\|}{M_n} \geq 1 - \frac{\sum_{m=1}^n \rho_m \|X_{m,N} - x_m\|}{M_n}. \quad (10.15)$$

Finally, observe from (10.12), (10.13), (10.14) and (10.15) that with probability converging to 1,

$$\begin{aligned} \|\widehat{y}_n\| &= \sup_{(\xi_1, \dots, \xi_n)' \in \Xi_n} \left\{ \frac{\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle}{\|\sum_{m=1}^n \xi_m X_{m,N}\|} \times \frac{\|\sum_{m=1}^n \xi_m X_{m,N}\|}{\|\sum_{m=1}^n \xi_m x_m\|} \right. \\ &\quad \left. - \frac{\langle y, \sum_{m=1}^n \xi_m (X_{m,N} - x_m) \rangle}{\|\sum_{m=1}^n \xi_m x_m\|} \right\} \\ &\geq \sup_{(\xi_1, \dots, \xi_n)' \in \Xi_n} \left\{ \frac{\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle}{\|\sum_{m=1}^n \xi_m X_{m,N}\|} \times \frac{\|\sum_{m=1}^n \xi_m X_{m,N}\|}{M_n} \right\} \\ &\quad - \|y\| \frac{\sum_{m=1}^n \rho_m \|X_{m,N} - x_m\|}{M_n} \\ &\geq \left(1 - \frac{\sum_{m=1}^n \rho_m \|X_{m,N} - x_m\|}{M_n} \right) \sup_{(\xi_1, \dots, \xi_n)' \in \Xi_n} \frac{\langle y, \sum_{m=1}^n \xi_m X_{m,N} \rangle}{\|\sum_{m=1}^n \xi_m X_{m,N}\|} \\ &\quad - \|y\| \frac{\sum_{m=1}^n \rho_m \|X_{m,N} - x_m\|}{M_n} \\ &= \left(1 - \frac{\sum_{m=1}^n \rho_m \|X_{m,N} - x_m\|}{M_n} \right) \|\widetilde{Y}_{n,N}\| - \|y\| \frac{\sum_{m=1}^n \rho_m \|X_{m,N} - x_m\|}{M_n} \\ &\geq \|\widetilde{Y}_{n,N}\| - 2\|y\| \frac{\sum_{m=1}^n \rho_m \|X_{m,N} - x_m\|}{M_n}, \end{aligned} \quad (10.16)$$

where the last inequality follows from $\|\widetilde{Y}_{n,N}\| \leq \|y\|$. Because by condition (10.3), $\liminf_{n \rightarrow \infty} M_n > 0$, it follows from (10.2) and (10.16) that (10.8) holds. The latter together with (10.7) imply (10.5).

References

Ackerberg, D., X. Chen & J. Hahn (2010), "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators", forthcoming in the *Review of Economics and Statistics*.

Adams, R.A. & J.J.F. Fournier (2003), *Sobolev Spaces*. Academic Press.

- Ai, C. & X. Chen (2007), "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables", *Journal of Econometrics* 141, 5-43.
- Andrews, D.W.K. (1994), "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity", *Econometrica* 62, 43-72.
- Bickel, P.J., C.A.J. Klaassen, Y. Ritov & J.A. Wellner (1998), *Efficient and Adaptive Estimation for Semiparametric Models*, Springer.
- Bierens, H.J. (2004), *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press.
- Bierens, H.J. (2008), "Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results", *Econometric Theory* 24, 749-794.
- Bierens, H.J. (2011), "Separate Appendix to: Consistency and Asymptotic Normality of Sieve Estimators Under Weak and Verifiable Conditions". http://econ.la.psu.edu/~hbierens/SNPMODELS_A.PDF
- Bierens, H.J. & J. R. Carvalho (2007), "Semi-Nonparametric Competing Risks Analysis of Recidivism", *Journal of Applied Econometrics* 22, 971-993.
- Billingsley, P. (1968), *Convergence of Probability Measures*. John Wiley.
- Chen, X. (2007), "Large Sample Sieve Estimation of Semi-Nonparametric Models". In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics, Vol. 6*, Ch. 76. Elsevier.
- Chen, X. (2011), "Penalized Sieve Estimation and Inference of Semi-Nonparametric Dynamic Models: A Selective Review", Cowles Foundation discussion paper no. 2804, Yale University.
- Chen, X. and X. Shen (1998), "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica* 66, 289-314.
- Chen, X., O. Linton & I. Van Keilegom (2003), "Estimation of Semiparametric Models when the Criterion Function is Not Smooth", *Econometrica*, 71, 1591-1608.
- Chen, X., Y. Fan & V. Tsyrennikov (2006), "Efficient Estimation of Semiparametric Multivariate Copula Models", *Journal of the American Statistical Association*, 101, 1228-1240.
- Eastwood, B.J. & A.R. Gallant (1991), "Adaptive Rules for Semi-Nonparametric Estimators that Achieve Asymptotic Normality", *Econometric Theory* 7, 307-340.
- Elbers, C. & G. Ridder (1982), "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model", *Review of Economic Studies* 49, 403-409.

Gabler, S, F. Laisney & M. Lechner (1993), "Seminonparametric Estimation of Binary-Choice Models with an Application to Labor-Force Participation", *Journal of Business & Economic Statistics* 11, 61-80.

Gallant, A. R. (1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form", *Journal of Econometrics* 15, 211-245.

Gallant, A.R. & D.W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation", *Econometrica* 55, 363-390.

Geman, S. & C. Hwang (1982), "Nonparametric Maximum Likelihood Estimation by the Method of Sieves", *Annals of Statistics* 10, 401-414

Gill, R.D. (1989), "Non- and Semi-Parametric Maximum Likelihood Estimators and the Von Mises Method (Part 1)", *Scandinavian Journal of Statistics* 16, 97-128.

Grenander, U. (1981), *Abstract Inference*. Wiley.

Hahn, J. (1994), "The Efficiency Bound of the Mixed Proportional Hazard Model", *Review of Economic Studies* 61, 607-629.

Hannan, E.J., & B.G. Quinn (1979), "The Determination of the Order of an Autoregression", *Journal of the Royal Statistical Society, Series B* 41, 190-195.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161.

Heckman, J. J. & B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 52, 271-320.

Horowitz, J.L. (1998), *Semiparametric Methods in Econometrics*, Springer.

Ichimura, H. & S. Lee (2010), "Characterization of the Asymptotic Distribution of Semiparametric M-Estimators", forthcoming in the *Journal of Econometrics*.

Jennrich, R.I. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics* 40, 633-643.

Kronmal, R. & M. Tarter (1968), "The Estimation of Densities and Cumulatives by Fourier Series Methods", *Journal of the American Statistical Association* 63, 925-952.

Lancaster, T. (1979), "Econometric Methods for the Duration of Unemployment", *Econometrica* 47, 939-956.

Manski, C.F. (1988), "Identification of Binary Response Models", *Journal of the American Statistical Association* 83, 729-738.

McLeish, D. L. (1974), "Dependent Central Limit Theorems and Invariance

Principles", *Annals of Probability* 2, 620-628.

Newey, W.K. (1990), "Semiparametric Efficiency Bounds", *Journal of Applied Econometrics* 5, 99-135.

Newey, W.K. (1994), "The Asymptotic Variance of Semiparametric Estimators", *Econometrica* 62, 1349-1382.

Newey, W.K. (1997), "Convergence Rates and Asymptotic Normality for Series Estimators", *Journal of Econometrics* 79, 147-168.

Ridder, G. & T. Woutersen (2003), "The Singularity of the Efficiency Bound of the Mixed Proportional Hazard Model", *Econometrica* 71, 1579-1589.

Schwarz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics* 6, 461-464.

Shen, X. (1997), "On the Method of Sieves and Penalization", *Annals of Statistics* 25, 2555-2591.

Van der Vaart, A.W. (1998), *Asymptotic Statistics*. Cambridge University Press.

White, H. & J. Wooldridge (1991), "Some Results on Sieve Estimation with Dependent Observations". In Barnett, W.A., J. Powell & G. Tauchen (eds.), *Non-parametric and Semi-parametric Methods in Econometrics and Statistics*, Ch. 18, Cambridge University Press.