A Unified Asymptotic Distribution Theory for Parametric and Non-Parametric Least Squares

Bruce E. Hansen*
University of Wisconsin†

November 2015
Preliminary

Abstract

This paper presents simple and general conditions for asymptotic normality of least squares estimators allowing for regressors vectors which expand with the sample size. Our assumptions include series and sieve estimation of regression functions, and any context where the regressor set increases with the sample size. The conditions are quite general, including as special cases the assumptions commonly used for both parametric and nonparametric sieve least squares. Our assumptions allow the regressors to be unbounded, and do not bound the conditional variances. Our assumptions allow the number of regressors K to be either fixed or increasing with sample size. Our conditions bound the allowable rate of growth of K as a function of the number of finite moments, showing that there is an inherent trade-off between the number of moments and the allowable number of regressors.

*Research supported by the National Science Foundation. I thank Victor Chernozhukov and Xiaohong Chen for helpful and detailed comments.

†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706.

1 Introduction

This paper presents simple and general conditions for asymptotic normality of least squares estimators allowing for regressors vectors which expand with the sample size. Our assumptions include series and sieve estimation of regression functions, and any context where the regressor set increases with the sample size.

We focus on asymptotic normality of studentized linear functions of the coefficient vector and regression function.

Our results are general and unified, in the sense that they include as special cases the conditions for asymptotic normality obtained in much of the previous literature. Some of the important features include the following. We allow the number of regressors K to be fixed or changing as a function of sample size, thus including parametric and nonparametric regression as special cases. We allow the regressors to be unbounded, which is a typical assumption in parametric regression but this is its first appearance in nonparmetric sieve regression. We allow for conditional heteroskedasticity, and do not require the conditional error variance to be bounded above zero nor below infinity. Again this is commonplace in parametric regression but new in nonparametric sieve theory.

We present two theorems concerning asymptotic normality. The first demonstrates the asymptotic normality of studentized linear functions of the coefficient vector. These are centered at the linear projection coefficients and thus represent inference on the latter. This result is similar to standard results for parametric regression.

Our second theorem demonstrates asymptotic normality of studentized linear functions of the regression function. These estimates have a finite sample bias due to the finite-K approximation error. Our asymptotic theory is explicit concerning this bias instead of assuming its negligability due to undersmoothing. We believe this to be a stronger distribution theory than one which assumes away the bias via undersmoothing.

Ours are the first results for the asymptotic normality of sieve regression which use moment conditions rather than boundedness. Bounded regressors and bounded variances appear as a special limiting case. Our results show that there is an effective trade-off between the number of finite moments and the allowable rate of expansion of the number of series terms. The previous literature which imposed boundedness has missed this trade-off.

This paper builds on an extensive literature developing an asymptotic distribution theory for series regression. Important contributions include Andrews (1991), Newey (1997), Chen and Shen (1998), Huang (2003), Chen, et. al. (2014), Chen and Liao (2014), Belloni et. al. (2014), and Chen and Christensen (2015). All of these papers assume bounded regressors and bounded conditional variances, with the interesting exception of Chen and Christensen (2015) who allow unbounded regressors but examine a trimmed least-squares estimator (which is effectively regression on bounded regressors), and Chen and Shen (1998) whose results are confined to root-n estimable functions. Chen and Shen (1998), Chen et. al. (2014) and Chen and Christensen (2015) allow for times series observations (while this paper only concerns iid data), and Belloni et. al. (2014) also consider uniform asymptotic approximations (while the results in this paper are only pointwise).

A word on notation. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalue of a positive semidefinite matrix A. Let $\|A\| = (\lambda_{\max}(A'A))^{1/2}$ denote the spectral norm of a matrix A. Note that if $A \geq 0$ then $\|A\| = \lambda_{\max}(A)$. When applied to a $K \times 1$ vector let $\|a\| = (a'a)^{1/2}$ denote the Euclidean norm. For $p \geq 1$, let $\|z\|_p = (\mathbb{E}\|z\|^p)^{1/p}$ denote the L^p norm of a random variable, vector, or matrix.

2 Least Squares Regression

Let (y_i, z_i), $i = 1, ..., n$ be a sample of iid observations with $y_i \in \mathbb{R}$. Define the conditional mean $g(z) = \mathbb{E}(y_i \mid z_i = z)$, the regression error $e_i = y_i - g(z_i)$, the conditional variance $\sigma^2(z) = \mathbb{E}\left(e_i^2 \mid z_i = z\right)$, and its realization $\sigma_i^2 = \sigma^2(z_i)$.

Consider the estimation of $g(z)$ by approximate linear regression. For $K = K(n)$ let $x_K(z)$ be a set of $K \times 1$ transformations of the regressor z. This can include a subset of observed variables z, transformations of z including basis function transformations, or combinations. For example, when $z \in \mathbb{R}$ a power series approximation sets $x_K(z) = \left(1, z, ..., z^{K-1}\right)$. Construct the regressors $x_{Ki} = x_K(z_i)$.

We approximate the conditional mean $g(z)$ by a linear function $x_K'(z)\beta_K$ for some $K \times 1$ coefficient vector β_K. We can write this approximating model as

$$y_i = x_{Ki}'\beta_K + e_{Ki}. \tag{1}$$

The projection approximation defines the coefficient by linear projection

$$\beta_K = \left(\mathbb{E}\left(x_{Ki}x_{Ki}'\right)\right)^{-1}\mathbb{E}\left(x_{Ki}y_i\right). \tag{2}$$

This has the properties that $x_K'(z)\beta_K$ is the best linear approximation (in L^2) to $g(z)$, and that $\mathbb{E}(x_{Ki}e_{Ki}) = 0$.

A vector-valued parameter of interest

$$\theta = a(g) \in \mathbb{R}^d \tag{3}$$

may be the regression function $g(z)$ at a fixed point z or some other linear function of g including derivatives and integrals over g. Linearity inplies that if we plug in the series approximation $x_K(z)'\beta_K$ into (3), then we obtain the approximating (or pseudo-true) parameter value

$$\theta_K = a(g_K) = a_K'\beta_K \tag{4}$$

for some $K \times d$ matrix a_K. For example, if the parameter of interest is the regression function $g(z)$, then $a_K = x_K(z)$.

The standard estimator of (2) is least-squares of y_i on x_{Ki}:

$$\widehat{\beta}_K = \left(\sum_{i=1}^n x_{Ki}x_{Ki}'\right)^{-1} \sum_{i=1}^n x_{Ki}y_i. \tag{5}$$

The corresponding estimator of $g(z)$ is

$$\widehat{g}_K(z) = x_K(z)'\widehat{\beta}_K$$

and that of θ is

$$\widehat{\theta}_K = a\left(\widehat{g}_K\right) = a_K'\widehat{\beta}_K. \tag{6}$$

Define the covariance matrices

$$\begin{aligned} Q_K &= \mathbb{E}\left(x_{Ki}x_{Ki}'\right) \\ S_K &= \mathbb{E}\left(x_{Ki}x_{Ki}'e_{Ki}^2\right) \\ V_K &= Q_K^{-1}S_KQ_K^{-1} \\ V_{\theta K} &= a_K'V_Ka_K. \end{aligned}$$

Thus V_K and $V_{\theta K}$ are the conventional asymptotic covariance matrices for the estimators $\widehat{\beta}_K$ and $\widehat{\theta}_K$ for fixed K.

The standard estimators for the above covariance matrices are

$$\begin{aligned} \widehat{Q}_K &= \frac{1}{n}\sum_{i=1}^n x_{Ki}x_{Ki}' \\ \widehat{S}_K &= \frac{1}{n}\sum_{i=1}^n x_{Ki}x_{Ki}'\widehat{e}_{Ki}^2 \\ \widehat{V}_K &= \widehat{Q}_K^{-1}\widehat{S}_K\widehat{Q}_K^{-1} \\ \widehat{V}_{\theta K} &= a_K'\widehat{V}_Ka_K \end{aligned}$$

where $\widehat{e}_{Ki} = y_i - x_{Ki}'\widehat{\beta}_K$ are the OLS residuals.

3 Matrix Convergence Theory

In this section we describe an important convergence result for sample design matrices. Unlike the existing literature, our convergence result does not require bounded regressors nor trimming.

Let u_i, $i = 1, ..., n$ be a sequence of independent $K \times 1$ vectors for $K \geq 2$, and set $\widehat{M}_n = \frac{1}{n}\sum_{i=1}^n u_iu_i'$ and $\overline{M}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{E}u_iu_i'$.

Theorem 1. *Suppose that for some $s > 2$ and some sequence of constants ξ_K and μ_K that*

$$\|u_i\|_s \le \xi_K, \tag{7}$$

$$\left\|\mathbb{E}u_i u_i'\right\| \le \mu_K,$$

and

$$\frac{\xi_K^{2s/(s-2)} \log K}{n} = o(1). \tag{8}$$

Then

$$\mathbb{E}\left\|\widehat{M}_n - \overline{M}_n\right\| = o\left(\sqrt{1+\mu_K}\right). \tag{9}$$

Furthermore, if $s \ge 4$ then

$$\mathbb{E}\left\|\widehat{M}_n - \overline{M}_n\right\| = O\left(\sqrt{\frac{\xi_K^{2s/(s-2)} \log K}{n}\left(1+\mu_K\right)}\right). \tag{10}$$

The proof of Theorem 1 builds on Theorem 1 of Rudelson (1999) and Lemma 6.2 of Belloni, Chernozhukov, Chetverikov, and Kato (2014) which treat the case of the deterministic bound $\|u_i\| \le \xi_K^*$. Existing convergence theory similarly require deterministic bounds. When the regressors are basis transformation of a random variable z_i with bounded support then the rates for ξ_K^* are known for common sieve transformations, for example, $\xi_K^* = O(K)$ for power series and $\xi_K^* = O(K^{1/2})$ for splines. In these cases we can set $s = \infty$ and $\xi_K = \xi_K^*$ and (8) simplifies to $n^{-1}\xi_K^2 \log K = o(1)$, which is identical to that obtained by Lemma 6.2 of Belloni, Chernozhukov, Chetverikov, and Kato (2014) and the in-probability analog of Lemma 2.1 of Chen and Christensen (2015).

The moment bound (7) is an important generalization of the deterministic bound, and in particular allows for regressors with unbounded support. An example for (7) is when the regressors have a uniformly bounded p^{th} moment. Suppose that $u_i = (u_{1i}, ..., u_{Ki})'$ and $\|u_{ji}\|_s \le C < \infty$. Then an application of Minkowski's inequality implies (7) with $\xi_K = CK^{1/2}$. This is the same rate as obtained, for example, by splines with bounded regressors.

The rate (8) shows that there is a trade-off between the number of finite moments and the allowable rate of divergence of the dimension K. As $s \downarrow 2$ the allowable rate (8) slows to fixed K, and as $s \to \infty$, the allowable rate increases. The moment bounds (9) and (10) imply that $\left\|\widehat{M}_n - \overline{M}_n\right\| = O_p(a_n)$, with a_n a function of the rate in (8) when $s \ge 4$. The difference when $s \ge 4$ is that the latter calculation involves the variance of the sample mean $\widehat{M}_n$ (which requires finite fourth moments) which is able to exploit standard orthogonality properties. When $s < 4$ the bound uses a more primitive calculation and hence obtains a less tight rate.

The bounds (9) and (10) depend on the norm of the covariance matrix $\mathbb{E}u_i u_i'$. In most applications this is either assumed bounded or the bounds (9) and (10) are applied to orthogonalized sequences so μ_K is bounded and can be omitted from these expressions.

Convergence results are often applied to the inverses of normalized moment matrices. Our next result provides a simple demonstration that inverses and square roots of normalized matrices inherit the same convergence rate.

Theorem 2. *Let $\widehat{M}_n$ be a sequence of positive definite $K \times K$ matrices. If $\left\| \widehat{M}_n - I_K \right\| = O_p(a_n)$, and $a_n = o(1)$, then for any r,*

$$\left\| \widehat{M}_n^r - I_K \right\| = O_p(a_n).$$

4 Assumptions

We now list the assumptions which are sufficient for asymptotic normality of the studentized estimators.

Assumption 1. *For some $q > 4$ and $p > 2$ such that $1/q + 1/p = 1/t < 1/2$,*

1. $\dfrac{\|x_{Ki}\|_q}{\sqrt{\lambda_{\min}(Q_K)}} \leq \zeta_K$
2. *Either*
   - *(a)* $\|e_i\|_p < \infty$
     or
   - *(b)* $\lim_{B\to\infty} \sup_z \mathbb{E}\left(e_i^2 1\left(e_i^2 \geq B\right) |z_i = z\right) = 0$
3. $\inf_{\beta \in \mathbb{R}^K} \|g(z_i) - x_{Ki}'\beta\|_p \leq \delta_K = O(1)$
4. $n^{-1}\zeta_K^{2t/(t-2)} \log K = o(1)$
5. $n^{-1}\left(\zeta_K^2 \delta_K\right)^{2q/(q-4)} \log K = O(1)$
6. $\lambda_{\min}\left(\overline{a}_K' V_K \overline{a}_K\right) \geq C > 0$ *where* $\overline{a}_K = a_K \left(a_K' Q_K^{-1} a_K\right)^{-1/2}$ *and* a_K *is defined in (4).*

Assumption 1.1 states that the regressors are bounded in the L^q norm. The bound ζ_K is increasing as the regressor set expands (typically at rate $K^{1/2}$ or K) and this rate will in part determine the allowable rate of divergence for K. Assumption 1.1 normalizes the L^q norm of the regressor by the square root of the smallest eigenvalue of the design matrix so that the assumption is invariant to rescaling.

Assumption 1.2 bounds the regression error e_i using either the unconditional p^{th} moment, or a conditional uniform square integrability bound. A sufficient condition for the latter is that $\mathbb{E}\left(|e_i|^{2+\eta} |x_i\right)$ is uniformly bounded for some $\eta > 0$.

Assumption 1.3 bounds the approximation error in the L^p norm. The L^p norm is weaker than the uniform norm which is conventional in the sieve literature. The L^p norm allows the regressors to have unbounded support. For the bound in Assumption 1.3 to exist, a sufficient condition is $\mathbb{E}|g(z_i)|^p < \infty$, for which a sufficient condition is $\mathbb{E}|y_i|^p < \infty$, which also implies Assumption 1.2(a). Explicit rates for the approximation error δ_K will be derived in Section 6.

Assumption 1.4 controls the rate of growth of the number of regressors K. The allowable rate is increasing in t. Note that when $q = p$ then $t = q/2$. As q and p diverge to infinity the rate simplifes to $n^{-1}\zeta_K^2 \log K = o(1)$ which is the rate obtained by Chen and Christensen (2015) under the assumption of bounded regressors ($q = \infty$) and uniform approximation error bounds ($p = \infty$). As $p \to \infty$, the rate in Assumption 1.4 simplifies to $n^{-1}\zeta_K^{2q/(q-2)} \log K = o(1)$, and similarly when $q \to \infty$ it simplifies to $n^{-1}\zeta_K^{2p/(p-2)} \log K = o(1)$.

Assumption 1.5 also controls the rate of growth of K, and trades off both the rate of growth in the regressor norms ζ_K and approximation errors δ_K. A simple sufficient condition for Assumption 1.5 is $p \leq 4$. (For then $n^{-1}\left(\zeta_K^2\delta_K\right)^{2q/(q-4)} \log K \leq O\left(n^{-1}\zeta_K^{4q/(q-4)} \log K\right) \leq O\left(n^{-1}\zeta_K^{2t/(t-2)} \log K\right) = o(1)$ by Assumption 1.4.). Another simple condition is $n^{-1}\zeta_K^{4q/(q-4)} \log K = O(1)$ (a strengthening of Assumption 1.4 which avoids conditions on δ_K). A third simple condition is $\zeta_K\delta_K = O(1)$ and $p \geq q$ for then $n^{-1}\left(\zeta_K^2\delta_K\right)^{2q/(q-4)} \log K \leq O\left(n^{-1}\zeta_K^{2q/(q-4)} \log K\right) \leq O\left(n^{-1}\zeta_K^{2t/(t-2)} \log K\right) = o(1)$ by Assumption 1.4. A fourth simple condition is $\delta_K n^{1/2} = o(1)$, the undersmoothing condition used in the existing nonparametric regression literature, for then $n^{-1}\left(\zeta_K^2\delta_K\right)^{2q/(q-4)} \log K = o\left(\left(n^{-1}\zeta_K^{2q/(q-2)}\right)^{2(q-2)/(q-4)} \log K\right) = o(1)$ by Assumption 1.4. The undersmoothing condition requires, however, $K \to \infty$, unlike Assumption 1.5 which does not require K to diverge.

Assumption 1.6 bounds the asymptotic covariance matrix $a_K' V_K a_K$ away from singularity. A sufficient (but not necessary) condition for Assumption 1.6 is $\lambda_{\min}\left(Q_K^{-1/2} S_K Q_K^{-1/2}\right) \geq C$. A sufficient condition for the latter is $\sigma_i^2 \geq \underline{\sigma}^2 > 0$, which is the standard assumption in the previous nonparametric sieve literature. Assumption 1.6 is much more general, allowing, for example, for $\sigma_i^2 = z_i^2$ (which is not bounded away from 0), for $\sigma_i^2 = 1(|z_i| \geq 1)$ (which allows $\sigma_i^2 = 0$ with positive probability), and allows for components of $\widehat{\beta}_K$ (those not in $a_K'\widehat{\beta}_K$) to converge at a faster than $n^{-1/2}$. The statement of Assumption 1.6 is not elegant but it is much milder, only requiring that the linear combination $a_K'\widehat{\beta}_K$ does not converge faster than $n^{-1/2}$.

5 Distribution Theory

Our first distribution result is for the least-squares coefficient estimate $\widehat{\beta}_K$ from (5).

Theorem 3. *Under Assumption 1,*

$$\sqrt{n}\widehat{V}_{\theta K}^{-1/2} a_K' \left(\widehat{\beta}_K - \beta_K\right) = \sqrt{n} V_{\theta K}^{-1/2} a_K' \left(\widehat{\beta}_K - \beta_K\right) + o_p(1)$$

$$\longrightarrow_d N(0, I_d).$$

Theorem 3 shows that linear functions of the least-squares estimate $\widehat{\beta}_K$ are asymptotically normal. The asymptotic distribution is centered at the projection coefficient β_K and the linear combination $\alpha_K'\widehat{\beta}_K$ has the conventional asymptotic variance $V_{\theta K}$. This theorem includes parametric regression (fixed K) and nonparametric regression (increasing K) as special cases. The theorem allows unbounded regressors and regression errors ($q < \infty$ and $p < \infty$) as is typical in parametric theory as well as bounded regressors and variances ($q=\infty$ and $p=\infty$) as has previously been assumed in the nonparametric theory. Theorem 1 shows that the assumption of boundedness is unnecessary for asymptotic normality. Instead, conventional moment bounds can be used, with the interesting implication that there is a trade-off between the number of finite moments q and p and the permitted number of regressors K.

Theorem 3 also shows that the asymptotic distribution is unaffected by estimation of the covariance matrix, and the assumptions for the theorem are unaffected.

Another new feature of Theorem 3 is that the distributional result concerns the projection errors e_{Ki}, rather than the regression errors e_i. This is an important distinction which has previously separated the existing parametric and non-parametric literatures. By establishing a CLT using the projection errors we are able to provide a foundation for a unified distribution theory.

Our second distribution result is for the plug-in parameter estimate $\widehat{\theta}_K$ from (6).

Theorem 4. *Under Assumption 1,*

$$\sqrt{n}\widehat{V}_{\theta K}^{-1/2}\left(\widehat{\theta}_K-\theta+a\left(r_K\right)\right)=\sqrt{n}V_{\theta K}^{-1/2}\left(\widehat{\theta}_K-\theta+a\left(r_K\right)\right)+o_p(1)$$

$$\longrightarrow_d N(0,I_d).$$

Theorem 4 shows that least-squares estimates of linear functionals are also asymptotically normal. The theorem includes parametric regression (fixed K) and nonparametric regression (increasing K) as special cases. This is in contrast to the current literature on nonparametric sieve estimation which invariably imposes the nonparametric assumption $K\to\infty$. Theorem 4 provides a richer distribution theory, by smoothly nesting the parametric and nonparametric cases. Theorem 4 shows that the appropriate asymptotic variance is $V_{\theta K}=a_K'Q_K^{-1}\mathbb{E}\left(x_{Ki}x_{Ki}'e_{Ki}^2\right)Q_K^{-1}\alpha_K$ which depends on the projection error e_{Ki}. This is an improvement on nonparametric sieve approximations which use the smaller asymptotic variance $a_K'Q_K^{-1}\mathbb{E}\left(x_{Ki}x_{Ki}'e_i^2\right)Q_K^{-1}\alpha_K$ which is only valid under the nonparametric assumption $K\to\infty$ and is a poorer finite sample approximation.

More importantly, Theorem 4 shows that the asymptotic distribution of the estimate $\widehat{\theta}_K$ is centered at $\theta-a\left(r_K\right)$, not at the desired parameter value θ. The term $a\left(r_K\right)$ is the (finite-sample) bias of the estimator $\widehat{\theta}_K$. This bias decreases as K increases, but not in the parametric case of fixed K. Conventional sieve asymptotic theory ignores the bias term $a(r_K)$ by imposing an undersmoothing assumption such as $\sqrt{n}\delta_K\to 0$ or $\sqrt{n}V_{\theta K}^{-1/2}a\left(r_K\right)\to 0$. An undersmoothing assumption such as this is typical in the nonparametric sieve literature and allows the asymptotic distribution in Theorem 4 to be written without the bias term $a(r_K)$. This is a deterioration in the asymptotic approximation, not an improvement. Theorem 4 is a better asymptotic approximation, as it characterizes the (asymptotic) bias in the nonparametric estimator due to inherent approximation in nonparametric estimation.

The presence of the bias term $a(r_K)$ in Theorem 4 is identical to the common inclusion of bias terms in the asymptotic distribution theory for nonparametric kernel estimation. Theorem 4 shows that the bias term can similarly be included in nonparametric sieve asymptotic theory.

Theorem 4 also shows that the asymptotic distribution is unaffected by estimation of the covariance matrix.

6 Approximation Rates

In this section we give primitive conditions for the approximation in Assumption 1.3, which we re-state here as

$$\inf_{\beta \in \mathbb{R}^K} \left\| g(z_i) - x_{Ki}'\beta \right\|_p \leq \delta_K \tag{11}$$

6.1 Series Moment Bound

Suppose that the regression function satisfies the series expansion $g(z) = \sum_{j=1}^{\infty} \beta_j x_j(z)$ with $\|x_j(z_i)\|_p \leq C < \infty$ and $|\beta_j| \leq Aj^{-a}$ for $a > 1$. Set $\boldsymbol{\beta}_K = (\beta_1, ..., \beta_K)$ so that $g(z_i) - x_{Ki}'\boldsymbol{\beta}_K = \sum_{j=K+1}^{\infty} \beta_j x_j(z_i)$. Then by Minkowski's inequality

$$\begin{aligned}
\inf_{\beta \in \mathbb{R}^K} \left\| g(z_i) - x_{Ki}'\beta \right\|_p &\leq \left\| g(z_i) - x_{Ki}'\boldsymbol{\beta}_K \right\|_p \\
&= \left\| \sum_{j=K+1}^{\infty} \beta_j x_j(z_i) \right\|_p \\
&\leq C \sum_{j=K+1}^{\infty} |\beta_j| \\
&\leq \frac{C}{a-1} K^{-a}
\end{aligned}$$

which is (11) with $\delta_K \sim K^{-a}$.

6.2 Weighted Sup Norm

Let $\mathcal{Z} \subset \mathbb{R}^d$ denote the support of z_i. Suppose that for some weight function function $w : \mathcal{Z} \to \mathbb{R}^+$

$$\inf_{\beta \in \mathbb{R}^K} \sup_{z \in \mathcal{Z}} \left| \frac{g(z) - x_K(z)'\beta}{w(z)} \right| \leq \psi_K \tag{12}$$

and

$$\|w(z_i)\|_p \leq C. \tag{13}$$

It is fairly straightforward to see that (12) and (13) imply (11) with $\delta_K = C\psi_K$.

The norm in (12) is known as the weighted sup norm.

If $\mathcal{Z}$ is bounded and $w(z) = 1$ then (12) is the conventional uniform approximation and (13) is automatically satisfied.

Chen, Hong and Tamer (2005) use (12) with $w(z) = (1 + \|z\|^2)^{\omega/2}$ to allow for regressors with unbounded support. They do not discuss primitive conditions for (12). A theory of approximation in weighted sup norm with this weight function is given in Chapter 6 of Triebel (2006).

We use (12)-(13) to provide a set of sufficient conditions for spline approximations with unbounded regressors in Section 6.5 below.

6.3 Polynomials

There is a rich literature in approximation theory (but apparently unknown in econometrics) giving conditions for polynomial weighted L^p approximations which can be used to establish (11).

Assume $d = 1$ and let $x_K(z) = (1, z, z^2, ..., z^{K-1})$ be powers of z. The following theorem is from the monograph of Mhaskar (1996)

Proposition 1. *(Mhaskar) If for some integers $1 \leq p \leq \infty$ and $s \geq 0$, and for some $\alpha > 1$ and $A > 0$, $g^{(s-1)}(z)$ is absolutely continuous and $\left(\int \left|g^{(s)}(z)\right|^p \exp\left(-A\left|z\right|^\alpha\right)\right)^{1/p} \leq C < \infty$, then there is a $c < \infty$ such that for every integer $K \geq s+1$,*

$$\inf_{\beta \in \mathbb{R}^K} \left(\int \left|g(z) - x_K'(z)\beta\right|^p \exp\left(-A\left|z\right|^\alpha\right) dz\right)^{1/p} \leq cK^{-s(1-1/\alpha)}.$$

Mhaskar's theorem can be used to directly imply a sufficient condition for (11).

Theorem 5. *If z_i has a density $f(z)$ which satisfies $f(z) \leq C\exp\left(-A\left|z\right|^\alpha\right)$ for some $C < \infty$, $\alpha > 1$, and $A > 0$, and for some integers s and p, $g^{(s-1)}(z)$ is absolutely continuous and $\int \left|g^{(s)}(z)\right|^p e^{-A|z|^\alpha} dz \leq C < \infty$, then (11) holds with $\delta_K = O\left(K^{-s(1-1/\alpha)}\right)$.*

Theorem 5 requires the regressor z_i to have a density with tails thinner than exponential, which includes the Gaussian case ($\alpha = 2$). The faster the decay rate α of the density, the faster the rate of decay of the bias bound δ_K. In the limiting case $\alpha \to \infty$ (bounded regressors) we obtain $\delta_K = O\left(K^{-s}\right)$ which is the rate known for bounded polynomial support. However as $\alpha \to 1$ the rate becomes arbitrarily slow. The intermediate Gaussian case ($\alpha = 2$) yields the rate $\delta_K = O\left(K^{-s/2}\right)$.

In the polynomial case we expect $\zeta_K = O(K)$ so to satisfy Assumption 1.5 we need $s(1-1/\alpha) \geq 1$, or $s \geq 2$ in the case of Gaussian z_i.

The assumptions require the s^{th} derivative of $g(z)$ to satisfy the weighted L^p bound $\int \left|g^{(s)}(z)\right|^p e^{-A|z|^\alpha} \leq C$. If the derivative is bounded, e.g. $\left|g^{(s)}(z)\right| \leq B < \infty$, then this requirement holds for all p, so (11) holds with the uniform norm. The weighted L^p bound is much weaker, allowing for functions $g(x)$ with unbounded derivatives. For example, Theorem 5 applies to the exponential regression function $g(z) = \exp(az)$

6.4 Multivariate Polynomials

Assume $d > 1$ and write $z = (z_1, ..., z_d)$. For integer m set $K = m^d$ and set $x_K(z)$ to be the $K \times 1$ vector of tensor products of $(1, z_j, z_j^2, ..., z_j^{m-1})$. Write $g^{(\mathbf{s})}(z) = \frac{\partial^s}{\partial z_1^{s_1} \cdots \partial z_d^{s_d}} g(z)$ where $\mathbf{s} = (s_1, \cdots, s_d)$ and set $|\mathbf{s}| = s_1 + \cdots + s_d$. Let $|z|_\alpha = |z_1|^\alpha + \cdots + |z_d|^\alpha$.

Proposition 2. *(Maioriv-Meir) If for some integers $1 \leq p \leq \infty$ and $s \geq 0$, and for some $\alpha \geq 2$ and $A > 0$, $\left(\int \left|g^{(\mathbf{s})}(z)\right|^p \exp\left(-A\left|z\right|_\alpha\right)\right)^{1/p} \leq C < \infty$ for all $|\mathbf{s}| \leq s$ then there is a $c < \infty$ such that for every integer K,*

$$\inf_{\beta \in \mathbb{R}^K} \left(\int \left|g(z) - x_K'(z)\beta\right|^p \exp\left(-A\left|z\right|_\alpha\right) dz\right)^{1/p} \leq cK^{-s(1-1/\alpha)/d}$$

Theorem 6. *If z_i has a density $f(z)$ which satisfies $f(z) \leq C\exp\left(-A\left|z\right|_\alpha\right)$ for some $C < \infty$, $\alpha \geq 2$ and $A > 0$, and for some integers s and p, $\left(\int \left|g^{(\mathbf{s})}(z)\right|^p \exp\left(-A\left|z\right|_\alpha\right) dz\right)^{1/p} \leq C < \infty$ for all $|\mathbf{s}| \leq s$, then (11) holds with $\delta_K = O\left(K^{-s(1-1/\alpha)/d}\right)$.*

6.5 Splines

Assume $d = 1$.

An s^{th} order polynomial takes the form $p(z) = \sum_{j=1}^{s} c_0 z^{j-1}$. An s^{th} order spline with $K+1$ evenly spaced knots on the interval $[-b, b]$ is an s^{th} order polynomial on each interval $I_j = [\tau_{j-1}, \tau_j]$ for $j = 1, ..., K$ where $\tau_j = -b + 2jb/K$ are called the knots, and the spline is constrained to have continuous $s-2$ derivatives. The spline is a piecewise polynomial.

There is a rich approximation literature for splines on bounded intervals $[-b, b]$ but none (to my knowledge) on unbounded sets such as $\mathbb{R}$. We make such an extension. For each K we define the spline as follows. For some $A > 0$ and $\alpha > 1$ set $b = BK^{1/\alpha}$ and the knots $\tau_j = -b + 2jb/K$. Then an s^{th} order spline on $\mathbb{R}$ is an s^{th} order polynomial on each interval $I_j = [\tau_{j-1}, \tau_j]$ for $j = 0, ..., K+1$ where $I_0 = (-\infty, \tau_0]$ and $I_{K+1} = [\tau_K, \infty)$, and is contrained to have continuous $s-2$ derivatives.

Theorem 7. *If $\sup_z \left|g^{(s)}(z)\right| \leq C$, $\sup_z \left|g^{(s-1)}(z)\right| \leq C$ and $\left\|z_i\right\|_{\alpha sp} \leq C < \infty$ then (11) holds with $\delta_K = O\left(K^{-s(1-1/\alpha)}\right)$.*

This result complements that for polynomial approximation. The polynomial approximation theorem impose weaker conditions on the regression function $g(x)$ (it does not require boundedness of the s^{th} derivative) but stronger moment conditions on the regressor z_i. The latter are because a polynomial necessarily requires that all moments of the regressor are finite, unlike a spline which is only an s^{th} order power. The weaker derivative condition on the regression function is allowed because of the rich literature on weighted polynomial approximation, while no analogous literature appears to exist for spline approximations.

7 Matrix Convergence Proofs

In this section we provide the proofs for the matrix convergence results of Section 3. Our proofs will making frequent use of a trimming argument which exploits the following simple (and classic) inequality

Lemma 1. *For any $s \geq v$ and $b > 0$ such that $\mathbb{E}|U_i|^s < \infty$,*

$$\mathbb{E}\left(|U_i|^v 1\left(|U_i| > b\right)\right) \leq \frac{\mathbb{E}|U_i|^s}{b^{s-v}}$$

Proof

$$\mathbb{E}\left(|U_i|^v 1\left(|U_i| > b\right)\right) = \mathbb{E}\left(\frac{|U_i|^s}{|U_i|^{s-v}} 1\left(|U_i|^{s-v} > b^{s-v}\right)\right) \leq \mathbb{E}\left(\frac{|U_i|^s}{b^{s-v}} 1\left(|U_i|^{s-v} > b^{s-v}\right)\right) \leq \frac{\mathbb{E}|U_i|^s}{b^{s-v}}.$$

■

We now establish Theorem 1. We extend the convergence result for bounded matrices established by Rudelson (1999) and Belloni, Chernozhukov, Chetverikov, and Kato (2014) to allow for unbounded regressors, using a trimming argument. The trimming error is bounded off using an L^2 bound when the regressors have four or greater moments, and using an L^1 bound when the regressors have less than four moments.

Proof of Theorem 1. Set

$$b_K = \left(\xi_K^{2s} \frac{n}{\log K}\right)^{1/(2s-2)}. \tag{14}$$

(8) implies

$$\frac{b_K^2 \log K}{n} = \left(\frac{\xi_K^{2s/(s-2)} \log K}{n}\right)^{(s-2)/(s-1)} = o(1). \tag{15}$$

Define

$$\widehat{M}_{1n} = \frac{1}{n}\sum_{i=1}^{n} u_i u_i' 1\left(\|u_i\| \leq b_K\right) \tag{16}$$

$$\widehat{M}_{2n} = \frac{1}{n}\sum_{i=1}^{n} u_i u_i' 1\left(\|u_i\| > b_K\right). \tag{17}$$

By the triangle inequality

$$\mathbb{E}\left\|\widehat{M}_n - \overline{M}_n\right\| \leq \mathbb{E}\left\|\widehat{M}_{1n} - \mathbb{E}\widehat{M}_{1n}\right\| + \mathbb{E}\left\|\widehat{M}_{2n} - \mathbb{E}\widehat{M}_{2n}\right\|. \tag{18}$$

Take the first term on the right side of (18). Since $\|u_i 1\left(\|u_i\| \leq b_K\right)\| \leq b_K$,

$$\left\|u_i u_i' 1\left(\|u_i\| \leq b_K\right)\right\| = \|u_i 1\left(\|u_i\| \leq b_K\right)\|^2 \leq b_K^2. \tag{19}$$

The assumption $\|\mathbb{E}u_iu_i'\| \le \mu_K$ implies that

$$\begin{aligned}
\|M_{1n}\| &\le \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbb{E}u_iu_i'1\left(\|u_i\|\le b_K\right)\right\| \\
&= \frac{1}{n}\sum_{i=1}^{n}\max_{\alpha'\alpha=1}\mathbb{E}\left(\left(\alpha'u_i\right)^2 1\left(\|u_i\|\le b_K\right)\right) \\
&\le \frac{1}{n}\sum_{i=1}^{n}\max_{\alpha'\alpha=1}\mathbb{E}\left(\alpha'u_i\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbb{E}u_iu_i'\right\| \le \mu_K. \qquad (20)
\end{aligned}$$

Equations (15), (19), and (20) establish the conditions for Theorem 1 of Rudelson (1999) (or Lemma 6.2 of Belloni, Chernozhukov, Chetverikov, and Kato (2014)), which states

$$\begin{aligned}
\mathbb{E}\left\|\widehat{M}_{1n}-\mathbb{E}\widehat{M}_{1n}\right\| &\le O\left(\sqrt{\frac{b_K^2\log K}{n}\left(1+\mu_K\right)}\right) \qquad (21) \\
&= o\left(\sqrt{(1+\mu_K)}\right) \qquad (22)
\end{aligned}$$

by (15).

Now take the second term on the right side of (18). By the triangle inequality, Lemma 1, (7) and (14),

$$\begin{aligned}
\mathbb{E}\left\|\widehat{M}_{2n}-\mathbb{E}\widehat{M}_{2n}\right\| &\le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\|u_iu_i'1\left(\|u_i\|>b_K\right)\right\| \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\|u_i\|^2 1\left(\|u_i\|>b_K\right)\right) \\
&\le \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{E}\|u_i\|^s}{b_K^{s-2}} \\
&\le \frac{\xi_K^s}{b_K^{s-2}} \\
&= \left(\frac{\xi_K^{2s/(s-2)}\log K}{n}\right)^{(s-2)/(2s-2)} = o(1) \qquad (23)
\end{aligned}$$

by (8) and $s>2$.

Equations (18), (22), and (23) show (9).

To establish (10), again make the decomposition (18) but with $b_K = \xi_K^{s/(s-2)}$. Equations (19),

(20) and (21) hold, implying

$$\mathbb{E}\left\|\widehat{M}_{1n}-\mathbb{E}\widehat{M}_{1n}\right\| \leq O\left(\sqrt{\frac{\xi_K^{2s/(s-2)}\log K}{n}\left(1+\mu_K\right)}\right). \tag{24}$$

Set

$$U_i = u_i u_i' 1\left(\|u_i\| > b_K\right) - \mathbb{E}\left(u_i u_i' 1\left(\|u_i\| > b_K\right)\right).$$

Using the inequality $\|A\|^2 = \lambda_{\max}(A'A) \leq \operatorname{tr}(A'A)$, the fact that U_i are independent and mean zero, $s \geq 4$, Lemma 1, (7), and $b_K = \xi_K^{s/(s-2)}$,

$$\begin{aligned}
\mathbb{E}\left\|\widehat{M}_{2n}-\mathbb{E}\widehat{M}_{2n}\right\|^2 &\leq \mathbb{E}\operatorname{tr}\left(\left(\widehat{M}_{2n}-\mathbb{E}\widehat{M}_{2n}\right)\left(\widehat{M}_{2n}-\mathbb{E}\widehat{M}_{2n}\right)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\operatorname{tr}\left(U_i U_j\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\operatorname{tr}\left(U_i U_i\right) \\
&\leq \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left(\|u_i\|^4 1\left(\|u_i\| > b_K\right)\right) \\
&\leq \frac{\xi_K^s}{n b_K^{s-4}} \\
&= \frac{\xi_K^{2s/(s-2)}}{n}.
\end{aligned}$$

Using Liapunov's inequality we conclude that

$$\mathbb{E}\left\|\widehat{M}_{2n}-\mathbb{E}\widehat{M}_{2n}\right\| \leq \left(\mathbb{E}\left\|\widehat{M}_{2n}-\mathbb{E}\widehat{M}_{2n}\right\|^2\right)^{1/2} \leq \sqrt{\frac{\xi_K^{2s/(s-2)}}{n}}. \tag{25}$$

Equations (18), (24), and (25) show (10). ■

Proof of Theorem 2. Take any $\varepsilon > 0$. For some $0 < \eta < 1$ define the event

$$E_n = \left\{\left\|\widehat{M}_n - I_K\right\| \leq B a_n \leq \eta\right\}.$$

The assumptions that $\left\|\widehat{M}_n - I_K\right\| = O_p(a_n)$ and $a_n = o(1)$ imply that we can pick $B < \infty$ and n sufficiently large so that $P(E_n) \geq 1 - \varepsilon$.

Since $\widehat{M}_n$ is positive definite we can write $\widehat{M}_n = H'\Lambda H$ where $H'H = I_K$ and $\Lambda = \operatorname{diag}(\lambda_1, ..., \lambda_K)$. Then

$$\left\|\widehat{M}_n - I_K\right\| = \left\|H'\left(\Lambda - I_K\right)H\right\| = \left\|\Lambda - I_K\right\| = \max_{1\leq j\leq K}\left|\lambda_j - 1\right|.$$

Thus the event E_n implies

$$\max_{1\leq j\leq K}|\lambda_j-1|\leq Ba_n\leq\eta.$$

Combined with a Taylor expansion we find

$$\max_{1\leq j\leq K}\left|\lambda_j^r-1\right|\leq|r|\left(1-\eta\right)^{-|r-1|}\max_{1\leq j\leq K}|\lambda_j-1|\leq|r|\left(1-\eta\right)^{-|r-1|}Ba_n.$$

This implies that on event E_n,

$$\left\|\widehat{M}_n^r-I_K\right\|=\left\|H'\left(\Lambda^r-I_K\right)H\right\|=\max_{1\leq j\leq K}\left|\lambda_j^r-1\right|\leq|r|\left(1-\eta\right)^{-|r-1|}Ba_n.$$

Since $P\left(E_n\right)\geq 1-\varepsilon$ this means $\left\|\widehat{M}_n^r-I_K\right\|=O_p\left(a_n\right)$ as claimed. ■

8 Structure of Proofs of Theorems 3 and 4

We start by replacing the regressor vector x_{Ki} with the orthogonalized vector $x_{Ki}^*=Q_K^{-1/2}x_{Ki}$, and similarly replace a_K with $a_K^*=Q_K^{-1/2}a_K$, $\overline{a}_K$ with $\overline{a}_K^*=a_K^*\left(a_K^{*\prime}a_K^*\right)^{-1/2}$, S_K and V_K with $S_K^*=Q_K^{-1/2}S_KQ_K^{-1/2}$, β_K with $\beta_K^*=Q_K^{1/2}\beta_K$ and $\widehat{\beta}_K^*=Q_K^{1/2}\widehat{\beta}_K$. With this replacement, the statistics of Theorems 3 and 4 remain unaltered. The only parts of Assumption 1 affected are parts 1 and 6. Since $\mathbb{E}\left(x_{Ki}^*x_{Ki}^{*\prime}\right)=I_K$, we see that

$$\frac{\left\|x_{Ki}^*\right\|_q}{\sqrt{\lambda_{\min}\left(\mathbb{E}\left(x_{Ki}^*x_{Ki}^{*\prime}\right)\right)}}\leq\left\|Q_K^{-1/2}\right\|\left\|x_{Ki}\right\|_q=\frac{\left\|x_{Ki}\right\|_q}{\sqrt{\lambda_{\min}\left(Q_K\right)}}\leq\zeta_K$$

and thus Assumption 1.1 for x_{Ki} implies the same bound for x_{Ki}^*. Since

$$\begin{aligned}\overline{a}_K'V_K\overline{a}_K&=\left(a_K'Q_K^{-1}a_K\right)^{-1/2}a_K'Q_K^{-1}S_KQ_K^{-1}a_K\left(a_K'Q_K^{-1}a_K\right)^{-1/2}\\&=\left(a_K^{*\prime}a_K^*\right)^{-1/2}a_K^{*\prime}S_K^*a_K^*\left(a_K^{*\prime}a_K^*\right)^{-1/2}\\&=\overline{a}_K^{*\prime}S_K^*\overline{a}_K^*\end{aligned}$$

we see that Assumption 1.6 is unaffected by this replacement. Hence there is no loss of generality in making this replacement, or equivalently in simply assuming that $Q_K=I_K$. We will thus impose this normalization for the remainder of the argument.

Now consider the statistics from Theorem 3. Define

$$A_K=a_KV_{\theta K}^{-1/2}\tag{26}$$

and

$$\widehat{V}_{AK}=A_K'\widehat{V}_KA_K\ .$$

This is convenient normalization for then $A_K' V_K A_K = I_d$. We find that

$$\begin{aligned}\sqrt{n}\widehat{V}_{\theta K}^{-1/2} a_K'\left(\widehat{\beta}_K - \beta_K\right) &= \sqrt{n}\widehat{V}_{AK}^{-1/2} A_K'\left(\widehat{\beta}_K - \beta_K\right) \\ &= \sqrt{n}A_K'\left(\widehat{\beta}_K - \beta_K\right) + \left(\widehat{V}_{AK}^{-1/2} - I_K\right)\sqrt{n}A_K'\left(\widehat{\beta}_K - \beta_K\right).\end{aligned}$$

Since $\widehat{\beta}_K = \left(X_K' X_K\right)^{-1} X_K' Y$ and $Y = X_K \beta_K + e_K$,

$$\begin{aligned}\sqrt{n}A_K'\left(\widehat{\beta}_K - \beta_K\right) &= \sqrt{n}A_K'\left(X_K' X_K\right)^{-1} X_K' e_K \\ &= A_K' \widehat{Q}_K^{-1} n^{-1/2} X_K' e_K \\ &= n^{-1/2} A_K' X_K' e_K + A_K'\left(\widehat{Q}_K^{-1} - I_K\right) n^{-1/2} X_K' e_K.\end{aligned}$$

The results of Theorem 3 then follow from the following

$$n^{-1/2} A_K' X_K' e_K \longrightarrow_d N(0, I_d) \tag{27}$$

$$A_K'\left(\widehat{Q}_K^{-1} - I_K\right) n^{-1/2} X_K' e_K = o_p(1) \tag{28}$$

$$\left\|\widehat{V}_{AK}^{-1/2} - I_K\right\| = o_p(1). \tag{29}$$

These are shown in Lemmas 3, 4, and 5 below.

Now consider the statistic from Theorem 4. Since $g = g_K + r_K$, then by linearity and (4)

$$\theta = a\left(g_K\right) + a\left(r_K\right) = a_K' \beta_K + a\left(r_K\right).$$

Thus

$$\widehat{\theta}_K - \theta = a_K' \widehat{\beta}_K - a_K' \beta_K - a\left(r_K\right)$$

and

$$\sqrt{n}\widehat{V}_{\theta K}^{-1/2}\left(\widehat{\theta}_K - \theta + a\left(r_K\right)\right) = \sqrt{n}\widehat{V}_{\theta K}^{-1/2} a_K'\left(\widehat{\beta}_K - \beta_K\right)$$

which is identical to the statistic in Theorem 3. It follows that Lemmas 3-5 are sufficient to establish both Theorems 3 and 4.

To establish these results, it will be convenient to establish a set of intermediate results, which we list and prove here. We start by decomposing the projection error e_{Ki}. Define the projection approximation error, the L^p best approximation coefficients, L^p approximation error, and differences

$$p_{Ki} = g(z_i) - x_K(z_i)'\beta_K \tag{30}$$

$$\beta_K^* = \underset{\beta\in\mathbb{R}^K}{\operatorname{argmin}} \left\| g(z_i) - x_{Ki}'\beta \right\|_p \tag{31}$$

$$r_{Ki} = g(z_i) - x_{Ki}'\beta_K^* \tag{32}$$

$$r_{Ki}^* = p_{Ki} - r_{Ki}. \tag{33}$$

Then we have the decompositions

$$e_{Ki} = e_i + p_{Ki} = e_i + r_{Ki} + r_{Ki}^*. \tag{34}$$

Next, define the following vectors and matrices

$$T_{1n} = n^{-1/2} \sum_{i=1}^n x_{Ki} e_i$$

$$T_{2n} = n^{-1/2} \sum_{i=1}^n x_{Ki} p_{Ki}$$

$$\overline{S}_K = n^{-1} \sum_{i=1}^n x_{Ki} x_{Ki}' \sigma_i^2$$

$$\widetilde{S}_K = \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' e_{Ki}^2$$

$$\widehat{Q}_{4K} = \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' \left(x_{Ki}' x_{Ki} \right).$$

Lemma 2. *Under Assumption 1 and* $Q_K = I_K$

1. $\|A_K\|^2 \leq C^{-1}$
2. $\|p_{Ki}\|_2 \leq \|r_{Ki}\|_2 \leq \|r_{Ki}\|_p \leq \delta_K$
3. $|r_{Ki}^*| \leq \|x_{Ki}\| \, \delta_K$
4. $\left\| \widehat{Q}_K - I_K \right\| = O_p\left(\sqrt{\frac{\zeta_K^{2q/(q-2)} \log K}{n}} \right)$ *and* $\left\| \widehat{Q}_K \right\| = O_p(1)$
5. $\left\| \widehat{Q}_K^{-1} - I_K \right\| = O_p\left(\sqrt{\frac{\zeta_K^{2q/(q-2)} \log K}{n}} \right)$ *and* $\left\| \widehat{Q}_K^{-1} \right\| = O_p(1)$
6. $\|\mathbb{E}(T_{1n} T_{1n}' | Z)\| = \left\| \overline{S}_K \right\| = O_p\left(\zeta_K^{4/(p-2)} \right)$
   where $Z = [z_1, ..., z_n]$

7. $\left\|\widetilde{S}_K\right\| = O_p\left(\zeta_K^{4/(p-2)}\right)$

8. $\left\|A_K'\widetilde{S}_K A_K - I_K\right\| = o_p(1)$

9. $\|T_{2n}\| = O_p\left(\left(\frac{\zeta_K^{2q/(q-2)}\log K}{n}\right)^{-(q-4)/2q}\right)$

10. $\left\|\widehat{Q}_{4K}\right\| = O_p\left(\zeta_K^{2q/(q-2)}\right)$

Proof of Lemma 2.1. Recall from Assumption 1.6 the definition $\overline{a}_K = a_K\left(a_K' a_K\right)^{-1/2}$. Then since $\|A'A\| = \|AA'\|$ and $\overline{a}_K'\overline{a}_K = I_d$,

$$\begin{aligned}
\|A_K\|^2 &= \left\|a_K\left(a_K'V_K a_K\right)^{-1} a_K'\right\| \\
&= \left\|\overline{a}_K\left(\overline{a}_K'V_K\overline{a}_K\right)^{-1}\overline{a}_K'\right\| \\
&= \left\|\left(\overline{a}_K'V_K\overline{a}_K\right)^{-1/2}\overline{a}_K'\overline{a}_K\left(\overline{a}_K'V_K\overline{a}_K\right)^{-1/2}\right\| \\
&= \left\|\left(\overline{a}_K'V_K\overline{a}_K\right)^{-1}\right\| \\
&= 1/\lambda_{\min}\left(\overline{a}_K'V_K\overline{a}_K\right) \\
&\leq C^{-1},
\end{aligned}$$

the final inequality using Assumption 1.6. ■

Proof of Lemma 2.2. The first inequality holds since the projection coefficient β_K minimizes the approximation error in L^2. The second inequality is Liapunov's. The final is Assumption 1.3. ■

Proof of Lemma 2.3. Since $y_i = x_{Ki}'\beta_K^* + e_i + r_{Ki}$ and $\mathbb{E}\left(x_{Ki}e_i\right) = 0$,

$$\begin{aligned}
r_{Ki}^* &= x_{Ki}'\left(\beta_K^* - \beta_K\right) \\
&= x_{Ki}'\left(\beta_K^* - \mathbb{E}\left(x_{Ki}x_{Ki}'\right)^{-1}\mathbb{E}\left(x_{Ki}y_i\right)\right) \\
&= -x_{Ki}'\mathbb{E}\left(x_{Ki}x_{Ki}'\right)^{-1}\mathbb{E}\left(x_{Ki}r_{Ki}\right).
\end{aligned}$$

Thus by the Schwarz inequality, $\mathbb{E}\left(x_{Ki}x'_{Ki}\right) = I_K$, the projection inequality, and Lemma 2.2

$$\begin{aligned}
|r^*_{Ki}| &\leq \|x_{Ki}\| \left|\mathbb{E}\left(r_{Ki}x'_{Ki}\right)\mathbb{E}\left(x_{Ki}x'_{Ki}\right)^{-1}\mathbb{E}\left(x_{Ki}x'_{Ki}\right)^{-1}\mathbb{E}\left(x_{Ki}r_{Ki}\right)\right|^{1/2} \\
&= \|x_{Ki}\| \left|\mathbb{E}\left(r_{Ki}x'_{Ki}\right)\mathbb{E}\left(x_{Ki}x'_{Ki}\right)^{-1}\mathbb{E}\left(x_{Ki}r_{Ki}\right)\right|^{1/2} \\
&\leq \|x_{Ki}\| \left(\mathbb{E}r^2_{Ki}\right)^{1/2} \\
&\leq \|x_{Ki}\| \delta_K
\end{aligned}$$

as stated. ■

Proof of Lemma 2.4. Setting $u_i = x_{Ki}$, $s = q > 4$, and $\xi_K = \zeta_K$, and noting $\|\mathbb{E}x_{Ki}x'_{Ki}\| = 1$ we can apply Theorem 1 and Markov's inequality to establish the bound for $\left\|\widehat{Q}_K - I_K\right\|$. By Assumption 1.4 and $q \geq t$ this bound is $o_p(1)$. Then by the triangle inequality

$$\left\|\widehat{Q}_K\right\| \leq \|I_K\| + \left\|\widehat{Q}_K - I_K\right\| \leq 1 + O_p\left(\sqrt{\zeta_K^{2q/(q-2)}\log K/n}\right) = O_p(1)$$

as stated. ■

Proof of Lemma 2.5. Since the bound in Lemma 2.4 is $o_p(1)$, Theorem 2 establishes the bound for $\left\|\widehat{Q}_K^{-1} - I_K\right\|$. The bound for $\left\|\widehat{Q}_K^{-1}\right\|$ follows the same argument as for $\left\|\widehat{Q}_K\right\|$. ■

Proof of Lemma 2.6. The first equality holds since

$$\mathbb{E}\left(T_{1n}T'_{1n}|Z\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{Ki}x'_{Kj}\mathbb{E}\left(e_ie_j|Z\right) = \frac{1}{n}\sum_{i=1}^{n} x_{Ki}x'_{Ki}\sigma_i^2 = \overline{S}_K.$$

Now suppose that Assumption 1.2(a) holds. By the conditional Jensen inequality

$$\|\sigma_i\|_p = \left(\mathbb{E}\left(\mathbb{E}\left(e_i^2 \mid z_i\right)\right)^{p/2}\right)^{1/p} \leq \|e_i\|_p \leq D \tag{35}$$

for some $D < \infty$. Set $v = 2q/(q-2)$ so that $1/v + 1/q = 1/2$, and note that since $\|\mathbb{E}x_{Ki}x'_{Ki}\| = 1$, Holder's inequality, Assumption 1.1, and (35)

$$\begin{aligned}\left\|\mathbb{E}\overline{S}_K\right\| &\leq \left\|\mathbb{E}\left(x_{Ki}x_{Ki}'\sigma_i^2 1\left(\|\sigma_i\| \leq \zeta_K^{2/(p-2)}\right)\right)\right\| + \mathbb{E}\left(\|x_{Ki}\|^2 \sigma_i^2 1\left(\|\sigma_i\| > \zeta_K^{2/(p-2)}\right)\right) \\ &\leq \zeta_K^{4/(p-2)} + \left(\mathbb{E}\|x_{Ki}\|^q\right)^{2/q}\left(\mathbb{E}\left(\sigma_i^v 1\left(\|\sigma_i\| > \zeta_K^{2/(p-2)}\right)\right)\right)^{2/v} \\ &\leq \zeta_K^{4/(p-2)} + \zeta_K^2\left(\mathbb{E}\left(\frac{\sigma_i^p}{\zeta_K^{2(p-v)/(p-2)}}\right)\right)^{2/\nu} \\ &\leq \zeta_K^{4/(p-2)} + \frac{\zeta_K^2}{\zeta_K^{2(v-4)/(p-2)}}D^{2p/v} \\ &= \left(1 + D^{2p/v}\right)\zeta_K^{4/(p-2)}. \end{aligned} \tag{36}$$

Using Holder's inequality, Assumption 1.1, and (35)

$$\left(\mathbb{E}\|x_{Ki}\sigma_i\|^t\right)^{1/t} \leq \|x_{Ki}\|_q \|\sigma_i\|_p \leq D\zeta_K.$$

Theorem 1 with $u_{Ki} = x_{Ki}\sigma_i$ and $s = t$ implies that

$$\mathbb{E}\left\|\overline{S}_K - \mathbb{E}\overline{S}_K\right\| = o\left(\sqrt{1 + \left\|\mathbb{E}\overline{S}_K\right\|}\right) = o\left(\zeta_K^{2/(p-2)}\right) \tag{37}$$

where the final bound uses (36). Equations (36) and (37) imply that

$$\left\|\overline{S}_K\right\| \leq \left\|\overline{S}_K - \mathbb{E}\overline{S}_K\right\| + \left\|\mathbb{E}\overline{S}_K\right\| \leq O_p(\zeta_K^{4/(p-2)}) \tag{38}$$

as claimed.

Alternatively, assume that Assumption 1.2(b) holds and thus $\sigma_i^2 \leq \sigma^2(0)$ where the latter is defined in (51). Then using Lemma 2.4

$$\left\|\overline{S}_K\right\| \leq \left\|\widehat{Q}_K\right\| \max_i \sigma_i^2 \leq O_p(1) \leq O_p(\zeta_K^{4/(p-2)}.$$

as claimed. ■

Proof of Lemma 2.7. The proof is identical to that of Lemma 2.5, except that $|e_i|$ replaces σ_i. Note that (35) applies to both. ■

Proof of Lemma 2.8. Notice that

$$\mathbb{E}\left(A_K'\widetilde{S}_K A_K\right) = A_K' S_K A_K = I_K.$$

Suppose that Assumption 1.2(a) holds. Then

$$\left(\mathbb{E}\left\|A_K' x_{Ki} e_i\right\|^t\right)^{1/t} \leq \|A_K\| \left\|x_{Ki}\right\|_q \left\|e_i\right\|_p \leq D\zeta_K$$

by Lemma 2.1, Assumption 1.1 and (35). Theorem 1 with $u_{Ki} = A_K' x_{Ki} e_i$ and $s = t$ implies that

$$\mathbb{E}\left\|A_K' \widetilde{S}_K A_K - I_K\right\| = o(1)$$

Markov's inequality implies the stated result.

Now suppose that Assumpiton 1.2(b) holds. [Proof incomplete.] ■

Proof of Lemma 2.9. Since $x_{Ki} p_{Ki}$ is iid and mean zero, the definition $p_{Ki} = r_{Ki} + r_{Ki}^*$, and the C_r inequality

$$\begin{aligned}
\mathbb{E}\left\|T_{2n}\right\|^2 &= \mathbb{E}\left[\frac{1}{n}\left(\sum_{i=1}^n x_{Ki} p_{Ki}\right)' \left(\sum_{j=1}^n x_{Kj} p_{Kj}\right)\right] \\
&= \mathbb{E}\left(\left\|x_{Ki}\right\|^2 p_{Ki}^2\right) \\
&\leq 2\mathbb{E}\left(\left\|x_{Ki}\right\|^2 r_{Ki}^2\right) + 2\mathbb{E}\left(\left\|x_{Ki}\right\|^2 r_{Ki}^{*2}\right). \qquad (39)
\end{aligned}$$

We now examine the two terms on the right-side of (39). By Liapunov's inequality, Holder's inequality, Assumption 1.1, and Lemma 2.2

$$\begin{aligned}
\mathbb{E}\left(\left\|x_{Ki}\right\|^2 r_{Ki}^2\right) &\leq \left(\mathbb{E}\left(\left\|x_{Ki}\right\|^t \left|r_{Ki}\right|^t\right)\right)^{2/t} \\
&\leq \left\|x_{Ki}\right\|_q^2 \left\|r_{Ki}\right\|_p^2 \\
&\leq \zeta_K^2 \delta_K^2. \qquad (40)
\end{aligned}$$

Applying the C_r inequality to (33) and two applications of Lemma 2.2,

$$\mathbb{E} r_{Ki}^{*2} \leq 2\mathbb{E} p_{Ki}^2 + 2\mathbb{E} r_{Ki}^2 \leq 4\delta_K^2. \qquad (41)$$

Then using (41), Lemma 2.3, Lemma 1, and Assumption 1.1,

$$\begin{aligned}
\mathbb{E}\left(\left\|x_{Ki}\right\|^2 r_{Ki}^{*2}\right) &= \mathbb{E}\left(\left\|x_{Ki}\right\|^2 r_{Ki}^{*2} 1\left(\left\|x_{Ki}\right\| \leq \zeta_K^{q/(q-2)}\right)\right) + \mathbb{E}\left(\left\|x_{Ki}\right\|^2 r_{Ki}^{*2} 1\left(\left\|x_{Ki}\right\| > \zeta_K^{q/(q-2)}\right)\right) \\
&\leq \zeta_K^{2q/(q-2)} \mathbb{E}\left(r_{Ki}^{*2}\right) + \mathbb{E}\left(\left\|x_{Ki}\right\|^4 1\left(\left\|x_{Ki}\right\| > \zeta_K^{q/(q-2)}\right)\right) \delta_K^2 \\
&\leq 4\zeta_K^{2q/(q-2)} \delta_K^2 + \frac{\zeta_K^q}{\left(\zeta_K^{q/(q-2)}\right)^{q-4}} \delta_K^2 \\
&= 5\zeta_K^{2q/(q-2)} \delta_K^2. \qquad (42)
\end{aligned}$$

Equation (39), (40), (42), and $2 \leq 2q/(q-2)$ imply

$$\mathbb{E}\left\|T_{2n}\right\|^2 \leq 12\zeta_K^{2q/(q-2)}\delta_K^2 \leq O\left(\zeta_K^{(8-2q)/(q-2)}\left(\frac{\log K}{n}\right)^{4/q-1}\right)$$

where the second inequality uses Assumption 1.5 which implies $\delta_K^2 = O\left(\zeta_K^{-4}\left(\log K/n\right)^{4/q-1}\right)$. Markov's inequality establishes the result. ■

Proof of Lemma 2.10. By the triangle inequality

$$\begin{aligned}\left\|\widehat{Q}_{4K}\right\| &\leq \left\|\frac{1}{n}\sum_{i=1}^n x_{Ki}x_{Ki}'\left\|x_{Ki}\right\|^2 1\left(\left\|x_{Ki}\right\| \leq \zeta_K^{q/(q-2)}\right)\right\| \\ &+ \left\|\frac{1}{n}\sum_{i=1}^n x_{Ki}x_{Ki}'\left\|x_{Ki}\right\|^2 1\left(\left\|x_{Ki}\right\| > \zeta_K^{q/(q-2)}\right)\right\| \\ &\leq \zeta_K^{q/(q-2)}\left\|\widehat{Q}_K\right\| \\ &+ \frac{1}{n}\sum_{i=1}^n \left\|x_{Ki}\right\|^4 1\left(\left\|x_{Ki}\right\| > \zeta_K^{q/(q-2)}\right).\end{aligned}$$

The first term is $O_p\left(\zeta_K^{2q/(q-2)}\right)$ by Lemma 2.4. The expectation of the second term is bounded (using Lemma 1) by $\zeta_K^{2q/(q-2)}$ and hence the term is $O_p\left(\zeta_K^{2q/(2-q)}\right)$ by Markov's inequality. This establishes the result. ■

9 Central Limit Theory

In this section we establish equations (27) and (28) which establish asymptotic normality for the non-studentized estimators. What is unconventional in the proof is that we apply the Lindeberg condition separately to the three error components from the decomposition (34). This is useful as they have differing moment properties. Also, our approach allows but does not require the number of series terms K to diverge, and hence the approximation errors must be handled explicitly.

Lemma 3. *Under Assumption 1 and $Q_K = I_K$, $n^{-1/2}A_K'X_K'e_K \longrightarrow_d N(0, I_d)$*

Proof: Take any $\eta \in \mathbb{R}^d$ such that $\eta'\eta = 1$ and set $\alpha_K = A_K\lambda$. By the Cramer-Wold device it is sufficient to show that $n^{-1/2}\sum_{i=1}^n \alpha_K'x_{Ki}e_{Ki} \longrightarrow_d N(0,1)$.

It will be convenient to observe that by the norm inequality $\|AB\| \leq \|A\|\,\|B\|$, $\|\lambda\|^2 = 1$, and Lemma 2.1

$$\left\|\alpha_K\right\|^2 \leq \left\|A_K\right\|^2\left\|\lambda\right\|^2 \leq C^{-1}.$$

The Schwarz inequality then implies

$$\left(\alpha_K'x_{Ki}\right)^2 \leq C^{-1}\left\|x_{Ki}\right\|^2. \tag{43}$$

Since the random variable $\alpha_K' x_{Ki} e_{Ki}$ is iid, mean zero, and unit variance, it is sufficient to verify Lindeberg's condition: that for any $\varepsilon > 0$,

$$\mathbb{E}\left[\left(\alpha_K' x_{Ki} e_{Ki}\right)^2 1\left(\left(\alpha_K' x_{Ki} e_{Ki}\right)^2 > 9\varepsilon n\right)\right] \to 0. \tag{44}$$

By the C_r inequality

$$e_{Ki}^2 \leq 3\left(e_i^2 + r_{Ki}^2 + r_{Ki}^{*2}\right). \tag{45}$$

This implies that the right-hand-side of (44) is bounded by 3 multiplied by

$$\mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 \left(e_i^2 + r_{Ki}^2 + r_{Ki}^{*2}\right) 1\left(\left(\alpha_K' x_{Ki}\right)^2 \left(e_i^2 + r_{Ki}^2 + r_{Ki}^{*2}\right) > 3\varepsilon n\right)\right]. \tag{46}$$

The inequality

$$\left(\sum_{j=1}^J a_j\right) 1\left(\sum_{j=1}^J a_j > b\right) \leq 2\sum_{j=1}^J a_j 1\left(a_j > b/J\right)$$

for $a_j \geq 0$ shows that (46) is bounded by 2 times

$$\mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 e_i^2 1\left(\left(\alpha_K' x_{Ki}\right)^2 e_i^2 > \varepsilon n\right)\right] \tag{47}$$

$$+\,\mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 r_{Ki}^2 1\left(\left(\alpha_K' x_{Ki}\right)^2 r_{Ki}^2 > \varepsilon n\right)\right] \tag{48}$$

$$+\,\mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 r_{Ki}^{*2} 1\left(\left(\alpha_K' x_{Ki}\right)^2 r_{Ki}^{*2} > \varepsilon n\right)\right]. \tag{49}$$

The proof is completed by showing that (47)-(49) are $o_p(1)$.

First, we show (47) supposing that Assumption 1.2(a) holds. Using (43), Lemma 1, Holder's inequality, Assumption 1.1, and $\|e_i\|_p \leq D$ for some $D < \infty$ by Assumption 1.2(a), (47) is bounded by

$$\begin{aligned} C^{-1}\mathbb{E}\left[\|x_{Ki}\|^2 e_i^2 1\left(\|x_{Ki}\|^2 e_i^2 > \varepsilon n C\right)\right] &\leq \frac{\mathbb{E}\left(\|x_{Ki}\|^t |e_i|^t\right)}{C\left(\varepsilon n C\right)^{(t-2)/2}} \\ &\leq \frac{\|x_{Ki}\|_q^t \|e_i\|_p^t}{C\left(\varepsilon n C\right)^{(t-2)/2}} \\ &\leq \frac{D^t}{C^{t/2}\varepsilon^{(t-2)/2}}\left(\frac{\zeta_K^{2t/(t-2)}}{n}\right)^{(t-2)/2} \\ &= o(1). \end{aligned} \tag{50}$$

the final bound by Assumption 1.4. Hence (47) is $o_p(1)$.

Now consider (47) supposing instead that Assumption 1.2(b) holds. Define

$$\sigma^2(B) = \sup_x \mathbb{E}\left[e_i^2 1\left(e_i^2 > B\right) | x_i = x\right] \tag{51}$$

which satisfies $\sigma^2(B) \to 0$ as $B \to \infty$. Set $b_K = \zeta_K^{2q/(q-2)} \log K$. Then (47) equals

$$\mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 e_i^2 1\left(\left(\alpha_K' x_{Ki}\right)^2 e_i^2 > \varepsilon n\right) 1\left(\|x_{Ki}\|^2 \le b_K\right)\right] \tag{52}$$
$$+\mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 e_i^2 1\left(\left(\alpha_K' x_{Ki}\right)^2 e_i^2 > \varepsilon n\right) 1\left(\|x_{Ki}\|^2 > b_K\right)\right]. \tag{53}$$

Using (51), the first term (52) is bounded by

$$\begin{aligned}
\mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 e_i^2 1\left(e_i^2 > C\varepsilon n/b_K\right)\right] &= \mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 \mathbb{E}\left(e_i^2 1\left(e_i^2 > \frac{C\varepsilon n}{b_K}\right)|z_i\right)\right] \\
&\le \mathbb{E}\left[\left(\alpha_K' x_{Ki}\right)^2 \sigma^2\left(\frac{C\varepsilon n}{b_K}\right)\right] \\
&= \|\alpha_K\|^2 \sigma^2\left(\frac{C\varepsilon n}{b_K}\right) \\
&\le o(1)
\end{aligned}$$

since $\mathbb{E}\left(\alpha_K' x_{Ki}\right)^2 = \alpha_K' \mathbb{E}\left(x_{Ki} x_{Ki}'\right) \alpha_K = \|\alpha_K\|^2$. The final inequality holds since Assumption 1.4 implies $b_K/n = n^{-1}\zeta_K^{2q/(q-2)} \log K = o(1)$, and $\|\alpha_K\|^2 \le C^{-1}$. Since $\mathbb{E}\left(e_i^2|z_i\right) \le \sigma^2(0) < \infty$ and using Lemma 1, the second term (53) is bounded by

$$\begin{aligned}
C^{-1}\mathbb{E}\left(\|x_{Ki}\|^2 e_i^2 1\left(\|x_{Ki}\|^2 > b_K\right)\right) &= C^{-1}\mathbb{E}\left(\|x_{Ki}\|^2 1\left(\|x_{Ki}\|^2 > b_K\right)\mathbb{E}\left(e_i^2|z_i\right)\right) \\
&\le C^{-1}\sigma^2(0)\mathbb{E}\left(\|x_{Ki}\|^2 1\left(\|x_{Ki}\|^2 > b_K\right)\right) \\
&\le C^{-1}\sigma^2(0)\frac{\zeta_K^q}{b_K^{(q-2)/2}} \\
&= \frac{\sigma^2(0)}{C\left(\log K\right)^{(q-2)/2}} \to 0.
\end{aligned}$$

Together, these show that (47)=(52)+(53) is $o_p(1)$, and thus under either Assumption 1.2(a) or (b), (47) is $o_p(1)$.

Second, take (48). Using (43), Lemma 1, Holder's inequality, Assumption 1.1, and Lemma 2.2, (48) is bounded by

$$\begin{aligned}
C^{-1}\mathbb{E}\left[\|x_{Ki}\|^2 r_{Ki}^2 1\left(\|x_{Ki}\|^2 r_{Ki}^2 > \varepsilon n C\right)\right] &\le \frac{\mathbb{E}\left(\|x_{Ki}\|^t |r_{Ki}|^t\right)}{C\left(\varepsilon n C\right)^{(t-2)/2}} \\
&\le \frac{1}{C^{t/2}\varepsilon^{(t-2)/2}} \frac{\|x_{Ki}\|_q^t \|r_{Ki}\|_p^t}{n^{(t-2)/2}} \\
&\le \frac{1}{C^{t/2}\varepsilon^{(t-2)/2}} \left(\frac{\zeta_K^{2t/(t-2)}}{n}\right)^{(t-2)/2} \delta_K^t \\
&\le o(1)
\end{aligned} \tag{54}$$

where the final inequality is Assumption 1.4 and $\delta_K = O(1)$. Thus (48) is $o_p(1)$.

Now take (49). Using (43), Lemma 1, Lemma 2.3 and Assumption 1.1, (49) is bounded by

$$\begin{aligned} C^{-1}\mathbb{E}\left[\|x_{Ki}\|^2 r_{Ki}^{*2} 1\left(\|x_{Ki}\|^2 r_{Ki}^{*2} > \varepsilon n C_2\right)\right] &\le \frac{\mathbb{E}\left(\|x_{Ki}\|^{q/2}|r_{Ki}^*|^{q/2}\right)}{C(\varepsilon n C)^{(q-4)/4}} \\ &\le \frac{1}{C^{q/4}\varepsilon^{(q-4)/4}} \frac{\mathbb{E}\|x_{Ki}\|^q \delta_K^{q/2}}{n^{(q-4)/4}} \\ &\le \frac{1}{C^{q/4}\varepsilon^{(q-4)/4}} \left(\frac{\left(\zeta_K^2\delta_K\right)^{2q/(q-4)}}{n}\right)^{(q-4)/4} \\ &= o(1) \end{aligned}$$

the final inequality by Assumption 1.5. Thus (49) is $o_p(1)$.

We have shown that (44) is bounded by 6 times the sum of (47)-(49) which are $o_p(1)$. This establishes Lindeberg's condition, completing the proof. ■

Lemma 4. *Under Assumption 1 and $Q_K = I_K$,*

$$A_K'\left(\widehat{Q}_K^{-1} - I_K\right) n^{-1/2} X_K' e_K = o_p(1)$$

Proof: Using (34),

$$A_K'\left(\widehat{Q}_K^{-1} - I_K\right) n^{-1/2} X_K' e_K = R_{1n} + R_{2n} \tag{55}$$

where $R_{1n} = A_K'\left(\widehat{Q}_K^{-1} - I_K\right) T_{1n}$ and $R_{2n} = A_K'\left(\widehat{Q}_K^{-1} - I_K\right) T_{2n}$.

First take R_{1n}. We use four algebraic inequalities: (1) that for any matrix A, $\|A\|^2 \le \operatorname{tr} AA'$, (2) for any $K \times d$ matrix A, any conformable matrix B and symmetric matrix C

$$\operatorname{tr}\left(A'B'CBA\right) = \operatorname{tr}\left(AA'B'CB\right) \le \operatorname{tr}\left(AA'\right)\left\|B'CB\right\| \le d\|A\|^2\|B\|^2\|C\|,$$

(3) for any $d \times d$ positive semi-definite matrix A, $\operatorname{tr} A \le d\|A\|$, and (4) the norm inequality

$\|AB\| \leq \|A\| \, \|B\|$. Using these results, conditioning, Lemma 2.1, Lemma 2.4, and Lemma 2.6,

$$
\begin{aligned}
\mathbb{E}\left[\|R_{1n}\|^2 |Z\right] &\leq \mathbb{E}\left[\operatorname{tr} R_{1n}R_{1n}'|Z\right] \\
&= \operatorname{tr}\left[A_K'\left(\widehat{Q}_K^{-1} - I_K\right)\mathbb{E}\left(T_{1n}T_{1n}'|Z\right)\left(\widehat{Q}_K^{-1} - I_K\right)A_K\right] \\
&\leq d\left\|A_K'\left(\widehat{Q}_K^{-1} - I_K\right)\mathbb{E}\left(T_{1n}T_{1n}'|Z\right)\left(\widehat{Q}_K^{-1} - I_K\right)A_K\right\| \\
&\leq d\,\|A_K\|^2\left\|\widehat{Q}_K^{-1} - I_K\right\|^2\left\|\mathbb{E}\left(T_{1n}T_{1n}'|Z\right)\right\| \\
&\leq \frac{d}{C}O_p\left(\frac{\zeta_K^{2q/(q-2)}\log K}{n}\right)O_p(\zeta_K^{4/(p-2)}) \\
&\leq O_p\left(\frac{\zeta_K^{2t/(t-2)}\log K}{n}\right) \\
&= o_p(1). \qquad (56)
\end{aligned}
$$

The fifth inequality in (56) holds since

$$
\frac{2q}{q-2} + \frac{4}{p-2} = \frac{2qp-8}{qp-2p-2q+4} \leq \frac{2qp}{qp-2p-2q} = \frac{2t}{t-2} \qquad (57)
$$

and the final bound in (56) is Assumption 1.4.

Markov's inequality implies that for any $\varepsilon > 0$

$$
p_n = P\left(\|R_{1n}\| > \varepsilon|Z\right) \leq \frac{\mathbb{E}\left[\|R_{1n}\|^2 |Z\right]}{\varepsilon^2} = o_p(1)
$$

and since p_n is bounded, it follows that

$$
P\left(\|R_{1n}\| > \varepsilon\right) = \mathbb{E}p_n \to 0
$$

and hence $R_{1n} = o_p(1)$.

Now take R_{2n}. By the norm inequality, Lemma 2.1, Lemma 2.4 and Lemma 2.9,

$$
\begin{aligned}
\|R_{2n}\| &\leq \|A_K\|\left\|\widehat{Q}_K^{-1} - Q_K^{-1}\right\|\|T_{2n}\| \\
&\leq O_p\left(\sqrt{\frac{\zeta_K^{2q/(q-2)}\log K}{n}}\right)O_p\left(\zeta_K^{(4-q)/(q-2)}\left(\frac{\log K}{n}\right)^{2/q-1/2}\right) \\
&= O_p\left(\left(\frac{\zeta_K^{2q/(q-2)}\log K}{n}\right)^{2/q}\right) \\
&= o_p(1)
\end{aligned}
$$

where the final bound is implied by Assumption 1.4 and $q \geq t$. Thus (55) is $o_p(1)$ as stated. ■

10 Covariance Matrix Estimation

In this section we establish (29), completing the proofs of Theorems 3 and 4. The main difficulty is handling the presence of the OLS residuals. Existing theory has dealt with this issue by imposing sufficient conditions to ensure uniform convergence of the regression function estimate, so that the residuals are uniformly close to the true errors and thus their substitution is asymptotically negligible. This approach requires substantially more restrictive assumptions. We avoid these restrictions by instead writing out the covariance matrix estimators explicitly without using uniform convergence bounds.

Lemma 5. *Under Assumption 1 and* $Q_K = I_k$*,*

$$\left\| \widehat{V}_{AK} - I_K \right\| = o_p(1) \tag{58}$$

and for any r

$$\left\| \widehat{V}_{AK}^{r} - I_K \right\| = o_p(1) \tag{59}$$

In particular, (59) with $r = -1/2$ is (29), which is required to complete the proofs of Theorems 3 and 4.

Proof: Since (59) follows directly from (58) by Theorem 2, the main difficult is in establishing (58).

By the triangle inequality

$$\left\| \widehat{V}_{AK} - I_K \right\| \leq \left\| A_K' \widehat{Q}_K^{-1} \left(\widehat{S}_K - \widetilde{S}_K \right) \widehat{Q}_K^{-1} A_K \right\| \tag{60}$$

$$+ \left\| A_K' \widehat{Q}_K^{-1} \widetilde{S}_K \widehat{Q}_K^{-1} A_K - I_K \right\|. \tag{61}$$

Using $\widehat{e}_{Ki}^2 = e_{Ki}^2 - 2 e_{Ki} x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) + \left(x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) \right)^2$ and the triangle inequality, the first term (60) equals

$$
\begin{aligned}
&\left\| A_K' \widehat{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' \left(\widehat{e}_{Ki}^2 - e_{Ki}^2 \right) \widehat{Q}_K^{-1} A_K \right\| \\
&\leq 2 \left\| A_K' \widehat{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' e_{Ki} x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) \widehat{Q}_K^{-1} A_K \right\| \\
&+ \left\| A_K' \widehat{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' \left(x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) \right)^2 \widehat{Q}_K^{-1} A_K \right\| \\
&\leq 2 \left[\left\| A_K' \widehat{Q}_K^{-1} \widetilde{S}_K \widehat{Q}_K^{-1} A_K - I_K \right\| + 1 \right]^{1/2} \\
&\times \left\| A_K' \widehat{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' \left(x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) \right)^2 \widehat{Q}_K^{-1} A_K \right\|^{1/2} \\
&+ \left\| A_K' \widehat{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' \left(x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) \right)^2 \widehat{Q}_K^{-1} A_K \right\|.
\end{aligned}
$$

The final inequality uses the norm inequality $\|X_1' X_2\| \leq \|X_1\| \|X_2\| = \|X_1' X_1\|^{1/2} \|X_2' X_2\|^{1/2}$ applied to the $n \times K$ matrices X_1 and X_2 whose i^{th} rows are $x_{Ki}' e_{Ki} \widehat{Q}_K^{-1} A_K$ and $x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) x_{Ki}' \widehat{Q}_K^{-1} A_K$, respectively.

The second term (61) equals

$$
\begin{aligned}
\left\| A_K' \widehat{Q}_K^{-1} \widetilde{S}_K \widehat{Q}_K^{-1} A_K - I_K \right\| &\leq \left\| A_K' \widetilde{S}_K A_K - I_K \right\| + 2 \left\| A_K' \left(\widehat{Q}_K^{-1} - I_k \right) \widetilde{S}_K A_K \right\| \\
&+ \left\| A_K' \left(\widehat{Q}_K^{-1} - I_k \right) \widetilde{S}_K \left(\widehat{Q}_K^{-1} - I_k \right) A_K \right\| \\
&\leq \left\| A_K' \widetilde{S}_K A_K - I_K \right\| \\
&+ 2 \left\| A_K' \left(\widehat{Q}_K^{-1} - I_k \right) \widetilde{S}_K \left(\widehat{Q}_K^{-1} - I_k \right) A_K \right\|^{1/2} \left[\left\| A_K' \widetilde{S}_K A_K - I_K \right\| + 1 \right]^{1/2} \\
&+ \left\| A_K' \left(\widehat{Q}_K^{-1} - I_k \right) \widetilde{S}_K \left(\widehat{Q}_K^{-1} - I_k \right) A_K \right\|.
\end{aligned}
$$

To establish that (60)-(61) is $o_p(1)$ it is sufficient to show the following three inequalities

$$
\left\| A_K' \widetilde{S}_K A_K - I_K \right\| = o_p(1) \tag{62}
$$

$$
\left\| A_K' \widehat{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^n x_{Ki} x_{Ki}' \left(x_{Ki}' \left(\widehat{\beta}_K - \beta_K \right) \right)^2 \widehat{Q}_K^{-1} A_K \right\| = o_p(1) \tag{63}
$$

$$
\left\| A_K' \left(\widehat{Q}_K^{-1} - I_k \right) \widetilde{S}_K \left(\widehat{Q}_K^{-1} - I_k \right) A_K \right\| = o_p(1). \tag{64}
$$

Equation (62) is Lemma 2.8.

We next show (63). Using

$$\widehat{\beta}_K - \beta_K = \widehat{Q}_K^{-1}\frac{1}{n}\sum_{i=1}^{n} x_{Ki}e_{Ki} = n^{-1/2}\widehat{Q}_K^{-1}\left(T_{1n}+T_{2n}\right)$$

and the C_r inequality, we find

$$\begin{aligned}\left(x_{Ki}'\left(\widehat{\beta}_K-\beta_K\right)\right)^2 &= \frac{1}{n}\left(x_{Ki}'\widehat{Q}_K^{-1}\left(T_{1n}+T_{2n}\right)\right)^2 \\ &\leq \frac{2}{n}\left(x_{Ki}'\widehat{Q}_K^{-1}T_{1n}\right)^2 + \frac{2}{n}\left(x_{Ki}'\widehat{Q}_K^{-1}T_{2n}\right)^2.\end{aligned}$$

Thus (63) is bounded by 2 multiplied by

$$\left\|A_K'\widehat{Q}_K^{-1}\frac{1}{n^2}\sum_{i=1}^{n} x_{Ki}x_{Ki}'\left(x_{Ki}'\widehat{Q}_K^{-1}T_{1n}\right)^2\widehat{Q}_K^{-1}A_K\right\| \tag{65}$$

$$+\left\|A_K'\widehat{Q}_K^{-1}\frac{1}{n^2}\sum_{i=1}^{n} x_{Ki}x_{Ki}'\left(x_{Ki}'\widehat{Q}_K^{-1}T_{2n}\right)^2\widehat{Q}_K^{-1}A_K\right\|. \tag{66}$$

Take (65). By conditioning, Lemma 2.4, Lemma 2.6, and Lemma 2.10

$$\begin{aligned}&\mathbb{E}\left[\left\|A_K'\widehat{Q}_K^{-1}\frac{1}{n^2}\sum_{i=1}^{n} x_{Ki}x_{Ki}'\left(x_{Ki}'\widehat{Q}_K^{-1}T_{1n}\right)^2\widehat{Q}_K^{-1}A_K\right\| |Z\right] \\ &\leq \operatorname{tr}\left[A_K'\widehat{Q}_K^{-1}\frac{1}{n^2}\sum_{i=1}^{n} x_{Ki}x_{Ki}'x_{Ki}'\widehat{Q}_K^{-1}\mathbb{E}\left(T_{1n}T_{1n}'|Z\right)\widehat{Q}_K^{-1}x_{Ki}\widehat{Q}_K^{-1}A_K\right] \\ &\leq \operatorname{tr}\left[A_K'\widehat{Q}_K^{-1}\frac{1}{n^2}\sum_{i=1}^{n} x_{Ki}x_{Ki}'\left\|x_{Ki}\right\|^2\widehat{Q}_K^{-1}A_K\right]\left\|\widehat{Q}_K^{-1}\right\|^2\left\|\mathbb{E}\left(T_{1n}T_{1n}'|Z\right)\right\| \\ &\leq d\left\|A_K\right\|^2\left\|\frac{1}{n^2}\sum_{i=1}^{n} x_{Ki}x_{Ki}'\left\|x_{Ki}\right\|^2\right\| O_p\left(\zeta_K^{4/(p-2)}\right) \\ &\leq O_p\left(\frac{\zeta_K^{2q/(q-2)}}{n}\right)O_p\left(\zeta_K^{4/(p-2)}\right) \\ &\leq O_p\left(\frac{\zeta_K^{2t/(t-2)}}{n}\right) \\ &= o_p(1)\end{aligned}$$

where the second-to-last inequality holds under (57). As discussed in the proof of Lemma 4, this implies that (65) is $o_p(1)$.

Similarly for (66), using Lemma 2.4, Lemma 2.8, and Lemma 2.10

$$\begin{aligned}
&\left\| A_K' \widehat{Q}_K^{-1} \frac{1}{n^2} \sum_{i=1}^{n} x_{Ki} x_{Ki}' \left(x_{Ki}' \widehat{Q}_K^{-1} T_{2n} \right)^2 \widehat{Q}_K^{-1} A_K \right\| \\
&\leq d \left\| A_K \right\|^2 \left\| \frac{1}{n^2} \sum_{i=1}^{n} x_{Ki} x_{Ki}' \left\| x_{Ki} \right\|^2 \right\| \left\| \widehat{Q}_K^{-1} \right\|^4 \left\| T_{2n} \right\|^2 \\
&\leq O_p\left(\frac{\zeta_K^{2q/(q-2)}}{n} \right) O_p\left(\left(\frac{\zeta_K^{2q/(q-2)} \log K}{n} \right)^{-(q-4)/2q} \right) \\
&\leq O_p\left(\left(\frac{\zeta_K^{2q/(q-2)}}{n} \right)^{(q+4)/2q} \right) \\
&\leq o_p(1).
\end{aligned}$$

This establishes (63).

To complete the proof we establish (64). By the matrix norm inequality, Lemma 2.1, Lemma 2.4, and Lemma 2.7

$$\begin{aligned}
\left\| A_K' \left(\widehat{Q}_K^{-1} - I_k \right) \widetilde{S}_K \left(\widehat{Q}_K^{-1} - I_k \right) A_K \right\| &\leq \left\| A_K \right\|^2 \left\| \widehat{Q}_K^{-1} - I_k \right\|^2 \left\| \widetilde{S}_K \right\| \\
&\leq O_p\left(\frac{\zeta_K^{2q/(q-2)} \log K}{n} \right) O_p\left(\zeta_K^{4/(p-2)} \right) \\
&\leq O_p\left(\frac{\zeta_K^{2t/(t-2)} \log K}{n} \right) \\
&= o_p(1)
\end{aligned}$$

again by (57) and Assumption 1.4. This is (64) as needed. ■

11 Proof for Spline Approximation Theory

Proof of Theorem 7. (Sketch) We show that for a properly constructed weight function $w(z)$,

$$\inf_{\beta \in \mathbb{R}^K} \sup_{z \in \mathcal{Z}} \left| \frac{g(z) - x_K(z)'\beta}{w(z)} \right| \leq C K^{-s(1-1/\alpha)} \tag{67}$$

and $\left\| w(z_i) \right\|_p \leq C$. The result follows via the discussion of weighted sup norms.

Recall that $b = BK^{1/\alpha}$. Let β_K^* be the coefficients of the best uniform spline approximation on the interval $[-b, b]$ with the coefficients for the intervals I_0 and I_{K+1} set to zero.

$$\beta_K^* = \underset{\beta \in \mathbb{R}^K}{\operatorname{argmin}} \sup_{|z| \leq b} \left| g(z) - x_K(z)'\beta \right|.$$

Set $g_K(z) = x_K(z)'\beta_K^*$. Since $\sup_z \left|g^{(s)}(z)\right| \leq C$, then by the standard approximation properties of splines (e.g. Corollary 6.21 of Schumaker (2007))

$$\sup_{|z|\leq b} |g(z) - g_K(z)| \leq C\left(\frac{b}{K}\right)^s = CB^s K^{-s(1-1/\alpha)}.$$

Furthermore, for $r = 0, 1, ..., s$

$$\sup_{|z|\leq b} \left|g^{(r)}(z) - g_K^{(r)}(z)\right| \leq C_1\left(\frac{b}{K}\right)^{s-r} = C_1 B^s K^{-(s-r)(1-1/\alpha)} \leq \varepsilon$$

where the final inequality is for sufficiently large K. This suggests that each segment of $g_K(z)$ in the interval $[-b, b]$ is an s^{th} order polynomial with coefficients bounded across segments. The assumption that $\sup_z \left|g^{(s-1)}(z)\right| \leq C$ also implies the function $g(z)$ can be globally bounded by a s^{th} order polynomial. Together, this means that we can globally bound $g(z)$ and each segment of $g_K(z)$ in the interval $[-b, b]$ by a common s^{th} order polynomial $\overline{g}(z) = \sum_{j=0}^{s-1} a_j |z|^j$. Since the coefficients of $g_K(z)$ on the segment I_0 equals the coefficients on I_1 and similarly the coefficients on I_K and I_{K+1} coincide, it follows the the polynomial coefficients for the segments I_0 and I_{K+1} are bounded by $\overline{g}(z)$ as well.

Now set $w(z) = \overline{g}(z) |z|^{s(\alpha-1)}$. Then

$$\begin{aligned}
\inf_{\beta\in\mathbb{R}^K} \sup_{z\in\mathcal{Z}} \left|\frac{g(z) - x_K(z)'\beta}{w(z)}\right| &\leq \sup_{z\in\mathcal{Z}} \left|\frac{g(z) - x_K(z)'\beta_K^*}{w(z)}\right| \\
&\leq \sup_{|z|\leq b} |w(z)|^{-1} CB^s K^{-s(1-1/\alpha)} + 2\sup_{|z|>b} |z|^{-s(\alpha-1)} \\
&\leq \left(a_0^{-1} CB^s + 2B^{-s(\alpha-1)}\right) K^{-s(1-1/\alpha)}.
\end{aligned}$$

This is (67). Finally

$$\|w(z_i)\|_p \leq \sum_{j=0}^{s-1} a_j \left(\mathbb{E} |z_i|^{(j+s(\alpha-1))p}\right)^{1/p} \leq \sum_{j=0}^{s-1} a_j C$$

is finite, as needed. ■

References

[1] Andrews, Donald W. K. (1991): "Asymptotic normality of series estimators for nonparametric and semiparametric models," *Econometrica,* 59, 307-345.

[2] Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato (2014): "Some new asymptotic theory for least squares series: Pointwise and uniform results," working paper.

[3] Chen, Xiaohong and Timothy M. Christensen (2015): "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions," *Journal of Econometrics*, 188, 447-465.

[4] Chen, Xiaohong, Han Hong, and Elie Tamer (2005): "Measurement error with auxiliary data," *Review of Economic Studies*, 72, 343-366.

[5] Chen, Xiaohong and Zhipeng Liao (2014): "Sieve M inference on irregular parameters," working paper.

[6] Chen, Xiaohong, Zhipeng Liao, and Yixiao Sun (2014): "Sieve inference on possibly misspecified semi-nonparametric time series models," *Journal of Econometrics*, 178, 639-658.

[7] Chen, Xiaohong and Xiaotong Shen (1998): "Sieve extremum estimates for weakly dependent data," *Econometrica,* 66, 289-314.

[8] Huang, Jianhua A. (2003): "Local asymptotics for polynomial spline regression," *The Annals of Statistics*, 31, 1600-1635.

[9] Li, Qi, and Jeffrey S. Racine (2006): *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

[10] Maioriv, Vitaly, and Ron Meir (1998) "Approximation bounds for smooth functions in $C(\mathbb{R}^d)$ by neural and mixture networks," *IEEE Transactions on Neural Networks*, 9, 969-978.

[11] Mhaskar, H. N. (1996): *Introduction to the Theory of Weighted Polynomial Approximation*, Singapore: World Scientific.

[12] Newey, Whitney K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147-168.

[13] Rudelson, M. (1999): "Random vectors in the isotropic position," *Journal of Functional Analysis*, 164, 60-72.

[14] Schumaker, Larry L. (2007): *Spline Functions: Basic Theory, Third Edition*, Cabridge: Cambridge University Press

[15] Triebel, Hans (2006): *Theorem of Function Spaces III*, Boston: Birkhauser.