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

Bruce E. Hansen*
University of Wisconsin†

March 2016
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 nonparametric 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 negligibility 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. (2015), 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. (2015) also consider
uniform asymptotic approximations (while the results in this paper are only pointwise).

A word on notation. Let \( \lambda_{\text{min}}(A) \) and \( \lambda_{\text{max}}(A) \) denote the smallest and largest eigenvalue of a positive semidefinite matrix \( A \). Let \( \|A\| = (\lambda_{\text{max}}(A'A))^{1/2} \) denote the spectral norm of a matrix \( A \). Note that if \( A \succeq 0 \) then \( \|A\| = \lambda_{\text{max}}(A) \). When applied to a \( K \times 1 \) vector let \( \|a\| = (\mathbb{E}\|a\|^p)^{1/p} \) 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 \( g(z) = \mathbb{E}(y_i | z_i = z) \), the regression error \( e_i = y_i - g(z_i) \), the conditional variance \( \sigma^2(z) = \mathbb{E}(e_i^2 | z_i = z) \), 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 \( x_K(z) = (1, z, ..., z^K) \). 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 \( \beta_K \). 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

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

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 implies 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 \( x_{Ki} \):

\[
\hat{\beta}_K = \left( \sum_{i=1}^{n} x_{Ki}x_{Ki}' \right)^{-1} \sum_{i=1}^{n} x_{Ki}y_i.
\] (5)

The corresponding estimator of \( g(z) \) is

\[
\hat{g}_K(z) = x_K(z)'\hat{\beta}_K
\]

and that of \( \theta \) is

\[
\hat{\theta}_K = a(\hat{g}_K) = a'_K\hat{\beta}_K.
\] (6)

Define the covariance matrices

\[
Q_K = \mathbb{E}(x_{Ki}x_{Ki}')
\]

\[
S_K = \mathbb{E}(x_{Ki}x_{Ki}'e^2_{Ki})
\]

\[
V_K = Q^{-1}_K S_K Q^{-1}_K
\]

\[
V_{\theta K} = a'_{K} V_{K} a_{K}.
\]

Thus for fixed \( K \), \( V_K \) and \( V_{\theta K} \) are the conventional asymptotic covariance matrices for the estimators \( \hat{\beta}_K \) and \( \hat{\theta}_K \).

The standard estimators for the above covariance matrices are

\[
\hat{Q}_K = \frac{1}{n} \sum_{i=1}^{n} x_{Ki}x_{Ki}'
\]

\[
\hat{S}_K = \frac{1}{n} \sum_{i=1}^{n} x_{Ki}x_{Ki}'e^2_{Ki}
\]

\[
\hat{V}_K = \hat{Q}^{-1}_K \hat{S}_K \hat{Q}^{-1}_K
\]

\[
\hat{V}_{\theta K} = a'_{K} \hat{V}_{K} a_{K}
\]

where \( e_{Ki} = y_i - x'_{Ki}\hat{\beta}_K \) are the OLS residuals.

We are interested in obtaining distributional approximations for the normalized coefficient estimate \( \hat{\beta}_K \), the parameter estimate \( \hat{\theta}_K \), and for test statistics constructed using the sample covariance matrix \( \hat{V}_{\theta K} \).

3 Matrix Convergence Theory

One important component for developing a distribution theory for the estimates and test statistics is convergence theory for the sample design matrix \( \hat{Q}_K \), its inverse \( \hat{Q}^{-1}_K \), and the central part of the covariance matrix estimate \( \hat{S}_K \). The technical challenge is dealing with matrix dimensions which change with sample size, without imposing arbitrary conditions such as boundedness.
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 \( u_{ni}, i = 1, ..., n \) be an array of mutually independent \( L(n) \times 1 \) vectors. Define \( \tilde{M}_n = \frac{1}{n} \sum_{i=1}^{n} u_{ni} u_{ni}' \).

**Theorem 1.** If \( \| E\tilde{M}_n \| \leq \mu_n \), and for all \( \varepsilon > 0 \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E \left( \| u_{ni} \|^2 1 \left( \| u_{ni} \|^2 > \varepsilon \frac{n}{\log (2L(n))} \right) \right) = 0,
\]

then

\[
E \left\| \tilde{M}_n - E\tilde{M}_n \right\| = o \left( \sqrt{1 + \mu_n} \right).
\]

In particular, if \( \mu_n \leq \mu < \infty \) then

\[
E \left\| \tilde{M}_n - E\tilde{M}_n \right\| = o(1)
\]

and

\[
\left\| \tilde{M}_n - E\tilde{M}_n \right\| = o_p(1).
\]

**Corollary 1.** Equation (7) holds under any of the following conditions

1. \( L \) is fixed, \( u_{ni} = u_i \) is i.i.d, and \( \mathbb{E} \| u_i \|^2 < \infty \)

2. \( L \) is fixed, and \( u_{ni} \) is uniformly square integrable: \( \lim_{B \to \infty} \sup_n \sup_{i \leq n} \mathbb{E} \left( \| u_{ni} \|^2 1 \left( \| u_{ni} \|^2 > B \right) \right) = 0 \)

3. \( \| u_{ni} \| \leq \xi_n \) and \( n^{-1} \xi_n^2 \log(2L(n)) = o(1) \)

4. For some \( s > 2 \), \( \| u_{ni} \|_s \leq \xi_n \) and \( n^{-1} \xi_n^{2s/(s-2)} \log(2L(n)) = o(1) \)

Theorem 1 gives conditions for convergence in probability, but does not give a rate of convergence. Our next result provides such a rate.

**Theorem 2.** If \( \| E\tilde{M}_n \| \leq \mu_n \), and for some \( s \geq 4 \),

\[
\| u_{ni} \|_s \leq \xi_n
\]

and

\[
\frac{\xi_n^{2s/(s-2)} \left( \log (2L(n)) \right)^{(s-4)/(s-2)}}{n} = o(1),
\]

then

\[
E \left\| \tilde{M}_n - E\tilde{M}_n \right\| = O \left( \sqrt{\frac{\xi_n^{2s/(s-2)} \left( \log (2L(n)) \right) \left( 1 + \mu_n \right)^{(s-4)/(s-2)}}{n}} \right).
\]
The proofs of Theorems 1 and 2 extend Lemma 6.2 of Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Theorem 1 of Rudelson (1999) which treat the case of the deterministic bound of Corollary 1.3. Standard convergence theory allowing for $L(n) \to \infty$ imposes such bounds. When $u_{ni}$ are basis transformation of a random variable $z_i$ with bounded support then the rates for $\xi_n$ are known for common sieve transformations, for example, $\xi_n = O(L)$ for power series and $\xi_n = O(L^{1/2})$ for splines. In these cases we can set $s = \infty$ and (10) simplifies to $n^{-1} \xi_n^2 \log 2L = o(1)$, which is identical to that obtained by Lemma 6.2 of Belloni, Chernozhukov, Chetverikov, and Kato (2015) and the in-probability analog of Lemma 2.1 of Chen and Christensen (2015).

The moment bound (9) is an important generalization beyond bounded variables, and in particular allows for variables with unbounded support. An example for (9) is when the variables have a uniformly bounded $s^{th}$ moment. Suppose that $u_{ni} = (u_{1i}, \ldots, u_{Ki})'$ and $\|u_{ji}\|_s \leq C < \infty$. Then an application of Minkowski’s inequality implies (9) with $\xi_n = CL(n)^{1/2}$. This is the same rate as obtained, for example, by splines with bounded variables.

The rate (10) shows that there is a trade-off between the number of finite moments and the allowable rate of divergence of the dimension $L$. As $s \downarrow 2$ the allowable rate (10) slows to fixed $L$, and as $s \to \infty$, the allowable rate increases. The moment bounds (8) and (11) imply that $\left\| \hat{M}_n - \mathbb{E} \hat{M}_n \right\| = O_p(a_n)$, with $a_n$ a function of the rate in (10) when $s \geq 4$. The difference when $s \geq 4$ is that the latter calculation involves the variance of the sample mean $\hat{M}_n$ (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 (8) and (11) depend on the norm $\mu_n$ of the covariance matrix $\mathbb{E}u_{ni}u_{ni}'$. In most applications this is either assumed bounded or the bounds (8) and (11) are applied to orthogonalized sequences so $\mu_n$ is bounded and can be omitted from these expressions.

It is useful to observe that Theorem 1 applies when the dimension $L$ of $u_{ni}$ is allowed to be fixed with $n$, growing, or varying. If $L = L(n)$ is bounded, then the term $\log (2L)$ can be omitted from (10) and (11).

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 3.** Let $\hat{M}_n$ be a sequence of positive definite $L(n) \times L(n)$ matrices. If $\left\| \hat{M}_n - I_{L(n)} \right\| = O_p(a_n)$, and $a_n = o(1)$, then for any $r \in \mathbb{R}$,

$$\left\| \hat{M}_n^r - I_{L(n)} \right\| = O_p(a_n).$$

Given Theorems 1-3, we can immediately develop an appropriate convergence theory for the sample design matrix $\hat{Q}_K$ and its inverse. As is conventional in the sieve literature, it is convenient to rotate the regressors so that they are orthogonal and thus the norm expected design matrix is
bounded. In particular, we define the orthogonalized regressor sequence

$$\tilde{x}_{Ki} = Q_K^{-1/2} x_{Ki}$$

and the orthogonalized sample design matrix

$$\tilde{Q}_K = Q_K^{-1/2} \tilde{Q}_K Q_K^{-1/2} = \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{Ki} \tilde{x}'_{Ki}.$$ 

By construction, $\mathbb{E}(\tilde{x}_{Ki}\tilde{x}'_{Ki}) = I_K$ so $\|\mathbb{E}(\tilde{x}_{Ki}\tilde{x}'_{Ki})\| = 1$.

**Assumption 1.** For some $4 \leq q \leq \infty$

1. $\frac{\|x_{Ki}\|}{\sqrt{\lambda_{\min}(Q_K)}} \leq \zeta_K$

2. $n^{-1} \zeta_K^{2q/(q-2)} (\log (2K))^{(q-4)/(q-2)} = o(1)$

**Theorem 4.** Under Assumption 1,

1. $\|\tilde{x}_{Ki}\| \leq \zeta_K$

2. $\mathbb{E} \left\| \tilde{Q}_K - I_K \right\| = O \left( \sqrt{n^{-1} \zeta_K^{2q/(q-2)} (\log (2K))^{(q-4)/(q-2)}} \right)$

3. $\|\tilde{Q}_K^{-1} - I_K\| = O_p \left( \sqrt{n^{-1} \zeta_K^{2q/(q-2)} (\log (2K))^{(q-4)/(q-2)}} \right)$

Assumption 1 states that the regressors are bounded in the $L^q$ norm. The bound $\zeta_K$ 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 bound is invariant to rescaling. It may be useful to observe that $\zeta_K$ cannot increase slower than $K^{1/2}$. To see this, define $\tilde{x}_{Ki} = Q_K^{-1/2} x_{Ki}$ so that $\mathbb{E}(\tilde{x}'_{Ki}\tilde{x}_{Ki}) = K$. Then

$$\frac{\|x_{Ki}\|}{\sqrt{\lambda_{\min}(Q_K)}} \geq \left( \frac{\mathbb{E}(\tilde{x}'_{Ki}Q_K\tilde{x}_{Ki})}{\lambda_{\min}(Q_K)} \right)^{1/2} \geq \left( \mathbb{E}(\tilde{x}'_{Ki}\tilde{x}_{Ki}) \right)^{1/2} \geq K^{1/2}.$$ 

4 **Assumptions**

We now list sets of assumptions which are sufficient for asymptotic normality of the studentized estimators. We start by defining the regression approximation error in the $L^p$ norm. For some $p \geq 2$, consider the $L^p$ norm of the difference between the regression approximation $x'_{Ki}\beta$ and $g(z_i)$

$$\|g(z_i) - x'_{Ki}\beta\|_p = \left( \mathbb{E} |g(z_i) - x'_{Ki}\beta|^p \right)^{1/p}.$$
For $p = \infty$ this is the sup norm (which is conventionally used in the sieve literature). $L^p$ with finite $p$ is weaker, and allows the regression function and regressors to have unbounded support. The best $L^p$ regression approximation to $g(z)$ is obtained by the coefficients

$$
\beta_K^* = \arg\min_{\beta \in \mathbb{R}^K} \| g(z_i) - x'_{ki} \beta \|_p
$$

with resulting approximation error

$$
r_{Ki}^* = g(z_i) - x'_{ki} \beta_K^*.
$$

Our first set of assumptions is based entirely on unconditional moments.

**Assumption 2.** For some $2 < t < \infty$ and $2 \leq p \leq \infty$ such that $1/q + 1/t = 1/\lambda_1 < 1/2$ and $1/q + 1/p = 1/\lambda_2 \leq 1/2$,

1. $\| e_i \|_t \leq D < \infty$
2. $n^{-1} \zeta^2_{Ki}/(\lambda_1 - 2) = o(1)$
3. $\| g(z_i) \|_p < \infty$ and $\| r_{Ki}^* \|_p \leq \delta_K$
4. $\frac{\zeta^2_{Ki}}{n(\lambda_2 - 2)/\lambda_2} = o(1)$
5. $\zeta^4_{Ki} \delta_{Ki}^2 \left( \frac{\log (2K)}{n} \right)^{(q-4)/4} = o(1)$
6. $\lambda_{\min}(\pi'_{K} V_{K} \pi_{K}) \geq C > 0$ where $\pi_{K} = a_{K} \left( a'_{K} Q^{-1}_{K} a_{K} \right)^{-1/2}$ and $a_{K}$ is defined in (4).

Assumption 2.1 states that the regression error has a finite $t^{th}$ moment. A sufficient condition for Assumption 2.1 is $\| y_i \|_t < \infty$.

Assumption 2.2 controls the rate of growth of the number of regressors $K$. The allowable rate is increasing in $t$ and $q$. As $q$ diverges to infinity the rate simplifies to $n^{-1} \zeta^{2p/(p-2)}_{K} = o(1)$ which is similar to the rate obtained by Chen and Christensen (2015) under the assumption of bounded regressors and uniform approximation error bounds. As $p$ diverges to infinity the rate simplifies to $n^{-1} \zeta^{2q/(q-2)}_{K} = o(1)$ which is implied by Assumption 1.2. As both diverge the rate simplifies to the best possible rate $n^{-1} \zeta^{4}_{K} \log (2K) = o(1)$ of Assumption 1.2.

Assumption 2.3 bounds the approximation error $r_{Ki}^*$ in the $L^p$ norm. Explicit rates for the approximation error $\delta_K$ will be derived in Section 6. The assumption that $g(z_i)$ is bounded in the $L^p$ norm is so that this approximation error is well defined.

Assumptions 2.4 and 2.5 control the rate of growth of $\zeta_K$ in relation to the rate of decay of $\delta_K$. If $t = p$ then 2.4 is redundant.

Assumption 2.4 allows $\lambda_2 = 2$ (which occurs when $q = \infty$ and $p = 2$) but then requires $\zeta_{K} \delta_{K} = o(1)$ which can only occur if $K$ is diverging, and is similar to an undersmoothing condition.
Assumption 2.5 allows \( q = 4 \) but then requires \( \zeta_K^2 \delta_K = o(1) \) which requires \( K \) to diverge.

Assumption 2.5 also controls the rate of growth of \( K \), and trades off both the rate of growth in the regressor norms \( \zeta_K \) and approximation errors \( \delta_K \). A simple sufficient condition for Assumption 2.5 is \( p \leq 4 \). (For then \( n^{-1} (\zeta_K^2 \delta_K)^{2q/(q-4)} \log (2K) \leq O \left( n^{-1} \zeta_K^{4q/(q-4)} \log (2K) \right) \leq O \left( n^{-1} \zeta_K^{2q/(q-4)} \log (2K) \right) = o(1) \) by Assumption 2.2.). Another simple condition is \( n^{-1} \zeta_K^{4q/(q-4)} \log K = O(1) \) (a strengthening of Assumption 2.2 which avoids conditions on \( \delta_K \)). A third simple condition is \( \zeta_K \delta_K = O(1) \) and \( p \geq q \) for then \( n^{-1} (\zeta_K^2 \delta_K)^{2q/(q-4)} \log (2K) \leq O \left( n^{-1} \zeta_K^{2q/(q-4)} \log (2K) \right) \leq O \left( n^{-1} \zeta_K^{2q/(q-4)} \log (2K) \right) = o(1) \) by Assumption 2.2. A fourth simple condition is \( \delta_K n^{1/2} = o(1) \), the undersmoothing condition used in the existing nonparametric regression literature, for then

\[
\begin{aligned}
&n^{-1} (\zeta_K^2 \delta_K)^{2q/(q-4)} \log (2K) = o \left( \left( n^{-1} \zeta_K^{2q/(q-2)} \right)^{2(q-2)/(q-4)} \log (2K) \right) = o(1) \text{ by Assumption 2.2.}
\end{aligned}
\]

The undersmoothing condition requires, however, \( K \to \infty \), unlike Assumption 2.5 which does not require \( K \) to diverge.

Assumption 2.6 bounds the asymptotic covariance matrix \( a'_K V_K a_K \) away from singularity. A sufficient (but not necessary) condition for Assumption 2.6 is \( \lambda_{\min} \left( Q_1^{-1/2} S_2 Q_2^{-1/2} \right) \geq C. \) A sufficient condition for the latter is \( \sigma_i^2 \geq \sigma^2 > 0 \), which is the standard assumption in the previous nonparametric sieve literature. Assumption 2.6 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 \( \beta_K \) (those not in \( a'_K \beta_K \)) to converge at a faster than \( n^{-1/2} \). The statement of Assumption 2.6 is not elegant but it is much milder, only requiring that the linear combination \( a'_K \beta_K \) does not converge faster than \( n^{-1/2} \).

An alternative set of assumptions uses a conditional moment bound for the error.

**Assumption 3.** Assumption 2 holds with parts 1 and 2 replaced by

\[
\lim_{B \to \infty} \sup_{z} \mathbb{E} \left[ e_i^2 1 (e_i^2 > B) \mid z_i = z \right] = 0
\]

Assumption 3 states that the regression error satisfies conditional squared uniform integrability, and is slightly stronger than assuming the conditional variance is uniformly bounded. A sufficient condition is that \( e_i \) has a bounded conditional moment slightly higher than 2. This assumption is weaker than, for example, Newey’s (1997) assumption of bounded conditional fourth moments. Neither Assumption 2.1 nor Assumption 3 is strictly stronger or weaker than the other. Assumption 2.1 is an unconditional moment bound, while Assumption 3 is a conditional moment bound.

### 5 Distribution Theory

We are interested in obtaining distributional approximations for the normalized coefficients \( \sqrt{n} V_{\theta K}^{-1/2} a'_K \left( \hat{\beta}_K - \beta_K \right) \), and understanding the impact of replacing the asymptotic covariance
matrix $V_{\theta K}$ with its estimate $\hat{V}_{\theta K}$. We start with the standard decomposition

$$
\sqrt{n} V_{\theta K}^{-1/2} a_K' \left( \hat{\beta}_K - \beta_K \right) = V_{\theta K}^{-1/2} a_K' \hat{Q}_K^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{Ki} e_{Ki} \right)
$$

where

$$
u_{ni} = V_{\theta K}^{-1/2} a_K' Q_K^{-1} x_{Ki} e_{Ki}
$$

and

$$R_n = V_{\theta K}^{-1/2} a_K' \left( \hat{Q}_K^{-1} - Q \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{Ki} e_{Ki} \right)
$$

We show below that $R_n$ is asymptotically negligible. Thus the asymptotic distribution of $\hat{\beta}_K$ is determined by the sum over $u_{ni}$. Furthermore, the vector $u_{ni}$ lies at the core of the estimated covariance matrix $\hat{V}_{\theta K}$. Hence understanding the asymptotic behavior of the normalized sums and outer products of $u_{ni}$ is key to understanding the distribution results. We now state perhaps the key result for our asymptotic theory.

**Theorem 5.** Under Assumption 1, and either 2 or 3,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{ni} \longrightarrow_d N(0, I_d)
$$

and

$$
\left\| \frac{1}{n} \sum_{i=1}^{n} u_{ni} u_{ni}' - I_K \right\| = o_p(1).
$$

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

**Theorem 6.** Under Assumption 1, and either 2 or 3,

$$
\sqrt{n} V_{\theta K}^{-1/2} a_K' \left( \hat{\beta}_K - \beta_K \right) \longrightarrow_d N(0, I_d)
$$

Theorem 6 shows that linear functions of the least-squares estimate $\hat{\beta}_K$ are asymptotically normal. The asymptotic distribution is centered at the projection coefficient $\beta_K$ and the linear combination $a_K' \hat{\beta}_K$ 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 $t < \infty$) as is typical in parametric theory as well as bounded regressors and variances ($q = \infty$ and $t = \infty$) as has previously been assumed in the nonparametric theory. Theorem 6 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 $t$ and
the permitted number of regressors \( K \).

Another new feature of Theorem 6 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 \( \hat{\theta}_K \) from (6).

Theorem 7. Under Assumption 1, and either 2 or 3,

\[
\sqrt{n}V_{\theta K}^{-1/2} \left( \hat{\theta}_K - \theta + a(r_K) \right) \longrightarrow_d N(0, I_d)
\]

where \( r_K(z) = g(z) - x_K(z)'\beta_K \).

Theorem 7 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 \( K \rightarrow \infty \). Theorem 7 provides a richer distribution theory, by smoothly nesting the parametric and nonparametric cases. Theorem 7 shows that the appropriate asymptotic variance is \( V_{\theta K} = a_K'Q_K^{-1}E(x_{Ki}x_{Ki}'e_{Ki}^2)Q_K^{-1}\alpha_K \) which depends on the projection error \( e_{Ki} \). This is an improvement on nonparametric sieve approximations which use the smaller asymptotic variance \( a_K'Q_K^{-1}E(x_{Ki}x_{Ki}'e_{i}^2)Q_K^{-1}\alpha_K \) which is only valid under the nonparametric assumption \( K \rightarrow \infty \) and is a poorer finite sample approximation.

More importantly, Theorem 7 shows that the asymptotic distribution of the estimate \( \hat{\theta}_K \) is centered at \( \theta - a(r_K) \), not at the desired parameter value \( \theta \). The term \( a(r_K) \) is the (finite-sample) bias of the estimator \( \hat{\theta}_K \). 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 \( \sqrt{n}\delta_K \rightarrow 0 \) or \( \sqrt{n}V_{\theta K}^{-1/2}a(r_K) \rightarrow 0 \). An undersmoothing assumption such as this is typical in the nonparametric sieve literature and allows the asymptotic distribution in Theorem 7 to be written without the bias term \( a(r_K) \). This is a deterioration in the asymptotic approximation, not an improvement. Theorem 7 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 7 is identical to the common inclusion of bias terms in the asymptotic distribution theory for nonparametric kernel estimation. Theorem 7 shows that the bias term can similarly be included in nonparametric sieve asymptotic theory.

We next investigate the impact of estimated covariance matrices.

Theorem 8. Under Assumption 1, and either 2 or 3,

\[
\sqrt{n}V_{\theta K}^{-1/2}a_K' \left( \hat{\beta}_K - \beta_K \right) \longrightarrow_d N(0, I_d)
\]
and
\[ \sqrt{n} \hat{V}_{\theta K}^{-1/2} \left( \hat{\theta}_K - \theta + a(r_K) \right) \rightarrow_d N(0, I_d) \]

Theorem 8 shows that the asymptotic distributions are unaffected by estimation of the covariance matrix. This is done without a strengthening of the assumptions.

6 Approximation Rates

In this section we give primitive conditions for the approximation in Assumption 1.3, which we re-state here as
\[ \inf_{\beta \in \mathbb{R}^K} \left\| g(z_i) - x'_{K_i} \beta \right\|_p \leq \delta_K \tag{18} \]

6.1 Series Moment Bound

Suppose that the regression function satisfies the series expansion \( g(x) = \sum_{j=1}^{\infty} \beta_j x_j(z) \) with \( \| x_j(z_i) \|_p \leq C < \infty \) and \( |\beta_j| \leq Aj^{-a} \) for \( a > 1 \). Set \( \beta_K = (\beta_1, ..., \beta_K) \) so that \( g(z_i) - x'_{K_i} \beta_K = \sum_{j=K+1}^{\infty} \beta_j x_j(z_i) \). Then by Minkowski’s inequality
\[
\inf_{\beta \in \mathbb{R}^K} \left\| g(z_i) - x'_{K_i} \beta \right\|_p \leq \left\| g(z_i) - x'_{K_i} \beta_K \right\|_p \\
= \left\| \sum_{j=K+1}^{\infty} \beta_j x_j(z_i) \right\|_p \\
\leq C \sum_{j=K+1}^{\infty} |\beta_j| \\
\leq \frac{C}{a-1} K^{-a}
\]

which is (18) with \( \delta_K \sim K^{-a} \).

6.2 Weighted Sup Norm

Let \( Z \subset \mathbb{R}^d \) denote the support of \( z_i \). Suppose that for some weight function \( w : Z \rightarrow \mathbb{R}^+ \)
\[ \inf_{\beta \in \mathbb{R}^K} \sup_{z \in Z} \frac{|g(z) - x_K(z)' \beta|}{w(z)} \leq \psi_K \tag{19} \]
and
\[ \| w(z_i) \|_p \leq C. \tag{20} \]

It is fairly straightforward to see that (19) and (20) imply (18) with \( \delta_K = C \psi_K \).

The norm in (19) is known as the weighted sup norm.

If \( Z \) is bounded and \( w(z) = 1 \) then (19) is the conventional uniform approximation and (20) is automatically satisfied.
Chen, Hong and Tamer (2005) use (19) with \( w(z) = (1 + \| z \|^2)^{\alpha/2} \) to allow for regressors with unbounded support. They do not discuss primitive conditions for (19). A theory of approximation in weighted sup norm with this weight function is given in Chapter 6 of Triebel (2006).

We use (19)-(20) 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 (18).

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 \((\int |g^{(s)}(z)|^p \exp(-A|z|^\alpha))^{1/p} \leq C < \infty \), then there is a \( c < \infty \) such that for every integer \( K \geq s + 1 \),

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

Mhaskar’s theorem can be used to directly imply a sufficient condition for (18).

**Theorem 9.** If \( z_i \) has a density \( f(z) \) which satisfies \( f(z) \leq C \exp(-A|z|^\alpha) \) for some \( C < \infty \), \( \alpha > 1 \), and \( A > 0 \), and for some integers \( s \) and \( p \), \( g^{(s-1)}(z) \) is absolutely continuous and \( \int |g^{(s)}(z)|^p e^{-A|z|^\alpha} \, dz \leq C < \infty \), then (18) holds with \( \delta_K = O\left(K^{-s(1-1/\alpha)}\right) \).

Theorem 9 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 \( \alpha \) of the density, the faster the rate of decay of the bias bound \( \delta_K \). 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 \( \delta_K = O\left(K^{-s/2}\right) \).

The assumptions require the \( s^{th} \) derivative of \( g(z) \) to satisfy the weighted \( L^p \) bound \( \int |g^{(s)}(z)|^p e^{-A|z|^\alpha} \leq C \). If the derivative is bounded, e.g. \( |g^{(s)}(z)| \leq B < \infty \), then this requirement holds for all \( p \), so (18) holds with the uniform norm. The weighted \( L^p \) bound is much weaker, allowing for functions \( g(x) \) with unbounded derivatives. For example, Theorem 9 applies to the exponential regression function \( g(z) = \exp(az) \).

We are also investigating conditions for the norm bounds \( \zeta_K \) of Assumption 1. The following result is preliminary.

**Proposition 2.** If \( z_i \) has a density \( f(z) \) which satisfies \( f(z) \leq C_1 \exp\left(-\frac{1}{2} |x/\sigma_1|^\alpha\right) \) and \( f(z) \geq C_2 \exp\left(-\frac{1}{2} |x/\sigma_2|^\alpha\right) \) for some constants \( C_1, C_2, \sigma_1, \sigma_2, \) and \( \alpha \geq 1 \), and \( p_{K_i} = (1, z, z^2, ..., z^{K-1}) \), there is a rotation \( x_{K_i} = B_{Kp_{K_i}} \) which satisfies Assumption 1.1 for any \( q \) with \( \zeta_K = O\left(\exp(\theta K)\right) \) where \( \theta \) depends on \( q \).
Proposition 1 is intriguing, for it shows that polynomial series in unbounded regressors which satisfy an exponential tail bound (including the normal) satisfy the moment bound of Assumption 1.1 but with an exponential rate. The proof of Proposition 1 is similar to that for the bounded case. The regressors \( x_{Ki} \) are formed as an orthogonal rotation of the power series \( p_{Ki} \), but in this case using a Freud polynomial, which is the orthogonal polynomial with respect to the weight function \( \exp \left( -\frac{1}{2} |x/\sigma_2|^\alpha \right) \). This ensures that the design matrix \( Q_K \) is uniformly bounded away from singularity. Next, the moments of \( x_{Ki} \) are bounded by the moments of Freud polynomial transformation of the variable \( z_i \). The exponential rate is a product of the factorial growth in the moments of the distributions with exponential tails (including the normal).

This exponential rate is a surprising difference from the bounded cases where the bounds satisfy the linear bound \( \zeta_K = O(K) \). Combined with the rate conditions of Assumptions 1 and 2, this implies that the number of variables \( K \) can only increase very slowly. Hence, polynomial regression with regressors with unbounded support seems to require a tight control on the number of regressors.

### 6.4 Multivariate Polynomials

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

**Proposition 3.** (Maioriv-Meir) If for some integers \( 1 \leq p \leq \infty \) and \( s \geq 0 \), and for some \( \alpha \geq 2 \) and \( A > 0 \), \( \left( \int |g^{(s)}(z)|^p \exp \left( -A |z|_\alpha \right) \right)^{1/p} \leq C < \infty \) for all \( |s| \leq s \) then there is a \( c < \infty \) such that for every integer \( K \),

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

**Theorem 10.** If \( z_i \) has a density \( f(z) \) which satisfies \( f(z) \leq C \exp \left( -A |z|_\alpha \right) \) for some \( C < \infty \), \( \alpha \geq 2 \) and \( A > 0 \), and for some integers \( s \) and \( p \), \( \left( \int |g^{(s)}(z)|^p \exp \left( -A |z|_\alpha \right) dz \right)^{1/p} \leq C < \infty \) for all \( |s| \leq s \), then (18) 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 \( p(z) = \sum_{j=1}^s c_j z^{j-1} \). 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 \( j = 1, \ldots, K \) 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 constrained to have continuous $s-2$ derivatives.

**Theorem 11.** If $\sup_z |g^{(s)}(z)| \leq C$, $\sup_z |g^{(s-1)}(z)| \leq C$ and $\|z_t\|_{ap} \leq C < \infty$ then (18) 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.** Suppose that $\mathbb{E}|U_i|^s < \infty$ for some $s > 0$. Then for any $s \geq v > 0$ and $b > 0$

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

**Proof**

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

\[ \blacksquare \]

Our proof of Theorem 1 is based on the following key inequality implicit in Rudelson (1999) and Oliveira (2010).

**Lemma 2.** For any sequence of independent $L(n) \times L(n)$ Hermitian matrices $Q_{ni}$

$$\mathbb{E}\left|\frac{1}{n} \sum_{i=1}^n (Q_{ni} - \mathbb{E}Q_{ni})\right| \leq C_1 \sqrt{\frac{\log(2L(n))}{n}} \mathbb{E}\left(\left|\frac{1}{n} \sum_{i=1}^n Q_{ni}Q_{ni}^*\right|^{1/2}\right)$$

where $C_1 = 2\sqrt{2} \left(1 + \sqrt{\frac{\pi}{\log 2}}\right) \simeq 8.85$.

**Proof:** Let $\varepsilon_i$ be a sequence of i.i.d. Rademacher random variables independent of the $Q_{ni}$. Since the $Q_{ni}$ are independent, by the Symmetrization Lemma (e.g. Lemma 2.3.6 of van der Vaart and
Wellner (1996)

$$
\mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} (Q_{ni} - \mathbb{E}Q_{ni}) \right\| \leq 2\mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} Q_{ni}\varepsilon_i \right\|
$$

(21)

where $\mathbb{E}_\varepsilon$ denotes expectation over the $\varepsilon_i$ only. Theorem 1 of Oliveira (2010) (for $p = 1$ in his notation) states that

$$
\mathbb{E}_\varepsilon \left\| \sum_{i=1}^{n} Q_{ni}\varepsilon_i \right\| \leq \left( \sqrt{\frac{2\log(2L) + \sqrt{2\pi}}{C_1^2 \log(2L)}} \right) \left\| \sum_{i=1}^{n} Q_{ni}Q_{ni} \right\|^{1/2}
$$

(22)

the second inequality since $L(n) \geq 1$. Combining (21) and (22) yields the result. □

We now establish Theorem 1. The proof extends Lemma 6.2 of Belloni, Chernozhukov, Chetverikov, and Kato (2015) by adding a trimming argument. The key is Lemma 2.

**Proof of Theorem 1.** Fix $0 < \delta < 1$. Write $L = L(n)$. Set

$$
\begin{align*}
    u_{1i} &= u_{ni} 1 \left( \|u_{ni}\|^2 \leq \frac{\delta^2 n}{C_1^2 \log(2L)} \right), \\
    u_{2i} &= u_{ni} 1 \left( \|u_{ni}\|^2 > \frac{\delta^2 n}{C_1^2 \log(2L)} \right).
\end{align*}
$$

By the triangle inequality

$$
\mathbb{E} \left\| \hat{M}_n - \mathbb{E} \hat{M}_n \right\| \leq \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} (u_{1i}u'_{1i} - \mathbb{E}(u_{1i}u'_{1i})) \right\| + \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} (u_{2i}u'_{2i} - \mathbb{E}(u_{2i}u'_{2i})) \right\|.
$$

(23)

Using the triangle inequality, (24) is bounded by

$$
\frac{2}{n} \sum_{i=1}^{n} \mathbb{E} \|u_{2i}u'_{2i}\| = \frac{2}{n} \sum_{i=1}^{n} \mathbb{E} \left( \|u_{ni}\|^2 1 \left( \|u_{ni}\|^2 > \frac{\delta^2 n}{C_1^2 \log(2L)} \right) \right) \leq \delta
$$

(25)

the final inequality by (7) for $n$ sufficiently large.
Applying Lemma 2 with $Q_{ni} = u_{1i}^t u_{1i}'$, (23) is bounded by

$$
C_1 \sqrt{\frac{\log(2L)}{n}} \mathbb{E} \left( \left| \frac{1}{n} \sum_{i=1}^{n} (u_{1i} u_{1i}') \left( u_{1i} u_{1i}' \right) \right|^{1/2} \right)
$$

$$
= C_1 \sqrt{\frac{\log(2L)}{n}} \mathbb{E} \left( \left| \frac{1}{n} \sum_{i=1}^{n} u_{ni} u_{ni}' \left| u_{ni} \right|^{2} 1 \left( \left| u_{ni} \right|^{2} \leq \frac{\delta^2 n}{C_1 \log(2L)} \right) \right|^{1/2} \right)
$$

$$
\leq C_1 \sqrt{\frac{\log(2L)}{n}} \frac{\delta^2 n}{C_1 \log(2L(n))} \mathbb{E} \left( \left| \frac{1}{n} \sum_{i=1}^{n} u_{ni} u_{ni}' \right|^{1/2} \right)
$$

$$
= \delta \mathbb{E} \left( \left| \tilde{M}_n \right|^{1/2} \right)
$$

$$
\leq \delta \left( \mathbb{E} \left| \tilde{M}_n \right| \right)^{1/2}
$$

$$
\leq \delta \left( \mathbb{E} \left| \tilde{M}_n - \mathbb{E} \tilde{M}_n \right| + \mu_n \right)^{1/2}
$$

$$
\leq \delta^{1/2} \left( \mathbb{E} \left| \tilde{M}_n - \mathbb{E} \tilde{M}_n \right| \right)^{1/2} + \delta \mu_n^{1/2}.
$$

(26)

The first inequality uses $\|u_{ni}\|^2 1 \left( \|u_{ni}\|^2 \leq \frac{\delta^2 n}{C_1 \log(2L)} \right) \leq \frac{\delta^2 n}{C_1 \log(2L(n))}$. The second inequality is Liapunov’s. The third inequality uses the triangle inequality and the assumption $\|\mathbb{E} \tilde{M}_n\| \leq \mu_n$. The fourth is the $C^r$ inequality and $\delta \leq \delta^{1/2}$ since $\delta < 1$.

Equations (23), (24), (25) and (26) combine to imply

$$
D_n = \mathbb{E} \left| \tilde{M}_n - \mathbb{E} \tilde{M}_n \right|
$$

$$
\leq \delta^{1/2} D_n^{1/2} + \left( 1 + \mu_n^{1/2} \right) \delta
$$

$$
\leq \delta^{1/2} D_n^{1/2} \left( 1 + \mu_n \right)^{1/4} + 2 \left( 1 + \mu_n \right)^{1/2} \delta.
$$

Factoring we find

$$
\left( D_n^{1/2} - 2 \left( 1 + \mu_n \right)^{1/4} \delta^{1/2} \right) \left( D_n^{1/2} + \left( 1 + \mu_n \right)^{1/4} \delta^{1/2} \right) \leq 0
$$

which implies

$$
D_n \leq 4 \left( 1 + \mu_n \right)^{1/2} \delta.
$$

Since $\delta$ is arbitrary this implies $D_n = o \left( \left( 1 + \mu_n \right)^{1/2} \right)$ as claimed.

Proof of Corollary 1.1. Set $B_n = \varepsilon n / \log(2L)$. Since $u_i$ is iid and $\mathbb{E} \|u_i\|^2 < \infty$

$$
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left( \|u_i\|^2 1 \left( \|u_i\|^2 > \frac{n}{\log(2L)} \right) \right) = \mathbb{E} \left( \|u_i\|^2 1 \left( \|u_i\|^2 > B_n \right) \right) \rightarrow 0
$$
as $n \to \infty$ since $B_n \to \infty$. ■

**Proof of Corollary 1.2.** Set $B_n = \varepsilon n / \log (2L)$ and

$$C(B) = \sup_n \sup_{i \leq n} \mathbb{E} \left( \|u_{ni}\|^2 1 \left( \|u_{ni}\|^2 > B \right) \right)$$

which satisfies $\lim_{B \to \infty} C(B) = 0$. Then

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left( \|u_{ni}\|^2 1 \left( \|u_{ni}\|^2 > \varepsilon \frac{n}{\log (2L)} \right) \right) \leq C(B_n) \to 0$$

as $n \to \infty$ since $B_n \to \infty$. ■

**Proof of Corollary 1.3.** Let $n$ be large enough so that $n^{-1} \xi_n^2 \log(2L(n)) < \varepsilon$. Then since $\|u_{ni}\| \leq \xi_n$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left( \|u_{ni}\|^2 1 \left( \|u_{ni}\|^2 > \varepsilon \frac{n}{\log (2L(n))} \right) \right) \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left( \|u_{ni}\|^2 1 \left( \frac{\xi_n^2 \log (2L(n))}{n} > \varepsilon \right) \right) = 0.$$

■

**Proof of Corollary 1.4.** Using Lemma 1,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left( \|u_{ni}\|^2 1 \left( \|u_{ni}\|^2 > \varepsilon \frac{n}{\log (2L(n))} \right) \right) \leq \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{E} \|u_{ni}\|^s}{(\varepsilon \log (2L(n)))^{s/2-1}}$$

$$\leq \frac{1}{\varepsilon^{s/2-1}} \left( \frac{\xi_n^{2s/(s-2)} \log (2L(n))}{n} \right)^{s/2-1} \to 0$$

since $s > 2$ and the assumption $n^{-1} \xi_n^{2s/(s-2)} \log (2L(n)) \to 0$. ■

**Proof of Theorem 2.** Define

$$b_n = \left( \frac{\xi_n^s}{\log (2L(n)) (1 + \mu_n)} \right)^{1/(s-2)}$$

$$M_{1n} = \frac{1}{n} \sum_{i=1}^{n} u_{ni} u_{ni}' 1 (\|u_{ni}\| \leq b_n)$$

$$M_{2n} = \frac{1}{n} \sum_{i=1}^{n} u_{ni} u_{ni}' 1 (\|u_{ni}\| > b_n)$$

17
so that
\[
\mathbb{E}\left\| \hat{M}_n - \mathbb{E}\hat{M}_n \right\| \leq \mathbb{E}\left\| \hat{M}_1n - \mathbb{E}\hat{M}_1n \right\| + \mathbb{E}\left\| \hat{M}_2n - \mathbb{E}\hat{M}_2n \right\| .
\] (27)

Since
\[
\|u_{ni}1(\|u_{ni}\| \leq b_n)\| \leq b_n,
\]
and
\[
\frac{b_n^2 \log (2L)}{n} = \frac{\xi_n^{2s/(s-2)} (\log (2L(n)))^{(s-4)/(s-2)}}{n} = o(1),
\]
the conditions for Lemma 6.2 of Belloni, Chernozhukov, Chetverikov, and Kato (2015) are satisfied, which states that
\[
\mathbb{E}\left\| \hat{M}_1n \right\| \leq \mathbb{E}\left\| \hat{M}_n \right\| \leq \mu_n,
\]
and
\[
\mathbb{E}\left\| \hat{M}_2n \right\| \leq \mathbb{E}\left\| \hat{M}_2n \right\| \leq \mu_n.
\]

[Their theorem is stated with \(\log n\) replacing \(\log (2L(n))\), and with the stronger assumption \(\|\mathbb{E}u_{ni}u'_{ni}\| \leq \mu_{ni}\) but it is clear from their proof that the statement holds with \(\log (2L(n))\) and the weaker bound \(\mathbb{E}\left\| \hat{M}_n \right\| \leq \mu_{ni}\).]

Set
\[
U_{ni} = u_{ni}u'_{ni}1(\|u_{ni}\| > b_n) - \mathbb{E} (u_{ni}u'_{ni}1(\|u_{ni}\| > b_n)).
\]

Using the inequality \(\|A\|^2 = \lambda_{\text{max}} (A' A) \leq \text{tr} (A' A)\), the fact that \(U_{ni}\) are independent and mean zero, \(s \geq 4\), Lemma 1, and (9),
\[
\mathbb{E}\left\| \hat{M}_2n - \mathbb{E}\hat{M}_2n \right\|^2 \leq \mathbb{E} \text{tr} \left( \left( \hat{M}_2n - \mathbb{E}\hat{M}_2n \right) \left( \hat{M}_2n - \mathbb{E}\hat{M}_2n \right) \right)
\]
\[
= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E} \text{tr} (U_{ni}U_{nj})
\]
\[
= \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E} \text{tr} (U_{ni}U_{ni})
\]
\[
\leq \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E} \left( \|u_{ni}\|^4 1(\|u_{ni}\| > b_n) \right)
\]
\[
\leq \frac{\xi_n^s}{nb_n^4}
\]
\[
= \frac{\xi_n^{2s/(s-2)} (\log (2L(n)))^{(s-4)/(s-2)}}{n}.
\]
Using Liapunov’s inequality we conclude that

\[
\mathbb{E} \left\| \hat{M}_{2n} - \mathbb{E} \hat{M}_{2n} \right\| \leq \left( \mathbb{E} \left( \left\| \hat{M}_{2n} - \mathbb{E} \hat{M}_{2n} \right\|^2 \right) \right)^{1/2} \leq \sqrt{\frac{2s/(s-2)}{n} \log \left( 2L(n) \right) \left( 1 + \mu_n \right) (s-4)/(s-2)}.
\]

Equations (27), (28), and (29) show (11). ■

**Proof of Theorem 3.** Take any \( \varepsilon > 0 \). For some \( 0 < \eta < 1 \) and \( 0 < B < \infty \) define the event

\[ E_n = \left\{ \left\| \hat{M}_n - I_{L(n)} \right\| \leq Ba_n \leq \eta \right\} . \]

The assumptions that \( \left\| \hat{M}_n - I_{L(n)} \right\| = O_p(a_n) \) 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 \( \hat{M}_n \) is positive definite we can write \( \hat{M}_n = H' \Lambda H \) where \( H'H = I_{L(n)} \) and \( \Lambda = \text{diag} \left( \lambda_1, \ldots, \lambda_{L(n)} \right) \).

Then

\[ \left\| \hat{M}_n - I_{L(n)} \right\| = \left\| H' \left( \Lambda - I_{L(n)} \right) H \right\| = \left\| \Lambda - I_{L(n)} \right\| = \max_{1 \leq j \leq L(n)} |\lambda_j - 1| . \]

Thus the event \( E_n \) implies

\[ \max_{1 \leq j \leq L(n)} |\lambda_j - 1| \leq Ba_n \leq \eta . \]

Combined with a Taylor expansion we find

\[ \max_{1 \leq j \leq L} |\lambda_j^r - 1| \leq |r| (1 - \eta)^{-|r-1|} \max_{1 \leq j \leq L(n)} |\lambda_j - 1| \leq |r| (1 - \eta)^{-|r-1|}Ba_n . \]

This implies that on event \( E_n \),

\[ \left\| \hat{M}^r_n - I_{L(n)} \right\| = \left\| H' \left( \Lambda^r - I_{L(n)} \right) H \right\| = \max_{1 \leq j \leq L(n)} |\lambda_j^r - 1| \leq |r| (1 - \eta)^{-|r-1|}Ba_n . \]

Since \( P(E_n) \geq 1 - \varepsilon \) this means \( \left\| \hat{M}^r_n - I_{L(n)} \right\| = O_p(a_n) \) as claimed. ■

**Proof of Theorem 4.1.** Using Schwarz matrix inequality \( \|AB\| \leq \|A\| \|B\| \) and Assumption 1.1,

\[ \|x_{Ki}\|_q \leq \|Q_K^{-1/2}\| \|x_{Ki}\|_q = \sqrt{\lambda_{\text{max}} (Q_K^{-1})} \|x_{Ki}\|_q = \frac{\|x_{Ki}\|_q}{\sqrt{\lambda_{\text{max}} (Q_K)}} \leq \zeta_K . \]

as stated. ■

**Proof of Theorem 4.2.** The conditions of Theorem 2 hold with \( u_{ni} = \bar{x}_{Ki} \), \( \xi_n = \zeta_K \), \( \mu_n = 1 \),
$L = K$, and $s = q \geq 4$. Theorem 2 implies

$$
E \| \tilde{Q}_K - I_K \| = E \left\| \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{K_i} \tilde{x}'_{K_i} - I_K \right\|
= O \left( \sqrt{\frac{2q/(q-2)}{s_K \log (2K)} \frac{(q-4)/(q-2)}{n}} \right)
$$

as stated. ■

**Proof of Theorem 4.3.** Theorem 4.2 and Markov’s inequality imply

$$
\| \tilde{Q}_K - I_K \| = O_p \left( \sqrt{\frac{2q/(q-2)}{s_K \log (2K)} \frac{(q-4)/(q-2)}{n}} \right)
$$

Theorem 3 with $r = -1$ yields

$$
\| \tilde{Q}_K^{-1} - I_K \| = \left\| \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{K_i} \tilde{x}'_{K_i} \right)^{-1} - I_K \right\|
= O_p \left( \sqrt{\frac{2q/(q-2)}{s_K \log (2K)} \frac{(q-4)/(q-2)}{n}} \right)
$$

as stated. ■

### 8 Proofs of Theorem 5

To establish Theorem 5 we start with some simple but useful bounds and this will require some new notation. In particular, it will be convenient to decompose the projection error $e_{K_i}$. Recall the definition of the $L^p$ approximation error $r_{K_i}^*$ from (13). Define the projection (or $L^2$) approximation error and differences between the projection and $L^p$ approximation errors

$$
r_{K_i} = g(z_i) - x'_i \beta_K
$$

$$
r_{K_i}^* = r_{K_i} - r_{K_i}^*
$$

Then we have the decompositions

$$
e_{K_i} = e_i + r_{K_i} = e_i + r_{K_i}^* + r_{K_i}^{**}.
$$
Notice that when \( p = 2 \) then \( r_{Ki} = r_{Ki}^* \) and \( r_{Ki}^{**} = 0 \), but when \( p > 2 \) then \( r_{Ki}^{**} \neq 0 \).

Also define

\[
A_K = Q_K^{-1/2} a_K V_{\theta K}^{-1/2}
\]

and observe that \( u_{ni} = A_K' x_{Ki} e_{Ki} \).

**Lemma 3.** Under Assumptions 1 and 2

1. \( \|A_K\|^2 \leq C^{-1} \)
2. \( \|r_{Ki}\| \leq \|r_{Ki}^*\| \leq \delta_K \)
3. \( |r_{Ki}^{**}| \leq \|\bar{x}_{Ki}\| \delta_K \)

**Proof of Lemma 3.1.** Recall from Assumption 2.6 the definition \( \bar{a}_K = a_K (a_K' Q_K^{-1} a_K)^{-1/2} \) and recall \( V_{\theta K} = a_K' V_K a_K \). Then since \( \|A' A\| = \|AA'\| \) and \( \bar{a}_K Q_K^{-1} \bar{a}_K = I_d \),

\[
\|A_K\|^2 = \left\| Q_K^{-1/2} a_K V_{\theta K}^{-1/2} \right\|^2 = \left\| Q_K^{-1/2} a_K (a_K' V_K a_K)^{-1} a_K' Q_K^{-1/2} \right\| = \left\| Q_K^{-1/2} \bar{a}_K (\bar{a}_K' V_K \bar{a}_K)^{-1} \bar{a}_K' Q_K^{-1/2} \right\| = \left\| (\bar{a}_K' V_K \bar{a}_K)^{-1/2} \bar{a}_K' Q_K^{-1} \bar{a}_K (\bar{a}_K' V_K \bar{a}_K)^{-1/2} \right\| = \left\| (\bar{a}_K' V_K \bar{a}_K)^{-1} \right\| = 1/\lambda_{\min} (\bar{a}_K' V_K \bar{a}_K) \leq C^{-1},
\]

the final inequality using Assumption 2.6. \( \square \)

**Proof of Lemma 3.2.** Since the projection coefficient \( \beta_K \) minimizes the approximation error in the \( L^2 \) norm,

\[
\|r_{Ki}\|_2 = \left( \mathbb{E} (y_i - x_{Ki}' \beta_K)^2 \right)^{1/2} \leq \left( \mathbb{E} (y_i - x_{Ki}' \beta_K^*)^2 \right)^{1/2} = \|r_{Ki}^*\|_2 \leq \|r_{Ki}^*\|_p \leq \delta_K.
\]

The second inequality is Liapunov's and the final inequality is Assumption 2.3. \( \square \)
Proof of Lemma 3.3. Since \( y_i = x_{Ki}' \beta_{Ki}^* + e_i + r_{Ki}^* \) and \( \mathbb{E}(x_{Ki}; e_i) = 0 \),

\[
\begin{align*}
  r_{Ki}^* &= x_{Ki}'(\beta_{Ki}^* - \beta_K) \\
  &= x_{Ki}'(\beta_{Ki}^* - \mathbb{E}(x_{Ki}; x_{Ki}')^{-1}\mathbb{E}(x_{Ki}; y_i)) \\
  &= -x_{Ki}'\mathbb{E}(x_{Ki}; x_{Ki}')^{-1}\mathbb{E}(x_{Ki}; r_{Ki}^*).
\end{align*}
\]

Thus by the Schwarz inequality, the projection inequality, and Lemma 3.2

\[
|r_{Ki}^*| \leq \left\| Q_{Ki}^{-1/2} x_{Ki} \right\| \left\| \mathbb{E}(r_{Ki}^* x_{Ki}'); \mathbb{E}(x_{Ki}; x_{Ki}')^{-1}\mathbb{E}(x_{Ki}; r_{Ki}^*) \right\|^{1/2}
\leq \left\| x_{Ki} \right\| (\mathbb{E} r_{Ki}^* x_{Ki}')^{1/2}
\leq \left\| x_{Ki} \right\| \delta_K
\]

as stated. ■

We now establish Theorem 4. What is unconventional in the proof is that we apply the Lindeberg condition separately to the three error components from the decomposition (32). 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.

Proof of Theorem 5: We show below that for any \( \varepsilon > 0 \)

\[
\lim_{n \to \infty} \mathbb{E} \left( \| u_{ni} \|^2 \mathbb{1} \left( \| u_{ni} \|^2 > 9\varepsilon n \right) \right) = 0.
\]

This is the classic Lindeberg condition, since the dimension \( d \) of \( u_{ni} \) is fixed. Since \( u_{ni} \) are mutually independent, \( \mathbb{E} u_{ni} = 0 \), and

\[
\mathbb{E} u_{ni} u_{ni}' = V_{\theta K}^{-1/2} a_K' Q_K^{-1} \mathbb{E} x_{Ki}; x_{Ki}' e_{Ki}^2 Q_K^{-1} a_K V_{\theta K}^{-1/2}
= V_{\theta K}^{-1/2} a_K' Q_K^{-1} S_K Q_K^{-1} a_K V_{\theta K}^{-1/2}
= I_d,
\]

the conditions for the Lindeberg central limit theorem are satisfied. Thus (16) holds.

We now show (34). Using definition (33) write \( u_{ni} = A_K^t \bar{x}_{Ki}; e_{Ki} \).

Applying the C_r inequality to decomposition (32) we find

\[
e_{Ki}^2 \leq 3(e_i^2 + r_{Ki}^2 + r_{Ki}^*)^2.
\]
This implies that
\[
\mathbb{E} \left[ \|u_{ni}\|^2 \left( \|u_{ni}\|^2 > 9\varepsilon n \right) \right] = \mathbb{E} \left[ \|A_K^T \tilde{x}_{Ki}\|^2 \epsilon_i^2 \left( \|A_K^T \tilde{x}_{Ki}\|^2 \epsilon_i^2 > 9\varepsilon n \right) \right] \\
\leq 3\mathbb{E} \left[ \|A_K^T \tilde{x}_{Ki}\|^2 (\epsilon_i^2 + r_{Ki}^2 + r_{Ki}^{**2}) \left( \|A_K^T \tilde{x}_{Ki}\|^2 (\epsilon_i^2 + r_{Ki}^2 + r_{Ki}^{**2}) > 3\varepsilon n \right) \right] 
\]  
(36)

The inequality
\[
\left( \sum_{j=1}^J a_j \right) \left( \sum_{j=1}^J a_j > b \right) \leq 2 \sum_{j=1}^J a_j (a_j > b/J) 
\]
for \(a_j \geq 0\) shows that (36) is bounded by 6 times
\[
\mathbb{E} \left[ \|A_K^T \tilde{x}_{Ki}\|^2 \epsilon_i^2 \left( \|A_K^T \tilde{x}_{Ki}\|^2 \epsilon_i^2 > \varepsilon n \right) \right] \]  
(37)

\[
+ \mathbb{E} \left[ \|A_K^T \tilde{x}_{Ki}\|^2 r_{Ki}^2 \left( \|A_K^T \tilde{x}_{Ki}\|^2 r_{Ki}^2 > \varepsilon n \right) \right] \]  
(38)

\[
+ \mathbb{E} \left[ \|A_K^T \tilde{x}_{Ki}\|^2 r_{Ki}^{**2} \left( \|A_K^T \tilde{x}_{Ki}\|^2 r_{Ki}^{**2} > \varepsilon n \right) \right]. \]  
(39)

The proof is completed by showing that (37)-(39) are \(o(1)\).

First, we examine (37) supposing that Assumption 1 holds. Using Lemma 3.1, Lemma 1, Holder’s inequality, Theorem 4.1, and Assumption 1.1, (37) is bounded by
\[
C^{-1} \mathbb{E} \left[ \|\tilde{x}_{Ki}\|^2 \epsilon_i^2 \left( \|\tilde{x}_{Ki}\|^2 \epsilon_i^2 > \varepsilon n C \right) \right] \leq \frac{\mathbb{E} \left( \|\tilde{x}_{Ki}\|^{\lambda_1} |e_i|^{\lambda_1} \right)}{C (\varepsilon n C)^{\lambda_1-2}/2} \leq \frac{\|\tilde{x}_{Ki}\|^{\lambda_1} \|e_i\|^{\lambda_1}}{C (\varepsilon n C)^{\lambda_1-2}/2} \leq \frac{D^{\lambda_1}}{2^{\lambda_1/2} (\varepsilon)^{(\lambda_1-2)/2}} \left( \frac{2^{(\lambda_1-1)}}{K n} \right)^{(\lambda_1-2)/2} 
\]
(40)

the final bound by Assumption 1.3. Hence (37) is \(o(1)\).

Now consider (37) supposing instead that Assumption 2 holds. Define
\[
\sigma^2(B) = \sup_z \mathbb{E} \left[ \epsilon_i^2 \left( \epsilon_i^2 > B \right) \right] 
\]
(41)

which satisfies \(\sigma^2(B) \to 0\) as \(B \to \infty\). Assumption 2.1 implies that there is some \(D < \infty\) sufficiently large such that \(\sup_z \mathbb{E} \left[ \epsilon_i^2 \left( \epsilon_i^2 > D/2 \right) \right] \leq D/2\) which means
\[
\sigma^2_i \leq D. \]  
(42)
Fix $\delta > 0$. Set

$$b_n = \left( \frac{D}{\delta C} \right)^{2/(q-2)} \zeta_K^{2q/(q-2)}.$$

Then (37) equals

$$\mathbb{E} \left[ \left\| A_K' \bar{x}_{Ki} \right\|^2 e_i^2 1 \left( \left\| A_K' \bar{x}_{Ki} \right\|^2 e_i^2 > \varepsilon n \right) 1 \left( \left\| \bar{x}_{Ki} \right\|^2 \leq b_n \right) \right] + \mathbb{E} \left[ \left\| A_K' \bar{x}_{Ki} \right\|^2 e_i^2 1 \left( \left\| A_K' \bar{x}_{Ki} \right\|^2 e_i^2 > \varepsilon n \right) 1 \left( \left\| \bar{x}_{Ki} \right\|^2 > b_n \right) \right].$$

Using Lemma 3.1 and (41), the first term (43) is bounded by

$$\mathbb{E} \left[ \left\| A_K' \bar{x}_{Ki} \right\|^2 e_i^2 1 \left( e_i^2 > \frac{\varepsilon n}{b_n} \right) \right] = \mathbb{E} \left[ \left\| A_K' \bar{x}_{Ki} \right\|^2 \mathbb{E} \left( e_i^2 1 \left( e_i^2 > \frac{\varepsilon n}{b_n} \right) \right) \left| z_i \right| \right] \leq \mathbb{E} \left[ \left\| A_K' \bar{x}_{Ki} \right\|^2 \sigma^2 \left( \frac{\varepsilon n}{b_n} \right) \right] \leq \frac{d}{C} \sigma^2 \left( \frac{\varepsilon n}{b_n} \right) \leq \delta$$

since

$$\mathbb{E} \left\| A_K' \bar{x}_{Ki} \right\|^2 \leq \operatorname{tr} \left( A_K' \mathbb{E} \left( \bar{x}_{Ki} A_K' \right) A_K \right) \leq d \left\| A_K \right\|^2 \leq d/C.$$

The final inequality in (45) holds for $n$ sufficiently large since $n/b_n \to \infty$. Using Lemma 3.1, (42), Lemma 1, and Theorem 4.1, the second term (44) is bounded by

$$C^{-1} \mathbb{E} \left( \left\| \bar{x}_{Ki} \right\|^2 e_i^2 1 \left( \left\| \bar{x}_{Ki} \right\|^2 > b_n \right) \right) = C^{-1} \mathbb{E} \left( \left\| \bar{x}_{Ki} \right\|^2 \left( \left\| \bar{x}_{Ki} \right\|^2 > b_n \right) \mathbb{E} \left( e_i^2 \left| z_i \right| \right) \right) \leq C^{-1} D \mathbb{E} \left( \left\| \bar{x}_{Ki} \right\|^2 1 \left( \left\| \bar{x}_{Ki} \right\|^2 > b_n \right) \right) \leq \frac{D \zeta_K^q}{C b_n^{(q-2)/2}} = \delta.$$

Together, (45) and (46) show that (37)=(43)+(44) is bounded by $2\delta$. Since $\delta$ is arbitrary this shows that (37) is $o(1)$ as claimed.

Second, take (38). Using Lemma 3.1, Lemma 1, Holder’s inequality, Theorem 4.1 and Assump-
tion 1.3, (38) is bounded by
\[ C^{-1} \mathbb{E} \left[ \|x_{Ki}\|^2 r_{Ki}^2 1 \left( \|x_{Ki}\|^2 r_{Ki}^* > \varepsilon nC \right) \right] \leq \frac{\mathbb{E} \left( \|x_{Ki}\|^\lambda_2 |r_{Ki}^*|^\lambda_2 \right)}{C (\varepsilon nC)^{(\lambda_2 - 2)/2}} \]
\[ \leq \frac{1}{C^{\lambda_2/2} \varepsilon^{(\lambda_2 - 2)/2} n^{(\lambda_2 - 2)/2}} \frac{\|x_{Ki}\|^\lambda_2 |r_{Ki}^*|^\lambda_2}{p} \]
\[ \leq \frac{1}{C^{\lambda_2/2} \varepsilon^{(\lambda_2 - 2)/2} n^{(\lambda_2 - 2)/2}} \left( \zeta K \delta K \right)^{\lambda_2} \]
\[ = o(1). \tag{47} \]

by Assumption 1.5. Thus (38) is \( o(1) \).

Now take (39). Using Lemma 3.1, Lemma 1, Theorem 4.1, and Lemma 3.3, (39) is bounded by
\[ C^{-1} \mathbb{E} \left[ \|x_{Ki}\|^2 r_{Ki}^{*2} 1 \left( \|x_{Ki}\|^2 r_{Ki}^{*2} > \varepsilon nC \right) \right] \leq \frac{\mathbb{E} \left( \|x_{Ki}\|^{q/2} |r_{Ki}^*|^{q/2} \right)}{C (\varepsilon nC)^{(q-4)/4}} \]
\[ \leq \frac{1}{C^{q/4} \varepsilon^{(q-4)/4} n^{(q-4)/4}} \frac{\mathbb{E} \|x_{Ki}\|^{q} \delta_{Ki}^{q/2}}{\delta_{Ki}^{q/2}} \]
\[ \leq \frac{1}{C^{q/4} \varepsilon^{(q-4)/4} n^{(q-4)/4}} \left( \zeta^2 K \delta K \right)^{q/2} \]
\[ = o(1). \tag{48} \]

by Assumption 1.6. Thus (39) is \( o(1) \).

We have shown that (36) is bounded by 6 times the sum of (37)-(39) which are \( o(1) \), completing the proof of (34) as needed. \( \blacksquare \)

9 Proofs of Theorems 6 and 7

Using (32), decompose \( R_n \) as defined in (15) as
\[ R_n = R_{1n} + R_{2n} \tag{49} \]
where

\[ R_{1n} = A_K' \left( Q_K^{-1} - I_K \right) T_{1n} \]
\[ R_{2n} = A_K' \left( Q_K^{-1} - I_K \right) T_{2n} \]
\[ T_{1n} = n^{-1/2} \sum_{i=1}^{n} \tilde{x}_{Ki} e_i \]
\[ T_{2n} = n^{-1/2} \sum_{i=1}^{n} \tilde{x}_{Ki} r_{Ki} . \]

Our goal is to show that \( R_{1n} = o_p(1) \) and \( R_{2n} = o_p(1) \).

Define

\[ \psi_n = \frac{\lambda_{2q/(q-2)} \log (2K)}{n} \leq o(1) \]

where the inequality holds under Assumption 1.

The following result will be useful. While it appears elementary, I am unaware of a previous statement.

\textbf{Lemma 4.} If \( \mathbb{E} [\| \phi_n \| | \Omega_n ] = o_p(1) \), then \( \| \phi_n \| = o_p(1) \).

\textbf{Proof of Lemma 4.} Fix \( \varepsilon > 0 \). Markov’s inequality implies that

\[ p_n = \mathbb{P} (\| \phi_n \| > \varepsilon | \Omega_n ) \leq \frac{\mathbb{E} [\| \phi_n \| | \Omega_n ]}{\varepsilon} . \]

Thus

\[ \mathbb{P} (p_n > \varepsilon ) \leq \mathbb{P} \left( \frac{\mathbb{E} [\| \phi_n \| | \Omega_n ]}{\varepsilon} > \varepsilon \right) \leq \varepsilon \]

where the second inequality holds for \( n \) sufficiently large since by assumption \( \mathbb{E} [\| \phi_n \| | \Omega_n ] = o_p(1) \).

Since \( p_n \leq 1 \) and (50)

\[ \mathbb{P} (\| \phi_n \| > \varepsilon ) = \mathbb{E} p_n \]
\[ \leq \varepsilon + \mathbb{P} (p_n > \varepsilon ) \]
\[ \leq 2 \varepsilon \]

Hence \( \| \phi_n \| = o_p(1) \).

Define \( \overline{S}_K = \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{Ki} \sigma_i^2 \) and \( Z_n = [z_1, ..., z_n] \).

\textbf{Lemma 5.} Under Assumptions 1 and 2,

1. \( \psi_n \| \mathbb{E} (T_{1n} T_{1n}' | Z_n) \| = \psi_n \| \overline{S}_K \| = o_p(1) \).
2. $\psi_n \mathbb{E} \| T_{2n} \|^2 = \psi_n \mathbb{E} \left( \| x_{Ki} \|^2 \sigma_i^2 \right) = o(1)$

3. $R_{1n} = o_p(1)$

4. $R_{2n} = o_p(1)$

Proof of Lemma 5.1. By conditioning and $\mathbb{E} (e_i | z_i) = 0$,

\[
\mathbb{E} \left( T_1 T_1' | Z_n \right) = \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^{n} \bar{x}_{Ki} e_i \sum_{j=1}^{n} \bar{x}_{Kj} e_j | Z_n \right) = \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}_{Ki}' \mathbb{E} \left( e_i^2 | z_i \right) = \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}_{Ki}' \sigma_i^2 = \mathcal{S}_K
\]

which is the first equality.

Suppose Assumption 1.1(a) holds. By the triangle inequality

\[
\mathbb{E} \left\| \mathcal{S}_K \right\| \leq \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}_{Ki}' \sigma_i^2 1 \left( \| \sigma_i \| \leq \zeta_i^{2/(t-2)} \right) \right\| + \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}_{Ki}' \sigma_i^2 1 \left( \| \sigma_i \| > \zeta_i^{2/(t-2)} \right) \right\|
\]

(52) is bounded by

\[
\zeta_i^{4/(t-2)} \mathbb{E} \left\| \bar{Q}_K \right\| \leq O \left( \zeta_i^{4/(t-2)} \right)
\]

where the inequality holds since

\[
\mathbb{E} \left\| \bar{Q}_K \right\| \leq \mathbb{E} \left\| \bar{Q}_K - I_K \right\| + \| I_K \| \leq O(1)
\]

by Theorem 4.2.

By the conditional Jensen inequality and Assumption 1.1,

\[
\| \sigma_i \| \leq \left( \mathbb{E} \left( e_i^2 | z_i \right) \right)^{1/t} \leq \| e_i \|_t \leq D.
\]
Set \( v = 2q/(q-2) \) so that \( 1/v + 1/q = 1/2 \). Then using the triangle inequality, (53) is bounded by

\[
\mathbb{E} \left( \| \bar{x}_{K_i} \|^2 \sigma_i^2 \mathbf{1} \left( \| \sigma_i \| > \zeta_K^{2/(t-2)} \right) \right) \leq \left( \mathbb{E} \| \bar{x}_{K_i} \|^2 q^{2/q} \left( \mathbb{E} \left( \sigma_i^2 \mathbf{1} \left( \| \sigma_i \| > \zeta_K^{2/(t-2)} \right) \right) \right)^{2/v}
\]

\[
\leq \zeta_K^2 \left( \mathbb{E} \left( \sigma_i^2 \zeta_K^{2/(t-2)} \right) \right)^{2/v}
\]

\[
\leq \frac{\zeta_K^2}{2^{1/(t-2)}} D^{2t/v}
\]

\[
\leq O \left( \zeta_K^{4/(t-2)} \right).
\]  

The first inequality is Hölder’s, the second is Lemma 1, and the third is (56). (52), (53), (54) and (57) show that \( \mathbb{E} \| \bar{S}_K \| \leq O \left( \zeta_K^{4/(t-2)} \right) \) and hence

\[
\psi_n \mathbb{E} \| \bar{S}_K \| = O \left( \zeta_K^{2q/(q-2)+4/(t-2)} \log (2K) \frac{1}{n} \right)
\]

\[
\leq O \left( \zeta_K^{2\lambda_1/(\lambda_1-2)} \log (2K) \frac{1}{n} \right)
\]

\[
\leq o(1).
\]

The first inequality in (58) holds since

\[
\frac{2q}{q-2} + \frac{4}{t-2} = \frac{2qt-8}{qt-2t-2q+4} \leq \frac{2qt}{qt-2t-2q} = \frac{2\lambda_1}{\lambda_1-2},
\]

and the second inequality in (58) is Assumption 1.4. Markov’s inequality implies \( \psi_n \| \bar{S}_K \| = o_p(1) \) as stated.

Alternatively, suppose Assumption 1.1(b) holds. Since \( \sigma_i^2 \leq D < \infty \) by (42)

\[
\| \bar{S}_K \| = \left\| \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \tilde{r}_{Ki} \right\| \leq D \| \bar{Q}_K \| \leq O_p(1)
\]

the final inequality by (55). Hence \( \psi_n \| \bar{S}_K \| = o_p(1) \) as stated.

**Proof of Lemma 5.2.** Since \( \bar{x}_{Ki} r_{Ki} \) is independent across \( i \) and mean zero,

\[
\mathbb{E} \| T_{2n} \|^2 = \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^{n} \bar{x}_{Ki} \tilde{r}_{Ki} \sum_{j=1}^{n} \bar{x}_{Kj} \tilde{r}_{Kj} \right) = \mathbb{E} \left( \| \bar{x}_{Ki} \|^2 r_{Ki}^2 \right)
\]

which is the first equality. Using the definition \( r_{Ki} = r_{Ki}^* + r_{Ki}^{**} \) and the \( C_t \) inequality

\[
\mathbb{E} \left( \| \bar{x}_{Ki} \|^2 r_{Ki}^2 \right) \leq 2\mathbb{E} \left( \| \bar{x}_{Ki} \|^2 r_{Ki}^{**2} \right) + 2\mathbb{E} \left( \| \bar{x}_{Ki} \|^2 r_{Ki}^{*2} \right),
\]

28
We now examine the two terms on the right-side of (61). By Liapunov's inequality, Holder's inequality, Theorem 4.1 and Assumption 1.3

\[
\mathbb{E} \left( \|x_{ki}\|^2 r_{K_i}^* \right) \leq \left( \mathbb{E} \left( \|x_{ki}\|^\lambda \|r_{K_i}^*\|^\lambda \right) \right)^{2/\lambda} \\
\leq \|x_{ki}\|_q^2 \|r_{K_i}^*\|_p^2 \\
\leq \zeta_K^2 \delta_K^2 \\
\leq O \left( \zeta_K^{2q/(q-2)} \delta_K^2 \right)
\]  

(62)

where the final inequality holds since \( q/(q-2) \geq 1 \).

Applying the \( C_p \) inequality to (31), using Lemma 3.2 and Assumption 1.3,

\[
\mathbb{E} r_{K_i}^* \leq 2\mathbb{E} r_{K_i}^* + 2\mathbb{E} r_{K_i}^* \leq 4\delta_K^2.
\]

(63)

Then using (63), Lemma 3.3, Lemma 1, and Theorem 4.1,

\[
\mathbb{E} \left( \|x_{ki}\|^2 r_{K_i}^* \right) = \mathbb{E} \left( \|x_{ki}\|^2 r_{K_i}^* 1 \left( \|x_{ki}\| \leq \zeta_K^{q/(q-2)} \right) \right) + \mathbb{E} \left( \|x_{ki}\|^2 r_{K_i}^* 1 \left( \|x_{ki}\| > \zeta_K^{q/(q-2)} \right) \right) \\
\leq \zeta_K^{2q/(q-2)} \mathbb{E} \left( r_{K_i}^* 1 \left( \|x_{ki}\| > \zeta_K^{q/(q-2)} \right) \right) \\
\leq 4\zeta_K^{2q/(q-2)} \delta_K^2 + \frac{\zeta_K^{q/(q-2)}}{(\zeta_K^{q/(q-2)})^{q-4}} \delta_K^2 \\
= 5\zeta_K^{2q/(q-2)} \delta_K^2.
\]

(64)

Assumption 1.6 implies \( \delta_K^2 = O \left( \zeta_K^{-4} \left( \log (2K) / n \right)^{4/q-1} \right) \). Hence (61), (62) and (64) imply

\[
\mathbb{E} \left( \|x_{ki}\|^2 r_{K_i}^* \right) \leq O \left( \zeta_K^{2q/(q-2)} \delta_K^2 \right) \\
\leq O \left( \zeta_K^{(8-2q)/(q-2)} \left( \log (2K) / n \right)^{4/q-1} \right)
\]  

(65)

Equations (60) and (65) imply that

\[
\psi_n \mathbb{E} |T_{2n}|^2 \leq O \left( \zeta_K^{2q/(q-2)} \log (2K) / n \right) \zeta_K^{(8-2q)/(q-2)} \left( \log (2K) / n \right)^{4/q-1} \\
= O \left( \left( \zeta_K^{2q/(q-2)} \log (2K) / n \right)^{4/q} \right) \\
= o(1)
\]

by Assumption 1.4.  ■
Proof of Lemma 5.3. We use four algebraic inequalities. First, that for any matrix $A$, $\|A\|^2 \leq \text{tr} AA'$. Second, that for any $K \times d$ matrix $A$, any conformable matrix $B$ and symmetric matrix $C$

$$\text{tr} (A'B'CBA) = \text{tr} (AA') \|B'C\| \leq d \|A\|^2 \|B\|^2 \|C\|.$$ 

Third, that for any $d \times d$ positive semi-definite matrix $A$, $\text{tr} A \leq d \|A\|$. Fourth, the norm inequality $\|AB\| \leq \|A\| \|B\|$.

Using these results, conditioning, Lemma 3.1, Theorem 4.3, and Lemma 5.1

$$\mathbb{E} \left[ \|R_{1n}\|^2 | Z \right] \leq \mathbb{E} \left[ \text{tr} R_{1n} R_{1n}' | Z \right]$$

$$= \text{tr} \left[ A_K' \left( \tilde{Q}_K^{-1} - I_K \right) \mathbb{E} (T_{1n} T_{1n}' | Z) \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right]$$

$$\leq d \|A_K'\|^2 \left\| \tilde{Q}_K^{-1} - I_K \right\|^2 \| \mathbb{E} (T_{1n} T_{1n}' | Z) \|$$

$$\leq d \|A_K\|^2 \left\| \tilde{Q}_K^{-1} - I_K \right\|^2 \| \mathbb{E} (T_{1n} T_{1n}' | Z) \|$$

$$\leq \frac{d}{C} O_p (\psi_n) \| \mathbb{E} (T_{1n} T_{1n}' | Z) \|$$

$$= o_p (1).$$

Lemma 4 implies that $\|R_{1n}\|^2 = o_p (1)$.

Proof of Lemma 5.4. By the norm inequality, Lemma 3.1, Theorem 4.3, and Lemma 5.2

$$\|R_{2n}\|^2 \leq \|A_K\|^2 \left\| \tilde{Q}_K^{-1} - Q_K^{-1} \right\|^2 \|T_{2n}\|^2$$

$$\leq C^{-1} O_p (\psi_n) \|T_{2n}\|^2$$

$$= o_p (1)$$

as stated.

Proof of Theorem 6:

Given the decomposition (14) and Theorem 5, to establish Theorem 6 is sufficient to show that $R_n = o_p (1)$ where $R_n$ is defined in (15). The proof is completed by showing that $R_{1n} = o_p (1)$ and $R_{2n} = o_p (1)$, which is established in Lemma 4 below.

Furthermore, consider the statistic from Theorem 7. Since $g(z) = g_K(z) + r_K(z)$, then by linearity and (4)

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

Thus

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

and

$$\sqrt{n} V_{\theta K}^{-1/2} \left( \hat{\theta}_K - \theta + a (r_K) \right) = \sqrt{n} V_{\theta K}^{-1/2} a_K' \left( \hat{\beta}_K - \beta_K \right)$$
which is identical to the statistic in Theorem 6. It follows that (??) (e.g. Lemma 4) is sufficient to establish both Theorems 6 and 7.

10 Covariance Matrix Estimation

In this section we establish Theorem 8 which shows that covariance matrix estimation does not affect the asymptotic distributions. Define

$$\hat{V}_{AK} = V_{\theta K}^{-1/2} \hat{\theta}_K V_{\theta K}^{-1/2}$$

so that

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

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

Theorem 6 established that

$$\sqrt{n} V_{\theta K}^{-1/2} a_K' \left( \hat{\beta}_K - \beta_K \right) \rightarrow^d N(0, I_d).$$

Theorem 8 follows if

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

By Theorem 3, it is sufficient to show that

$$\left\| \hat{V}_{AK} - I_K \right\| = o_p(1) \quad (66)$$

which is established in Lemma 7 below.

This section establishes (66). 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.

Define

Lemma 6. $\left\| \frac{1}{n} \sum_{i=1}^{n} \bar{x}_K i \bar{x}_K i' \left( \bar{x}_K i \bar{x}_K i \right) \right\| = O_p \left( s_{AK}^{2q/(q-2)} \right)$
**Proof.** By the triangle inequality

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} (\bar{x}'_{Ki} \bar{x}_{Ki}) \right\| \leq \left\| \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \| \bar{x}_{Ki} \|^{2} 1 \left( \| \bar{x}_{Ki} \| \leq \zeta_{K}^{q/(q-2)} \right) \right.
\]
\[
+ \left\| \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \| \bar{x}_{Ki} \|^{2} 1 \left( \| \bar{x}_{Ki} \| > \zeta_{K}^{q/(q-2)} \right) \right. \]
\[
\leq \zeta_{K}^{2q/(q-2)} \left\| \hat{Q}_{K} \right\|
\]
\[
+ \frac{1}{n} \sum_{i=1}^{n} \| \bar{x}_{Ki} \|^{4} 1 \left( \| \bar{x}_{Ki} \| > \zeta_{K}^{q/(q-2)} \right).
\]

The first term is \( O_p \left( \zeta_{K}^{2q/(q-2)} \right) \) by (55). The expectation of the second term is

\[
\mathbb{E} \left( \| \bar{x}_{Ki} \|^{4} 1 \left( \| \bar{x}_{Ki} \| > \zeta_{K}^{q/(q-2)} \right) \right) = \zeta_{K}^{q} \left( \frac{\zeta_{K}^{q/(q-2)}}{\zeta_{K}^{q/(q-2)}} \right)^{q-4} = \zeta_{K}^{2q/(q-2)}
\]

by Lemma 1 and Theorem 4.1 and hence the term is \( O_p \left( \zeta_{K}^{2q/(2-q)} \right) \) by Markov’s inequality. This establishes the result.  

**Lemma 7.** Under Assumption 1, \( \| \hat{V}_{AK} - I_{K} \| = o_p(1) \)

**Proof of Lemma 7.** Note if we recall the definition of \( A_{K} \) in (33).

\[
\hat{V}_{AK} = A'_{K} Q_{K}^{1/2} \hat{V} Q_{K}^{1/2} A_{K}
\]

By the triangle inequality

\[
\left\| \hat{V}_{AK} - I_{K} \right\| \leq \left\| A'_{K} Q_{K}^{1/2} \hat{Q}_{K}^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} x_{Ki} x'_{Ki} (\hat{e}^{2}_{Ki} - e_{Ki}^{2}) \right) \hat{Q}_{K}^{-1} Q_{K}^{1/2} A_{K} \right\| \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad (67)
\]
\[
+ \left\| A'_{K} Q_{K}^{1/2} \hat{Q}_{K}^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} x_{Ki} x'_{Ki} e_{Ki}^{2} \right) \hat{Q}_{K}^{-1} Q_{K}^{1/2} A_{K} - I_{K} \right\|. \quad \quad (68)
\]

Define \( \tilde{S}_{K} = n^{-1} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} e_{Ki}^{2} \). Using \( e_{Ki}^{2} = (\hat{e}_{Ki}^{2} - 2 e_{Ki} x'_{Ki} (\hat{\beta}_{K} - \beta_{K}) + (x'_{Ki} (\hat{\beta}_{K} - \beta_{K})^{2}
\[

32
and the triangle inequality, the first term (67) equals

\[
\left\| A_K' \tilde{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \tilde{\beta}_{Ki} - \beta_{Ki} \right) \tilde{Q}_K^{-1} A_K \right\|
\]

\[
\leq 2 \left\| A_K' \tilde{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \tilde{x}'_{Ki} \left( \tilde{\beta}_K - \beta_K \right) \tilde{Q}_K^{-1} A_K \right\|
\]

\[
+ \left\| A_K' \tilde{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \tilde{x}'_{Ki} \left( \tilde{\beta}_K - \beta_K \right) \right)^2 \tilde{Q}_K^{-1} A_K \right\|
\]

\[
\leq 2 \left[ \left\| A_K' \tilde{Q}_K^{-1} \tilde{S}_K \tilde{Q}_K^{-1} A_K - I_K \right\| + 1 \right]^{1/2}
\]

\[
\times \left\| A_K' \tilde{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \tilde{x}'_{Ki} \left( \tilde{\beta}_K - \beta_K \right) \right)^2 \tilde{Q}_K^{-1} A_K \right\|^{1/2}
\]

\[
+ \left\| A_K' \tilde{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \tilde{x}'_{Ki} \left( \tilde{\beta}_K - \beta_K \right) \right)^2 \tilde{Q}_K^{-1} A_K \right\|.
\]

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 \( n \times d \) matrices \( X_1 \) and \( X_2 \) whose \( i^{th} \) rows are \( \bar{x}'_{Ki} e_K \tilde{Q}_K^{-1} A_K \) and \( \bar{x}'_{Ki} \left( \tilde{\beta}_K - \beta_K \right) \bar{x}'_{Ki} \tilde{Q}_K^{-1} A_K \), respectively.

The second term (68) equals

\[
\left\| A_K' \tilde{Q}_K^{-1} \tilde{S}_K \tilde{Q}_K^{-1} A_K - I_K \right\| \leq \left\| A_K' \tilde{S}_K A_K - I_K \right\| + 2 \left\| A_K' \tilde{Q}_K^{-1} - \tilde{I}_K \right\| \tilde{S}_K A_K \right\|
\]

\[
+ \left\| A_K' \tilde{Q}_K^{-1} - \tilde{I}_K \right\| \tilde{S}_K \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right\|
\]

\[
\leq \left\| A_K' \tilde{S}_K A_K - I_K \right\| + 2 \left\| A_K' \tilde{Q}_K^{-1} - \tilde{I}_K \right\| \tilde{S}_K \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right\|^{1/2} \left[ \left\| A_K' \tilde{S}_K A_K - I_K \right\| + 1 \right]^{1/2}
\]

\[
+ \left\| A_K' \tilde{Q}_K^{-1} - \tilde{I}_K \right\| \tilde{S}_K \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right\|.
\]

To establish that (67)-(68) is \( o_p(1) \) it is sufficient to show the following three inequalities

\[
\left\| A_K' \tilde{S}_K A_K - I_K \right\| = o_p(1) \quad (69)
\]

\[
\left\| A_K' \tilde{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \tilde{x}'_{Ki} \left( \tilde{\beta}_K - \beta_K \right) \right)^2 \tilde{Q}_K^{-1} A_K \right\| = o_p(1) \quad (70)
\]

\[
\left\| A_K' \tilde{Q}_K^{-1} - \tilde{I}_K \right\| \tilde{S}_K \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right\| = o_p(1). \quad (71)
\]

Equation (69) is Theorem 5 (17), so it remains to show (70) and (71).
Take (70). Using
\[ Q_K^{1/2} (\beta_K - \beta_K) = \bar{Q}_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{Ki} e_{Ki} = n^{-1/2} \bar{Q}_K^{-1} (T_{1n} + T_{2n}) \]
and the $C_r$ inequality, we find
\[
\left( x'_{Ki} (\beta_K - \beta_K) \right)^2 = \frac{1}{n} \left( \bar{x}'_{Ki} \bar{Q}_K^{-1} (T_{1n} + T_{2n}) \right)^2 \\
\leq \frac{2}{n} \left( \bar{x}'_{Ki} \bar{Q}_K^{-1} T_{1n} \right)^2 + \frac{2}{n} \left( \bar{x}'_{Ki} \bar{Q}_K^{-1} T_{2n} \right)^2.
\]
Thus (70) is bounded by 2 multiplied by
\[
\left\| A_K' \bar{Q}_K^{-1} \frac{1}{n^2} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \bar{x}'_{Ki} \bar{Q}_K^{-1} T_{1n} \right)^2 \bar{Q}_K^{-1} A_K \right\| + \left\| A_K' \bar{Q}_K^{-1} \frac{1}{n^2} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \bar{x}'_{Ki} \bar{Q}_K^{-1} T_{2n} \right)^2 \bar{Q}_K^{-1} A_K \right\|. \tag{72}
\]
Take (72). Observe that
\[
\left\| \bar{Q}_K^{-1} \right\| \leq \left\| \bar{Q}_K^{-1} - I_K \right\| + \left\| I_K \right\| \leq O_p(1) \tag{74}
\]
by Theorem 4.3. Using the inequalities $\|A\| \leq \text{tr} A$ and $\text{tr} A \leq d \|A\|$ for positive semi-definite $A$, conditioning, Lemma 3.1, Lemma 6, and (74)
\[
\mathbb{E} \left[ \left\| A_K' \bar{Q}_K^{-1} \frac{1}{n^2} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \left( \bar{x}'_{Ki} \bar{Q}_K^{-1} T_{1n} \right)^2 \bar{Q}_K^{-1} A_K \right\| \right] \\
\leq \text{tr} \left[ A_K' \bar{Q}_K^{-1} \frac{1}{n^2} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \bar{x}'_{Ki} \bar{Q}_K^{-1} \mathbb{E} (T_{1n} T_{1n}' | Z) \bar{Q}_K^{-1} \bar{x}_{Ki} \bar{Q}_K^{-1} A_K \right] \\
\leq \text{tr} \left[ A_K' \bar{Q}_K^{-1} \frac{1}{n^2} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \| \bar{x}_{Ki} \|^2 \bar{Q}_K^{-1} A_K \right] \left\| \bar{Q}_K^{-1} \right\|^2 \left\| \mathbb{E} (T_{1n} T_{1n}' | Z) \right\| \\
\leq d \|A_K\|^2 \left\| \frac{1}{n^2} \sum_{i=1}^{n} \bar{x}_{Ki} \bar{x}'_{Ki} \| \bar{x}_{Ki} \|^2 \right\| \left\| \bar{Q}_K^{-1} \right\|^4 \left\| \mathbb{E} (T_{1n} T_{1n}' | Z) \right\| \\
\leq O_p(\psi_n) \left\| \mathbb{E} (T_{1n} T_{1n}' | Z) \right\| . \tag{75}
\]
\[
= o_p(1) \tag{76}
\]
the final bound by Lemma 5.1. Thus (72) is $o_p(1)$. 

34
Similarly for (73) but using the norm inequality $\|AB\| \leq \|A\| \|B\|$, 

\[
\left\| A'K\tilde{Q}_K^{-1} \frac{1}{n^2} \sum_{i=1}^{n} \tilde{x}_{Ki}\tilde{x}_{Ki}' \left( \tilde{x}_{Ki}'\tilde{Q}_K^{-1}T_{2n} \right)^2 \tilde{Q}_K^{-1}A_K \right\|
\leq \|A_K\|^2 \|\tilde{Q}_K^{-1}\|^4 \left\| \frac{1}{n^2} \sum_{i=1}^{n} \tilde{x}_{Ki}\tilde{x}_{Ki}' \left( \tilde{x}_{Ki}'\tilde{Q}_K^{-1}T_{2n} \right)^2 \right\|
\leq \|A_K\|^2 \|\tilde{Q}_K^{-1}\|^4 \left\| \frac{1}{n^2} \sum_{i=1}^{n} \tilde{x}_{Ki}\tilde{x}_{Ki}' \|\tilde{x}_{Ki}\|^2 \right\| \|T_{2n}\|^2
\leq O_p(\psi_n) \|T_{2n}\|^2
\leq o_p(1).
\]

the final bound by Lemma 5.2. This establishes that (73) is $o_p(1)$ and hence (70) holds.

To complete the proof we establish (71). Note that since $E(e_{Ki}^2|z_i) = \sigma_i^2 + r_{Ki}^2$, 

\[
\left\| E\left( \tilde{S}_K|Z_n \right) \right\| = \left\| \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{Ki}\tilde{x}_{Ki}' E\left( e_{Ki}^2|z_i \right) \right\| \leq \left\| S_K \right\| + \frac{1}{n} \sum_{i=1}^{n} \|x_{Ki}\|^2 r_{Ki}^2.
\]

Lemma 5.1 states that $\psi_n \left\| S_K \right\| = o_p(1)$, and Lemma 5.2 and Markov’s inequality implies that $\psi_n n^{-1} \sum_{i=1}^{n} \|x_{Ki}\|^2 r_{Ki}^2 = o_p(1)$. Together, we find that 

\[
\psi_n \left\| E\left( \tilde{S}_K|Z \right) \right\| = o_p(1).
\] (77)

Using the inequalities $\|A\| \leq \text{tr}A$ and $\text{tr}A \leq d \|A\|$ for positive semi-definite $A$, conditioning, the matrix norm inequality, Lemma 3.1, Theorem 4.3, and (77) 

\[
E\left( \left\| A'K \left( \tilde{Q}_K^{-1} - I_K \right) \tilde{S}_K \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right\| | Z \right) \leq \text{tr} \left( A'K \left( \tilde{Q}_K^{-1} - I_K \right) E\left( \tilde{S}_K|Z \right) \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right)
\leq d \left\| A'K \left( \tilde{Q}_K^{-1} - I_K \right) E\left( \tilde{S}_K|Z \right) \left( \tilde{Q}_K^{-1} - I_K \right) A_K \right\|
\leq d \left\| A_K \right\|^2 \left\| \tilde{Q}_K^{-1} - I_K \right\|^2 \left\| E\left( \tilde{S}_K|Z \right) \right\|
\leq O_p(\psi_n) \left\| E\left( \tilde{S}_K|Z \right) \right\|
= o(1).
\]

This implies (71) as needed. \[\square\]

11 Proof for Spline Approximation Theory
Proof of Theorem 7. (Sketch) We show that for a properly constructed weight function \( w(z) \),

\[
\inf_{\beta \in \mathbb{R}^K} \sup_{z \in \mathbb{Z}} \left| \frac{g(z) - x_K(z)\beta}{w(z)} \right| \leq CK^{-s(1-1/\alpha)} \tag{78}
\]

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

Recall that \( b = BK^{1/\alpha} \). Let \( \beta_K^* \) 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.

\[
\beta_K^* = \arg\min_{\beta \in \mathbb{R}^K} \sup_{|z| \leq b} \left| g(z) - x_K(z)\beta \right|.
\]

Set \( g_K(z) = x_K(z)\beta_K^* \). Since \( \sup_z |g^{(s)}(z)| \leq C \), 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 \( s^\text{th} \) order polynomial with coefficients bounded across segments. The assumption that \( \sup_z |g^{(s-1)}(z)| \leq C \) also implies the function \( g(z) \) can be globally bounded by a \( s^\text{th} \) 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 \( s^\text{th} \) 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

\[
\inf_{\beta \in \mathbb{R}^K} \sup_{z \in \mathbb{Z}} \left| \frac{g(z) - x_K(z)\beta}{w(z)} \right| \leq \sup_{z \in \mathbb{Z}} \left| \frac{g(z) - x_K(z)\beta \overline{y}}{w(z)} \right|
\]

\[
\leq \sup_{|z| \leq b} |w(z)|^{-1} CB^s K^{-s(1-1/\alpha)} + 2 \sup_{|z| > b} |z|^{-s(\alpha-1)}
\]

\[
\leq \left( a_0^{-1} CB^s + 2B^{-s(\alpha-1)} \right) K^{-s(1-1/\alpha)}.
\]

This is (78). Finally

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


