Chapter 4

Asymptotic Properties of the Substitution Versions of the Projective Scores

In this chapter the asymptotic consequences of having constructed a theory of estimating equations around Locally Minimum Variance Unbiased Estimation will be investigated. The question of interest is to evaluate the difference between the "optimal" score, which in general depends on the unknown λ_0 and the substitution version of the score, which is evaluated at some estimate of λ, which will usually be taken to be $\hat{\lambda}_\psi$.

The chapter can be crudely summarized, by saying that the U_1 substitution projective score suffers from an asymptotic bias, but by moving to the U_2 substitution projective score this asymptotic bias is removed. The value of sufficient bias removal will be taken up again in the next chapter, where its relevance will be displayed in the context of "rectangular array" asymptotics.

4.1 The strong substitution property

The first result to be presented provides a strong reason for using an estimative kernel, say $K(\psi, \lambda)$ that satisfies the $\mathcal{G}_1^r$ unbiasedness property, that is, it is orthogonal to the closed linear space generated by $\mathbf{V}_1$, and therefore, according to Corollary 3.1 of Theorem 3.2, has the property that $E_{\lambda_0}[\nabla_\lambda K(\psi, \lambda_0)] = 0$. It will be assumed that the estimative kernels in the space $\mathcal{G}_t^r$ behave like *i.i.d.* sums, so that if $K \in \mathcal{G}_t^r$ then since K is mean zero, and likewise its first t derivatives, it follows that $K_{i_1 i_2 \cdots i_k}/\sqrt{n} = O_p(1)$ for $k = 0, 1 \cdots t$. This assumption is reasonable since it is true for the first projective score U_1.

As a convention, when a statement is made about a vector or matrix, say for the vector Y, that $Y = o_p(1)$, then the statement is implied for *each* element of the vector or matrix.

Theorem 4.1 *Let $K(\psi,\lambda) \in \mathcal{G}_0{}^r$, and let $\tilde{\lambda}_\psi$ be any estimator of λ such that $(\tilde{\lambda}_\psi - \lambda) = o_p(n^{-\frac{1}{4}})$. Denoting the i^{th} component of the vector K by K^i assume that for each i, $i = 1, 2, \cdots, r$,*

1. $n^{-\frac{1}{2}} \nabla_\lambda K^i = O_p(1)$.
2. $n^{-1} |\nabla_{\lambda\lambda} K^i(\lambda)|$ *is bounded in a neighborhood of λ, by a function $H^i(\mathbf{x})$, such that $E[H^i(\mathbf{x})] < M^i$, and M^i is independent of λ.*

Then

$$\frac{1}{\sqrt{n}} \left(K(\psi, \tilde{\lambda}_\psi) - K(\psi, \lambda) \right) \xrightarrow{p} 0. \tag{4.1}$$

<u>Proof</u>

It is enough to show that the component-wise convergence holds in order to establish the convergence of the entire vector.

Consider a Taylor Series expansion of $\frac{1}{\sqrt{n}} \left(K^i(\psi, \tilde{\lambda}_\psi) - K^i(\psi, \lambda) \right)$ about λ, that is

$$\frac{1}{\sqrt{n}} \left(K^i(\psi, \tilde{\lambda}_\psi) - K^i(\psi, \lambda) \right) = \frac{1}{\sqrt{n}} (\tilde{\lambda}_\psi - \lambda)^T \nabla_\lambda K^i + \tfrac{1}{2} (\tilde{\lambda}_\psi - \lambda)^T \tfrac{1}{\sqrt{n}} \nabla_{\lambda\lambda} K^i(\lambda^*) (\tilde{\lambda}_\psi - \lambda), \tag{4.2}$$

where $|\lambda^* - \lambda| < |\tilde{\lambda}_\psi - \lambda|$.

Taking the first term on the RHS of (4.2), $\frac{1}{\sqrt{n}} (\tilde{\lambda}_\psi - \lambda)^T \nabla_\lambda K^i$, we have by assumption that $\frac{1}{\sqrt{n}} \nabla_\lambda K^i = O_p(1)$, and that $(\tilde{\lambda}_\psi - \lambda) = o_p(n^{-\frac{1}{4}})$, so that the product is $o_p(1)$.

For the second term, $(\tilde{\lambda}_\psi - \lambda)^T \frac{1}{\sqrt{n}} \nabla_{\lambda\lambda} K^i(\lambda^*)(\tilde{\lambda}_\psi - \lambda) = o_p(1)\frac{1}{n}\nabla_{\lambda\lambda}K^i(\lambda^*)o_p(1)$, and from assumption 2, $\frac{1}{n}\nabla_{\lambda\lambda}K^i(\lambda^*) = O_p(1)$. Consequently the second term is $o_p(1)$ as well, so that (4.2) goes in probability to zero, as required. ■

This result essentially says that the substitution version of the kernel possesses all the important properties of the original kernel. For example, if $K(\lambda)/\sqrt{n}$ is asymptotically normal, then so is $K(\tilde{\lambda}_\psi)/\sqrt{n}$. The result establishes a strong substitution property for the kernel U_t, in the sense that using the substitution version is asymptotically equivalent to using the optimal score.

Clearly Theorem 4.1 shows that no difference between the substitution and the original version of the kernel can be discerned at this level of resolution. In order to locate where they differ, the resolution must be increased. That is, the difference between $K(\lambda)$ and $K(\tilde{\lambda}_\psi)$ must be looked at on the level where a limiting distribution exists for the difference. To achieve this, (4.1) will be multiplied by a factor of $\sqrt{n}$, and the limiting distribution of the quantities

$$U_1(\psi, \hat{\lambda}_\psi) - U_1(\psi, \lambda_0) \tag{4.3}$$

$$U_2(\psi, \hat{\lambda}_\psi) - U_2(\psi, \lambda_0) \tag{4.4}$$

will be considered. Here, $\hat{\lambda}_\psi$ is the maximum likelihood estimate of λ for ψ fixed, and λ_0 is the true value of λ. As usual, the appropriate regularity conditions will be assumed, and an *i.i.d.* sampling context used.

4.2 The U_1 projective score for univariate ψ

An important result needed for the proofs of this section will be an asymptotically valid expression for the term $\hat{\lambda}^i_\psi - \lambda^i_0$. Recall here, that the superscript i denotes the i^{th} component of the vector λ. The relevant result is

Proposition 4.1

$$\hat{\lambda}^i_\psi - \lambda^i_0 = W^{i,j} V_j / n + O_p(n^{-1}),$$

or in matrix notation

$$\hat{\lambda}_\psi - \lambda_0 = W^{-1}\mathbf{V}_1 / n + O_p(n^{-1}).$$

This result is taken from McCullagh (1987, p.209,7.10). The matrix W is the per observation covariance matrix between the functions V_i^k and V_j^k and does not depend on k because of the "identically distributed" assumption, see (3.5).

The following result is required in order to apply the classical limiting distribution theorem results, to the terms in the Taylor Series expansions of (4.3) and (4.4).

Lemma 4.1 *Denoting the c^{th} element of the vector U_1 by U_1^c, then U_1^c is an i.i.d. sum, and likewise each component of $\nabla_\lambda U_1^c$, and $\nabla_{\lambda\lambda} U_1^c$.*

Proof

Using the definition of U_1^c (see 2.16):

$$U_1^c = U_0^c - E[U_0^c \mathbf{V}_1^T](E[\mathbf{V}_1 \mathbf{V}_1^T])^{-1}\mathbf{V}_1. \tag{4.5}$$

Since both U_0^c and $\mathbf{V}_1^T$ are *i.i.d.* sums, their covariance can be written as $n\, E[U_0^{ck}\mathbf{V}_1^{kT}] = n\, T^c$ say. The vector with i^{th} element $E[U_0^{ck} V_i^k]$ will be denoted by T_i^c, so that (4.5) can be written as

$$U_1^c = U_0^c - T_i^c W^{i,j} V_j. \tag{4.6}$$

Expressing both U_0^c and V_j explicitly as *i.i.d.* sums gives

$$\begin{aligned} U_1^c &= \sum_k U_0^{ck} - T_i^c W^{i,s} \sum_k V_s^k \\ &= \sum_k (U_0^{ck} - T_i^c W^{i,s} V_s^k), \end{aligned} \tag{4.7}$$

which establishes that U_1^c is an *i.i.d.* sum. Differentiating (4.7) with respect to any λ^j will result in a quantity that is again an *i.i.d.* sum, likewise for $\nabla_{\lambda\lambda}U_1^c$ and any higher order derivatives. ■

In order to find the limiting distribution of $U_1(\psi, \hat{\lambda}_\psi) - U_1(\psi, \lambda_0)$, it is initially useful to consider the case where $\dim(\psi) = 1$, the extension to multivariate ψ being accomplished through the use of the Cramér-Wold device. However, the superscript c will still be maintained in order to facilitate the move to the case where ψ is multivariate.

The following assumptions concerning the derivatives of the log-likelihood will be made:

1. Regularity conditions R(3) hold, (see Section 4, Chapter 2).
2. Denoting the derivative of U_1^{ck} with respect to $\lambda^{i_1}, \lambda^{i_2}$ and λ^{i_3} as $U_{1,i_1 i_2 i_3}^{ck}$, then for all λ in a neighborhood of λ_0, the function $|U_{1,i_1 i_2 i_3}^{ck}| < H^c(x_k)$, where $E[H^c(x_k)] < M^c$, and M^c is a constant independent of λ. This is the usual bounded third derivative assumption.

Theorem 4.2 *Under the previously stated regularity conditions,*

$$U_1^c(\psi, \hat{\lambda}_\psi) - U_1^c(\psi, \lambda_0) \xrightarrow{\ell} Y^c,$$

where

$$Y^c = \left(Z_1^T, Z_{2c}^T\right) \begin{pmatrix} \frac{1}{2}W^{-1}T^{c*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_{2r} \end{pmatrix},$$

and

$$\begin{pmatrix} Z_1 \\ Z_{2c} \end{pmatrix} \sim N_{2s}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W & -T^{c*} \\ -T^{c*} & \Sigma^c \end{pmatrix}\right].$$

The matrix Σ^c *is defined as* $E[\nabla_\lambda U_1^{ck}(\nabla_\lambda U_1^{ck})^T]$, *and* T^{c*} *as the matrix with* $(i,j)^{th}$ *element equal to* $E[U_0^{ck} V_{ij}^{k*}]$. *Note that both these matrices are symmetric.*

This result shows that the limiting distribution of $U_1^c(\psi, \hat{\lambda}_\psi) - U_1^c(\psi, \lambda_0)$ is a homogeneous central quadratic form in normal random variables.

Proof

Expanding the term $U_1^c(\hat{\lambda}_\psi) - U_1^c(\lambda_0)$ in a Taylor series about λ_0 yields

$$U_1^c(\hat{\lambda}_\psi) - U_1^c(\lambda_0) = (\hat{\lambda}_\psi - \lambda_0)^T \nabla_\lambda U_1^c + \tfrac{1}{2}(\hat{\lambda}_\psi - \lambda_0)^T \nabla^2_{\lambda\lambda} U_1^c (\hat{\lambda}_\psi - \lambda_0) + R_n,$$

where

$$R_n = \tfrac{1}{6} \textstyle\sum_{i=1}^s \sum_{j=1}^s \sum_{k=1}^s U_{1ijk}^c(\lambda^*)(\hat{\lambda}_\psi^i - \lambda_0^i)(\hat{\lambda}_\psi^j - \lambda_0^j)(\hat{\lambda}_\psi^k - \lambda_0^k),$$

and $|\lambda^* - \lambda_0| < |\hat{\lambda}_\psi - \lambda_0|$.

Substituting in $\hat{\lambda}_\psi - \lambda_0 = \frac{1}{n}\mathbf{V}_1^T W^{-1} + O_p(n^{-1})$ the relevant terms in the limiting distribution are

$$\tfrac{1}{n}\mathbf{V}_1^T W^{-1} \nabla_\lambda U_1^c \tag{4.8}$$

and

$$\tfrac{1}{2}\tfrac{1}{n}\mathbf{V}_1^T W^{-1} \nabla_{\lambda\lambda} U_1^c W^{-1} \tfrac{1}{n}\mathbf{V}_1. \tag{4.9}$$

In order to justify the statement that these are the only asymptotically relevant relevant terms it must be shown that the following quantities are $o_p(1)$:

$$O_p(n^{-1})\nabla_\lambda U_1^c$$

$$O_p(n^{-1})\nabla_{\lambda\lambda} U_1^c \tfrac{1}{n}\mathbf{V}_1^T W^{-1}$$

$$O_p(n^{-2})\nabla_{\lambda\lambda} U_1^c$$

$$R_n.$$

These are established by using the facts that $\nabla_\lambda U_1^c$ and $\mathbf{V}_1$ are mean zero *i.i.d.* sums and that the law of large numbers holds for $\frac{1}{n}\nabla_{\lambda\lambda}U_1^c$. The fact that R_n is $o_p(1)$ is a direct consequence of the bounded third derivative assumption.

Now consider the (i,j) element of the matrix $\nabla_{\lambda\lambda}U_1^c$. Denoting it as $U_{1,ij}^c$ and using the fact that it is the *i.i.d.* sum $\sum_k U_{1,ij}^{ck}$, then since U_1^{ck} is orthogonal to the Bhattacharyya basis $\{\mathbf{V}_1^k\}$, Theorem 3.2 implies that

$$E[U_{1,ij}^{ck}] = E[U_1^{ck}V_{ij}^k],$$

but $E[U_1^{ck}V_{ij}^k] = E[U_1^{ck}V_{ij}^{k*}]$ where V_{ij}^{k*} is the orthogonalized version of V_{ij}^k, so that $E[U_1^{ck}V_{ij}^k] = E[U_0^{ck}V_{ij}^{k*}]$. Applying the Weak Law of Large Numbers to $\frac{1}{n}U_{1,ij}^c$ shows that

$$\tfrac{1}{n}U_{1,ij}^c = \tfrac{1}{n}\textstyle\sum_k U_{1,ij}^{ck} \xrightarrow{p} E[U_0^{ck}V_{ij}^{k*}],$$

so that $\frac{1}{n}\nabla_{\lambda\lambda}U_1^c \xrightarrow{p} T^{c*}$.

Now using Slutsky's theorem it follows that

$$\tfrac{1}{2}\tfrac{1}{n}\mathbf{V}_1^T W^{-1}\nabla_{\lambda\lambda}U_1^c W^{-1}\tfrac{1}{n}\mathbf{V}_1 = \tfrac{1}{2}\tfrac{1}{n}\mathbf{V}_1^T W^{-1}T^{c*}W^{-1}\mathbf{V}_1 + o_p(1),$$

where T^{c*} has been substituted for $\frac{1}{n}\nabla_{\lambda\lambda}U_1^c$.

Consequently the limiting distribution of $U_1^c(\psi,\hat{\lambda}_\psi) - U_1^c(\psi,\lambda_0)$ is obtained from

$$\tfrac{1}{\sqrt{n}}\mathbf{V}_1^T W^{-1}\tfrac{1}{\sqrt{n}}\nabla_\lambda U_1^c + \tfrac{1}{2}\tfrac{1}{\sqrt{n}}\mathbf{V}_1^T W^{-1}T^{c*}W^{-1}\tfrac{1}{\sqrt{n}}\mathbf{V}_1,$$

which can be expressed in matrix form as

$$\left(\tfrac{1}{\sqrt{n}}\mathbf{V}_1^T, \tfrac{1}{\sqrt{n}}(\nabla_\lambda U_1^c)^T\right)\begin{pmatrix} \frac{1}{2}W^{-1}T^{c*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{n}}\mathbf{V}_1 \\ \frac{1}{\sqrt{n}}\nabla_\lambda U_1^c \end{pmatrix}.$$

Since the vectors $\mathbf{V}_1$ and $\nabla_\lambda U_1^c$ have elements that are mean 0 *i.i.d.* sums, an application of the Multivariate Central limit theorem gives

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\mathbf{V}_1 \\ \frac{1}{\sqrt{n}}\nabla_\lambda U_1^c \end{pmatrix} \xrightarrow{\ell} N_{2r}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W & E[\mathbf{V}_1^k(\nabla_\lambda U_1^{ck})^T] \\ E[\nabla_\lambda U_1^{ck}\mathbf{V}_1^{kT}] & E[\nabla_\lambda U_1^{ck}(\nabla_\lambda U_1^{ck})^T] \end{pmatrix}\right].$$

Now consider the $(i,j)^{th}$ element of the matrix $E[\mathbf{V}_1^k(\nabla_\lambda U_1^{ck})^T]$, that is $E[V_i^k U_{1,j}^{ck}]$. An application of Theorem 3.2 shows that this expectation is equal to $-E[V_{ij}^k U_1^{ck}]$ or equivalently $-E[V_{ij}^{k*}U_0^{ck}]$, so that $E[\mathbf{V}_1^k(\nabla_\lambda U_1^{ck})^T] = -T^{c*}$. Consequently the limiting normal distribution is

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\mathbf{V}_1 \\ \frac{1}{\sqrt{n}}\nabla_\lambda U_1^c \end{pmatrix} \xrightarrow{\ell} N_{2s}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W & -T^{c*} \\ -T^{c*} & \Sigma^c \end{pmatrix}\right].$$

Applying the continuous function theorem (Serfling, 1980, p.25,Corollary) to the vector $\left(\frac{1}{\sqrt{n}}\mathbf{V}_1, \frac{1}{\sqrt{n}}\nabla_\lambda U_1^c\right)$ implies that

$$\left(\tfrac{1}{\sqrt{n}}\mathbf{V}_1^T, \tfrac{1}{\sqrt{n}}(\nabla_\lambda U_1^c)^T\right)\begin{pmatrix} \frac{1}{2}W^{-1}T^{c*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{n}}\mathbf{V}_1 \\ \frac{1}{\sqrt{n}}\nabla_\lambda U_1^c \end{pmatrix} \xrightarrow{\ell}$$

$$\left(Z_1^T, Z_{2c}^T\right)\begin{pmatrix} \frac{1}{2}W^{-1}T^{c*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix}\begin{pmatrix} Z_1 \\ Z_{2c} \end{pmatrix},$$

where

$$\begin{pmatrix} Z_1 \\ Z_{2c} \end{pmatrix} \sim N_{2s}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W & -T^{c*} \\ -T^{c*} & \Sigma^c \end{pmatrix}\right],$$

which proves the result. ■

Since the limiting distribution is a quadratic form in normal random variables, the mean and variance of the limiting random variable are easily established:

Theorem 4.3 *Let Y^c be defined as in Theorem 4.2, then*

$$
\begin{aligned}
E[Y^c] &= -\tfrac{1}{2} tr(T^{c*} W^{-1}). && (4.10)\\
\mathrm{Var}[Y^c] &= tr\left(\left[\Sigma^c - \tfrac{1}{2} T^{c*} W^{-1} T^{c*}\right] W^{-1}\right). && (4.11)
\end{aligned}
$$

Proof

The distribution of Y^c is a quadratic form in normal random variables, it is also a "central" quadratic form, by which is meant that the variables Z_1 and Z_{2c} are both mean zero. Consequently, the cumulants of Y^c have a particularly simple form: Denoting the variance-covariance matrix of the random variable (Z_1, Z_{2c}) by B, and the matrix

$$
\begin{pmatrix} \frac{1}{2} W^{-1} T^{c*} W^{-1} & \frac{1}{2} W^{-1} \\ \frac{1}{2} W^{-1} & 0 \end{pmatrix}
$$

by A, then the κ_s cumulant of Y^c is given by

$$
\kappa_s = 2^{s-1}(s-1)! \mathrm{tr}(B\,A)^s \tag{4.12}
$$

(see, Johnson and Kotz (1970, p.153, (22))). Since the mean is the first cumulant, κ_1, and the variance is the second cumulant κ_2, all that is required to calculate them is matrix multiplication. To this end:

$$
B\,A = \frac{1}{2} \begin{pmatrix} 0 & I \\ \left(\Sigma^c - T^{c*} W^{-1} T^{c*}\right) W^{-1} & -T^{c*} W^{-1} \end{pmatrix},
$$

where I denotes an $s \times s$ identity matrix, and

$$
B\,A\,B\,A = \frac{1}{4} \begin{pmatrix} \left(\Sigma^c - T^{c*} W^{-1} T^{c*}\right) W^{-1} & -T^{c*} W^{-1} \\ -\left(\Sigma^c - T^{c*} W^{-1} T^{c*}\right)\left(W^{-1} - T^{c*} W^{-1} T^{c*}\right) & \Sigma^r W^{-1} \end{pmatrix}.
$$

Taking the traces of these matrices and using (4.12) provides the result. ∎

The important fact being demonstrated here is that the asymptotic distribution of $U_1^c(\psi, \hat{\lambda}_\psi) - U_1^c(\psi, \lambda_0)$ is not mean zero, which is the root cause of the bias in the *m.l.e.* of ψ. The expression derived for the asymptotic bias can be shown to be identical to that derived by McCullagh and Tibshirani (1990, (7) and (9)).

4.3 Multivariate extensions

Since U_1 and U_2 are vectors, the limiting distributions of $U_1(\hat{\lambda}_\psi) - U_1(\lambda_0)$ and $U_2(\hat{\lambda}_\psi) - U_2(\lambda_0)$ are conveniently dealt with using the Cramér-Wold device. The relevant theorem states (see Billingsley (1968, Theorem 7.7, p.49)): *"In $\mathbb{R}^k$, X_n converges in distribution to X if and only if each linear combination of X_n converges in distribution to the corresponding linear combination of the components of X".*

Define the r dimensional vector $\mathbf{X}$ to be a central multivariate quadratic form in normal random variables if

$$X_i = Z_i^T \Xi_i Z_i, \quad i = 1, 2, \cdots, r, \tag{4.13}$$

where the vector $(Z_1^T, Z_2^T, \cdots, Z_r^T)$ has a mean zero multivariate normal distribution

$$\left(Z_1^T, Z_2^T, \cdots Z_r^T,\right) \sim N\left[0, \Omega\right].$$

As a particular instance of the distribution consider the case where $Z_i = (Z_1, Z_{2i})^T$ and the vector of normals $(Z_1, Z_{21}, Z_{22}, Z_{23}, \cdots, Z_{2r})^T$ has variance-covariance matrix

$$\begin{pmatrix} W & T^{1*} & T^{2*} & \cdots & T^{r*} \\ T^{1*} & \Sigma^1 & E[\nabla_\lambda U_1^{1k}(\nabla_\lambda U_1^{2k})^T] & \cdots & E[\nabla_\lambda U_1^{1k}(\nabla_\lambda U_1^{rk})^T] \\ T^{2*} & E[\nabla_\lambda U_1^{2k}(\nabla_\lambda U_1^{1k})^T] & \Sigma^2 & \cdots & E[\nabla_\lambda U_1^{2k}(\nabla_\lambda U_1^{rk})^T] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T^{r*} & E[\nabla_\lambda U_1^{rk}(\nabla_\lambda U_1^{1k})^T] & E[\nabla_\lambda U_1^{rk}(\nabla_\lambda U_1^{2k})^T] & \cdots & \Sigma^r \end{pmatrix} \tag{4.14}$$

with Y_i defined as

$$Y_i = Z_i^T \begin{pmatrix} \frac{1}{2}W^{-1}T^{i*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix} Z_i, \tag{4.15}$$

so that

$$\Xi_i = \begin{pmatrix} \frac{1}{2}W^{-1}T^{i*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix}.$$

The random vector with i^{th} component given by (4.15) will be denoted by $\mathbf{Y}$. Consider an arbitary linear combination of the components of $\mathbf{Y}$, say $a^T\mathbf{Y}$. Then

$$a^T\mathbf{Y} = \sum_{i=1}^{r} Z_i^T a_i \begin{pmatrix} \frac{1}{2}W^{-1}T^{i*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix} Z_i. \tag{4.16}$$

This can be written as a central univariate quadratic form. For example if $r = 2$ then (4.16) is equal to

$$(Z_1^T, Z_{21}^T, Z_{22}^T) \begin{pmatrix} \sum_{i=1}^{2} \frac{1}{2}a_iW^{-1}T^{i*}W^{-1} & \frac{1}{2}a_1W^{-1} & \frac{1}{2}a_2W^{-1} \\ \frac{1}{2}a_1W^{-1} & 0 & 0 \\ \frac{1}{2}a_2W^{-1} & 0 & 0 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_{21} \\ Z_{22} \end{pmatrix}, \tag{4.17}$$

where the variance-covariance matrix of $(Z_1^T, Z_{21}^T, Z_{22}^T)$ is given by (4.14).

Now consider the linear combination

$$a^T(U_1(\hat{\lambda}_\psi) - U_1(\lambda_0)).$$

If this can be shown to have the same limiting distribution as $a^T\mathbf{Y}$ where $\mathbf{Y}$ is defined in (4.15), then by the Cramér-Wold device, the limiting distribution of $U_1(\hat{\lambda}_\psi) - U_1(\lambda_0)$ must be the same as the distribution of $\mathbf{Y}$.

To establish that this is in fact the case, a Taylor series expansion of $a^T(U_1(\hat{\lambda}_\psi) - U_1(\lambda_0))$ about λ_0 will suffice.

Theorem 4.4 *Assume that the regularity conditions of Theorem 4.2, hold for each component of the vector U_1. Denoting the components of the vector Y by Y_i, $i = 1, 2\cdots r$, and define*

$$Y_i = Z_i^T \begin{pmatrix} \frac{1}{2}W^{-1}T^{i*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix} Z_i,$$

where the $Z_i = (Z_1, Z_{2i})^T$ are vector valued mean zero normal random variables, with the variance-covariance matrix of $(Z_1, Z_{21}, Z_{22}, \cdots, Z_{2r})$ as

$$\begin{pmatrix} W & T^{1*} & T^{2*} & \cdots & T^{r*} \\ T^{1*} & \Sigma^1 & E[\nabla_\lambda U_1^{1k}(\nabla_\lambda U_1^{2k})^T] & \cdots & E[\nabla_\lambda U_1^{1k}(\nabla_\lambda U_1^{rk})^T] \\ T^{2*} & E[\nabla_\lambda U_1^{2k}(\nabla_\lambda U_1^{1k})^T] & \Sigma^2 & \cdots & E[\nabla_\lambda U_1^{2k}(\nabla_\lambda U_1^{rk})^T] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T^{r*} & E[\nabla_\lambda U_1^{rk}(\nabla_\lambda U_1^{1k})^T] & E[\nabla_\lambda U_1^{rk}(\nabla_\lambda U_1^{2k})^T] & \cdots & \Sigma^r \end{pmatrix}.$$

Then

$$U_1(\hat{\lambda}_\psi) - U_1(\lambda_0) \xrightarrow{\ell} Y,$$

where Y is defined as above.

__**Proof**__

For an arbitary vector $a \in \mathbb{R}^r$, consider a Taylor series expansion of $a^T(U_1(\hat{\lambda}_\psi) - U_1(\lambda_0))$ about λ_0. That is

$$\sum_{i=1}^r a_i(U_1^i(\hat{\lambda}_\psi) - U_1^i(\lambda_0)) = \sum_i a_i(\hat{\lambda}_\psi - \lambda_0)^T \nabla_\lambda U_1^i + \sum_i a_i(\hat{\lambda}_\psi - \lambda_0)^T \nabla_{\lambda\lambda} U_1^i(\hat{\lambda}_\psi - \lambda_0) + R_n.$$

Applying the same argumants used as in Theorem 4.2 establishes that

$$\sum_i a_i(\hat{\lambda}_\psi - \lambda_0)^T \nabla_\lambda U_1^i + \sum_i a_i(\hat{\lambda}_\psi - \lambda_0)^T \nabla_{\lambda\lambda} U_1^i (\hat{\lambda}_\psi - \lambda_0) =$$
$$\tfrac{1}{\sqrt{n}}\mathbf{V}_1^T W^{-1} \textstyle\sum_i a_i \tfrac{1}{\sqrt{n}} \nabla_\lambda U_1^i + \tfrac{1}{2}\tfrac{1}{\sqrt{n}}\mathbf{V}_1^T W^{-1} \textstyle\sum_i T^{i*} W^{-1} \tfrac{1}{\sqrt{n}}\mathbf{V}_1 + o_p(1), \quad (4.18)$$

which can be written in matrix form as

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\mathbf{V}_1 \\ \frac{1}{\sqrt{n}}U_1^i \\ \frac{1}{\sqrt{n}}U_1^2 \\ \vdots \\ \frac{1}{\sqrt{n}}U_1^r \end{pmatrix}^T \begin{pmatrix} \frac{1}{2}W^{-1}\sum_i a_i T^{i*} W^{-1} & \frac{1}{2}a_1 W^{-1} & \frac{1}{2}a_2 W^{-1} & \cdots & \frac{1}{2}a_r W^{-1} \\ \frac{1}{2}a_1 W^{-1} & 0 & 0 & \cdots & 0 \\ \frac{1}{2}a_2 W^{-1} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_r W^{-1} & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}}\mathbf{V}_1 \\ \frac{1}{\sqrt{n}}U_1^i \\ \frac{1}{\sqrt{n}}U_1^2 \\ \vdots \\ \frac{1}{\sqrt{n}}U_1^r \end{pmatrix} + o_p(1).$$

Applying the Multivariate Central Limit Theorem to the vector of random variables $(\frac{1}{\sqrt{n}}\mathbf{V}_1, \frac{1}{\sqrt{n}}U_1^1, \frac{1}{\sqrt{n}}U_1^2, \cdots, \frac{1}{\sqrt{n}}U_1^r)$ establishes that it has the same limiting distribution as the random variables $(Z_1, Z_{21}, Z_{22}, \cdots, Z_{2r})$. So by the Continuous Function Theorem the limiting distribution of (4.18) is the same as that of $a^T Y$. The theorem will follow by an aplication of the Cramér-Wold device so long as the remander term R_n is negligible, that is $R_n = o_p(1)$. But this follows from the bounded third derivative assumption. ∎

4.4 The second projective score, U_2

In this section the limiting distribution of $U_2(\hat{\lambda}_\psi) - U_2(\lambda_0)$ will be investigated. In order to do this a series of technical lemmas are needed; their proofs have been relegated to the last section of the chapter, in order to keep some form of continuity in the presentation of results. The result appears in a slightly weaker version in Small and McLeish (1989).

The lemmas are as follows:

Lemma 4.2 *Differentiation of the coefficients in the Extended Information Matrix,* W_{v_1,v_2}*, can at most decrease their order of magnitude.*

The next lemma presents an asymptotic approximation to the derivatives of the second-order elements in the Bhattacharyya basis, given in terms of the first order elements.

Lemma 4.3

$$\frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij}^*\right]\right] = \frac{1}{\sqrt{n}}\left[-\left(W_{j,k}V_i + W_{i,k}V_j\right)\right] + o_p(1).$$

The following result is very interesting: it states that from a stochastic viewpoint the derivative of the difference between the first and second projective scores is a linear combination of the elements of $\mathbf{V}_1$. So that even though the functions U_1 and U_2 are orthogonal to $\mathbf{V}_1$, the derivative of their difference approximately lies in the space generated by $\{\mathbf{V}_1\}$.

Lemma 4.4

$$\frac{1}{\sqrt{n}}\left[\nabla_\lambda\left(U_2^c - U_1^c\right) - T^{c*}W^{-1}\mathbf{V}_1\right] = o_p(1).$$

The final lemma proves that even though $U_{2,ij}^c$ is not itself an *i.i.d.* sum, a law of large numbers holds for it, and the value of the limiting constant is 0.

Lemma 4.5

$$\lim_{n\to\infty}\frac{U_{2,ij}^c}{n} \xrightarrow{p} 0. \tag{4.19}$$

These lemmas are all used in the proof of Theorem 4.5, which is the analog of Theorem 4.2 for the function U_2:

Theorem 4.5 *Under the previously stated regularity conditions,*

$$U_2^c(\psi, \hat{\lambda}_\psi) - U_2^c(\psi, \lambda_0) \xrightarrow{\ell} Y^c,$$

where

$$Y^c = \left(Z_1^T, Z_{2r}^T\right) \begin{pmatrix} W^{-1}T^{c*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_{2r} \end{pmatrix},$$

and

$$\begin{pmatrix} Z_1 \\ Z_{2r} \end{pmatrix} \sim N_{2r}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W & -T^{c*} \\ -T^{c*} & \Sigma^c \end{pmatrix}\right].$$

The matrices Σ^c and T^{c*} are defined as $E[\nabla_\lambda U_1^{ck}(\nabla_\lambda U_1^{ck})^T]$, and $E[U_0^{ck}V_{ij}^{k*}]$ respectively.

Note that the only difference between this limiting distribution and the limiting distribution of $U_1^c(\psi, \hat{\lambda}_\psi) - U_1^c(\psi, \lambda_0)$ is that the upper left matrix in the quadratic form is $W^{-1}T^{c*}W^{-1}$ rather than $\frac{1}{2}W^{-1}T^{c*}W^{-1}$! However, this apparently minor difference has some very relevant consequences:

Theorem 4.6 *Let Y^c be defined as in Theorem 4.5, then*

$$E[Y^c] = 0. \tag{4.20}$$

$$\mathrm{Var}[Y^c] = tr\left(\left[\Sigma^c - T^{c*}W^{-1}T^{c*}\right] W^{-1}\right). \tag{4.21}$$

The critical fact is that the limiting distribution of $U_2^c(\psi, \hat{\lambda}_\psi) - U_2^c(\psi, \lambda_0)$ has zero mean, so that $U_2(\hat{\lambda}_\psi)$ will provide a consistent root.

<u>Proof</u> of Theorem 4.5.

Taylor series expansion of $U_2^c(\psi, \hat{\lambda}_\psi) - U_2^c(\psi, \lambda_0)$ about λ_0 yields

$$U_2^c(\hat{\lambda}_\psi) - U_2^c(\lambda_0) = (\hat{\lambda}_\psi - \lambda_0)^T \nabla_\lambda U_2^c + \tfrac{1}{2}(\hat{\lambda}_\psi - \lambda_0)^T \nabla^2_{\lambda\lambda} U_2^c (\hat{\lambda}_\psi - \lambda_0) + R_n,$$

where

$$R_n = \frac{1}{6}\sum_{i=1}^{s}\sum_{j=1}^{s}\sum_{k=1}^{s} U^c_{2ijk}(\lambda^*)(\hat{\lambda}^i_\psi - \lambda^i_0)(\hat{\lambda}^j_\psi - \lambda^j_0)(\hat{\lambda}^k_\psi - \lambda^k_0),$$

and $|\lambda^* - \lambda_0| < |\hat{\lambda}_\psi - \lambda_0|$.

Substituting in $\hat{\lambda}_\psi - \lambda_0 = \frac{1}{n}\mathbf{V}_1^T W^{-1} + O_p(n^{-1})$ the relevant terms in the limiting distribution are

$$\frac{1}{n}\mathbf{V}_1^T W^{-1}\nabla_\lambda U_2^c, \tag{4.22}$$

and

$$\frac{1}{2}\frac{1}{n}\mathbf{V}_1^T W^{-1}\nabla_{\lambda\lambda} U_2^c W^{-1}\frac{1}{n}\mathbf{V}_1. \tag{4.23}$$

This statement is again justified by showing that the following quantities are $o_p(1)$:

$$O_p(n^{-1})\nabla_\lambda U_2^c$$
$$O_p(n^{-1})\nabla_{\lambda\lambda} U_2^c \frac{1}{n}\mathbf{V}_1^T W^{-1}$$
$$O_p(n^{-2})\nabla_{\lambda\lambda} U_2^c$$
$$R_n.$$

They are established by using the facts that $\nabla_\lambda U_2^c$ and $\mathbf{V}_1$ are mean zero *i.i.d.* sums and that the law of large numbers holds for $\frac{1}{n}\nabla_{\lambda\lambda} U_2^c$. The fact that R_n is $o_p(1)$ is a direct consequence of the bounded third derivative assumption.

Concentrating on the term (4.22), and using the result of Lemma 4.4, indicates that it can be re-expressed as

$$\frac{1}{\sqrt{n}}\mathbf{V}_1^T W^{-1}\frac{1}{\sqrt{n}}\nabla_\lambda U_1^c + \frac{1}{\sqrt{n}}\mathbf{V}_1^T W^{-1} T^{c*} W^{-1}\frac{1}{\sqrt{n}}\mathbf{V}_1 + o_p(1). \tag{4.24}$$

By using Lemma 4.19 it is clear that (4.23) is $o_p(1)$, so that the limiting distribution is determined by (4.24). This can be expressed in matrix form as

$$\left(\frac{1}{\sqrt{n}}\mathbf{V}_1^T, \frac{1}{\sqrt{n}}(\nabla_\lambda U_1^c)^T\right)\begin{pmatrix} W^{-1}T^{c*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{n}}\mathbf{V}_1 \\ \frac{1}{\sqrt{n}}\nabla_\lambda U_1^c \end{pmatrix}.$$

Exactly the same arguments that were used to prove Theorem 1, can be re-employed now to establish the result. ■

<u>Proof</u> of Theorem 4.

Using the formula for the mean and variance of a central quadratic form we have:

$$\begin{pmatrix} W & -T^{c*} \\ -T^{c*} & \Sigma^c \end{pmatrix}\begin{pmatrix} W^{-1}T^{c*}W^{-1} & \frac{1}{2}W^{-1} \\ \frac{1}{2}W^{-1} & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}T^{c*}W^{-1} & \frac{1}{2}I \\ \frac{1}{2}(\Sigma^c - 2T^{c*}W^{-1}T^{c*})W^{-1} & -\frac{1}{2}T^{c*}W^{-1} \end{pmatrix}$$

The mean is simply the trace of the previous matrix, which is 0. The variance follows in a similar fashion using the formula in (4.12).

The limiting distribution of $U_2^c(\psi, \hat{\lambda}_\psi) - U_2^c(\psi, \lambda_0)$ for multivariate ψ can be handled using the Cramér-Wold device. Since this presents no new ideas or difficulties the result will not be explicitly presented, but suffice to say that the multivariate limiting distribution is the same as that presented in (4.13),(4.14) and (4.15), with the matrix $\frac{1}{2}W^{-1}T^{i*}W^{-1}$ in (4.15) being replaced by $W^{-1}T^{i*}W^{-1}$.

It is of interest to examine the form of the correction term to the U_1 projective score, that results in the U_2 projective score. First note that since the nuisance parameters are evaluated at $\hat{\lambda}_\psi$, $\mathbf{V}_1(\hat{\lambda}_\psi) = 0$, since by definition, $\hat{\lambda}_\psi$ is the root of the equation $\mathbf{V}_1(\lambda) = 0$. Consequently

$$U_2(\hat{\lambda}_\psi) = U_0 - E[U_0\mathbf{V}_2^{*T}]E[\mathbf{V}_2^*\mathbf{V}_2^{*T}]\mathbf{V}_2\Big|_{\lambda=\hat{\lambda}_\psi}.$$

The $(i,j)^{th}$ element of $\mathbf{V}_2$ is by definition V_{ij}, but using the recurrance relation (Theorem 3.1), $V_{ij} = \frac{\partial}{\partial\lambda^j}V_i + V_iV_j$, so that $V_{ij}|_{\lambda=\hat{\lambda}_\psi} = \frac{\partial}{\partial\lambda^j}V_i|_{\lambda=\hat{\lambda}_\psi}$. Therefore the correction term is simply a matrix multiple of the second derivative of the log-likelihood.

Throughout this chapter the matrix $T^{c*}W^{-1}$ has made a frequent appearance. Its trace, a scalar quantity, which appears in the bias of the limiting distribution of $U_1(\hat{\lambda}_\psi) - U_1(\lambda_0)$, is a contraction of a tensor, and therefore invariant under transformations of the nuisance parameters. Chapter 6, will reveal this quantity to be a certain model curvature, which therefore plays a key role in the asymptotic theory of the substitution projective scores.

4.5 Technical lemmas

The first of these lemmas, Lemma 4.2 states that:

Differentiation of the coefficients in the Extended Information Matrix, $W_{\upsilon_1,\upsilon_2}$, can at most decrease their order of magnitude.

Proof Any element in the extended information matrix can be expressed as a polynomial in n, see for example (3.10). Consequently a term of order t, $W_{\upsilon_1,\upsilon_2}$ can be written as

$$W_{\upsilon_1,\upsilon_2} = \sum_{i=0}^{t} a_i n^i,$$

for some coefficients a_i not a function of n. Differentiating this polynomial establishes that the order of magnitude cannot increase, but may possibly decrease if $a_t' = 0$. ■

Lemma 4.3 states that:

$$\frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij}^*\right]\right] = \frac{1}{\sqrt{n}}\left[-\left(W_{j,k}V_i + W_{i,k}V_j\right)\right] + o_p(1).$$

Proof

By definition

$$\begin{aligned}
\frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij}^*\right]\right] &= \frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij} - E[V_{ij}\mathbf{V}_1^T]E[\mathbf{V}_1\mathbf{V}_1^T]^{-1}\mathbf{V}_1\right]\right] \\
&= \frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij} - W_{ij,r}W^{r,s}V_s\right]\right] \\
&= \frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij}\right]\right] - \frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[W_{ij,r}W^{r,s}V_s\right]\right].
\end{aligned}$$

The coefficient $W_{ij,r}W^{r,s}$ is not a function of n, so that it is $O(1)$, and likewise its derivative is $O(1)$. Since both V_s and $\frac{\partial}{\partial\lambda^k}V_s$ are *i.i.d.* sums it follows that

$$\frac{\partial}{\partial\lambda^k}(W_{ij,r}W^{r,s})\frac{1}{n^{\frac{3}{2}}}V_s \xrightarrow{p} 0,$$

and

$$W_{ij,r}W^{r,s}\frac{1}{n^{\frac{3}{2}}}\frac{\partial}{\partial\lambda^k}V_s \xrightarrow{p} 0,$$

which establishes that

$$\frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij}^*\right]\right] = \frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij}\right]\right] + o_p(1).$$

Now using the definition of V_{ij}:

$$\begin{aligned}
\frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[V_{ij}\right]\right] &= \frac{1}{n^{\frac{3}{2}}}\left[\frac{\partial}{\partial\lambda^k}\left[\sum_{r_1} V_{ij}^{r_1} + \sum_{r_1\neq r_2} V_i^{r_1}V_j^{r_2}\right]\right] \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{r_1}\left[\frac{\partial}{\partial\lambda^k}V_{ij}^{r_1}\right] + \frac{1}{n^{\frac{3}{2}}}\sum_{r_1\neq r_2} V_i^{r_1}\frac{\partial}{\partial\lambda^k}V_j^{r_2} + \frac{\partial}{\partial\lambda^k}V_i^{r_1}V_j^{r_2}.
\end{aligned}$$

The term $\frac{1}{n^{\frac{3}{2}}}\sum_{r_1}\left[\frac{\partial}{\partial\lambda^k}V_{ij}^{r_1}\right]$ is $o_p(1)$, so the only part that contributes to the limiting distribution is

$$\frac{1}{n^{\frac{3}{2}}}\sum_{r_1\neq r_2} V_i^{r_1}\frac{\partial}{\partial\lambda^k}V_j^{r_2} + \frac{\partial}{\partial\lambda^k}V_i^{r_1}V_j^{r_2}.$$

Re-expressing this as

$$\frac{n-1}{2\sqrt{n}}\,\frac{2}{n(n-1)}\left(\sum_{r_1<r_2} V_i^{r_1}\frac{\partial}{\partial\lambda^k}V_j^{r_2} + V_i^{r_2}\frac{\partial}{\partial\lambda^k}V_j^{r_1} + \sum_{r_1<r_2}\frac{\partial}{\partial\lambda^k}V_i^{r_1}V_j^{r_2} + \frac{\partial}{\partial\lambda^k}V_i^{r_2}V_j^{r_1}\right), \tag{4.25}$$

identifies the U-statistic structure.

The projection of the U-statistic, U_n is defined by Serfling (1980, p.188 (1)) as

$$\hat{U}_n = \sum_{i=1}^{n} E_F\{U_n \mid X_i\} - (n-1)\theta,$$

where θ is the expectation of the U-statistic kernel. For the kernel in (4.25), θ is 0. The projection can also be represented as

$$\hat{U}_n = \frac{m}{n}\sum_{i=1}^{n} \tilde{h}_1(X_i),$$

where m is the order of the kernel. The idea of the projection is to approximate the U-statistic with an *i.i.d.* sum, which is of course highly desirable from the analytic perspective.

Denoting the kernel $V_i^{r_1}\frac{\partial}{\partial\lambda^k}V_j^{r_2} + V_i^{r_2}\frac{\partial}{\partial\lambda^k}V_j^{r_1}$ by $h(V_i^{r_1}, V_j^{r_2})$, the function $\tilde{h}_1$ is by definition (see Serfling (1980, p.177))

$$\begin{aligned}
\tilde{h}_1(V_i^{r1}) &= E[h(V_i^{r_1}, V_j^{r_2}) \mid V_i^{r_1}] \\
&= E[V_i^{r_1}\frac{\partial}{\partial\lambda^k}V_j^{r_2} + V_i^{r_2}\frac{\partial}{\partial\lambda^k}V_j^{r_1} \mid V_i^{r_1}] \\
&= V_i^{r_1}E[\frac{\partial}{\partial\lambda^k}V_j^{r_2}].
\end{aligned}$$

The expectation $E[\frac{\partial}{\partial\lambda^k}V_j^{r_2}] = -E[V_j^{r_2}V_k^{r_2}] = -W_{j,k}$, so that $\tilde{h}_1(V_i^{r1}) = -W_{j,k}V_i^{r_1}$. The same arguments apply to the kernel $\frac{\partial}{\partial\lambda^k}V_i^{r_1}V_j^{r_2} + \frac{\partial}{\partial\lambda^k}V_i^{r_2}V_j^{r_1}$, so that the projection is

$$\frac{n-1}{2\sqrt{n}}\,\frac{2}{n}\left(\sum_{r_1} -W_{j,k}V_i^{r_1} - W_{i,k}V_j^{r_1} + o_p(n^{-1}(\log n)^{\delta})\right), \tag{4.26}$$

where Serfling (1980, p.189, 5.3.3) has been used to evaluate the difference between the projected and original U-statistic. If the kernels are square integrable, which is true in this case by assumption, then δ can be set to 1, and (4.26) becomes

$$\frac{n-1}{\sqrt{n}}\,\frac{1}{n}\left(\sum_{r_1} -W_{j,k}V_i^{r_1} - W_{i,k}V_j^{r_1} + o_p(n^{-1}(\log n))\right)$$

$$\frac{1}{\sqrt{n}}\left(-W_{j,k}V_i - W_{i,k}V_j\right) + o_p(n^{-\frac{3}{2}}\log n)$$
$$\frac{1}{\sqrt{n}}\left(-W_{j,k}V_i - W_{i,k}V_j\right) + o_p(1)$$

as required. ■

Lemma 4.4 states

$$\frac{1}{\sqrt{n}}\left[\nabla_\lambda\left(U_2^c - U_1^c\right) - T^{c*}W^{-1}\mathbf{V}_1\right] = o_p(1).$$

Proof

Writing U_2^c as

$$U_2^c = U_1^c - E[U_1^c\mathbf{V}_2^{*T}]E[\mathbf{V}_2^*\mathbf{V}_2^{*T}]^+\mathbf{V}_2^*, \tag{4.27}$$

and writing nT_{ij}^{c*} for the elements of $E[U_1^c\mathbf{V}_2^{*T}]$, and substituting in the approximation to the inverse of $E[\mathbf{V}_2^*\mathbf{V}_2^{*T}]$ given by Theorem 3.7, then

$$U_2^c = U_1^c - \frac{n}{4n(n-1)}T_{ij}^{c*}(W^{i,u}W^{j,t} + W^{i,t}W^{j,u})V_{ut}^* + O(n^{-2})V_{ut}^*. \tag{4.28}$$

So that on differentiation by λ^k, denoted by $'$, we obtain

$$\begin{aligned} U_2^{c\prime} = U_1^{c\prime} - \left(\frac{n}{4n(n-1)}T_{ij}^{c*}(W^{i,u}W^{j,t} + W^{i,t}W^{j,u})\right)' V_{ut}^* - \\ \frac{n}{4n(n-1)}T_{ij}^{c*}(W^{i,u}W^{j,t} + W^{i,t}W^{j,u})(V_{ut}^*)' + (O(n^{-2})V_{ut}^*)'. \end{aligned} \tag{4.29}$$

Now

$$\tfrac{1}{\sqrt{n}}\left(\tfrac{1}{4(n-1)}T_{ij}^{c*}(W^{i,u}W^{j,t} + W^{i,t}W^{j,u})\right)' V_{ut}^* \xrightarrow{p} 0,$$

and by Lemma 4.2, the order of the coefficients in the term $O(n^{-2})V_{ut}^*$ are unaffected by differentiation so that

$$\tfrac{1}{\sqrt{n}}(O(n^{-2})V_{ut}^*)' \xrightarrow{p} 0.$$

Consequently the only term from (4.29) of stochastic importance is

$$-\frac{1}{4(n-1)}T^{c*}_{ij}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})(V^*_{ut})'.$$

The k^{th} element in $T^{c*}W^{-1}\mathbf{V}_1$ can be written as $T^{c*}_{kj}W^{j,s}V_s$ so that the k^{th} element in the left hand side of 4.27 on division by $\sqrt{n}$ can be re-expressed as

$$\frac{1}{\sqrt{n}}\left[-\frac{1}{4(n-1)}T^{c*}_{ij}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})(V^*_{ut})'-T^{c*}_{kj}W^{j,s}V_s\right]+o_p(1). \qquad (4.30)$$

Using Lemma 4.3, that is replacing $n^{-\frac{3}{2}}V^{*}_{ut}{}'$ with $\frac{1}{\sqrt{n}}(-W_{t,k}V_u-W_{u,k}V_t)+o_p(1)$, (4.30) becomes

$$\frac{1}{\sqrt{n}}\left[\tfrac{1}{4}T^{c*}_{ij}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})(W_{t,k}V_u+W_{u,k}V_t)-T^{c*}_{kj}W^{j,s}V_s\right]+o_p(1).$$

Terms of the form $T^{c*}_{ij}W^{i,u}W^{j,t}W_{t,k}V_u$ on summation become $T^{c*}_{ij}W^{i,u}\delta^j_kV_u=T^{c*}_{ik}W^{i,u}V_u$, but by the symmetry of T^{c*}_{ik} this is equivalent to $T^{c*}_{ki}W^{i,u}V_u$. There are 4 of these terms, so that

$$\tfrac{1}{4}T^{c*}_{ij}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})(W_{t,k}V_u+W_{u,k}V_t)-T^{c*}_{kj}W^{j,s}V_s=0,$$

which provides the result. ∎

The final technical lemma, Lemma 4.19 states that

$$\lim_{n\to\infty}\frac{U^c_{2,ij}}{n}\xrightarrow{p}0. \qquad (4.31)$$

Proof

By definition

$$U^c_2=U^c_1-\frac{n}{4n(n-1)}T^{c*}_{ij}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})V^*_{ut}+O(n^{-2})V^*_{ut},$$

so that

$$U^c_{2,ij}=U^c_{1,ij}-\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}(\frac{1}{4(n-1)}T^{c*}_{ij}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u}))V^*_{ut}-$$

$$\frac{\partial}{\partial\lambda^i}\left(\frac{1}{4(n-1)}T_{ij}^{c*}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})\right)\frac{\partial}{\partial\lambda^j}V_{ut}^{*}-$$
$$\frac{\partial}{\partial\lambda^j}\left(\frac{1}{4(n-1)}T_{ij}^{c*}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})\right)\frac{\partial}{\partial\lambda^i}V_{ut}^{*}-$$
$$\frac{1}{4(n-1)}T_{ij}^{c*}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}V_{ut}^{*}+\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}\left(O(n^{-2})V_{ut}^{*}\right).$$

Using the result (Proposition 3.3) for the order of magnitude of the coefficients of the extended information matrix one can verify, using the weak law of large numbers and the fact that the Bhattacharyya basis elements are mean zero, that on division by n all of the above terms are $o_p(1)$, with the exception of

$$\frac{U_{1,ij}^{c}}{n} \quad\text{and}\quad \frac{1}{4(n-1)n}T_{ij}^{c*}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}V_{ut}^{*}.$$

Consider now the term $\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}\frac{V_{ut}^{*}}{n^2}$ which is stochastically dominated by

$$\frac{1}{n^2}\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}\sum_{r_1}\sum_{r_1\neq r_2}V_u^{r_1}V_t^{r_2}$$
$$=\frac{1}{n^2}\sum_{r_1\neq r_2}\left(\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}V_u^{r_1}\right)V_t^{r_2}+\frac{\partial}{\partial\lambda^i}V_u^{r_1}\frac{\partial}{\partial\lambda^j}V_t^{r_2}+\frac{\partial}{\partial\lambda^j}V_u^{r_1}\frac{\partial}{\partial\lambda^i}V_t^{r_2}+V_u^{r_1}\left(\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}V_u^{r_1}V_t^{r_2}\right).$$

An application of the law of large numbers indicates that this is equal to

$$W_{i,u}W_{j,t}+W_{j,u}W_{i,t}+o_p(1),$$

so that

$$\frac{1}{4(n-1)n}T_{ij}^{c*}(W^{i,u}W^{j,t}+W^{i,t}W^{j,u})\frac{\partial^2}{\partial\lambda^i\partial\lambda^j}V_{ut}^{*}=T_{ij}^{c*}+o_p(1). \tag{4.32}$$

Finally, applying the law of large numbers to $\frac{U_{1,ij}^{c}}{n}$ establishes that it goes in probability to $E[U_{1,ij}^{ck}]$, but

$$E[U_{1,ij}^{ck}]=E[U_1^{ck}V_{ij}^{k}]=E[U_0^{ck}V_{ij}^{k*}]=T_{ij}^{c*}. \tag{4.33}$$

Substitution of (4.32) and (4.33) into (4.31) gives the result. ■