Faculteit der Economische Wetenschappen en Econometrie

Serie Research Memoranda

On the Limit Behavior of a Chi-Square Type Test if the Number of Conditional Moments Tested Approaches Infinity
Preliminary Version

R.M. de Jong
H.J. Bierens

Research Memorandum 1991-35
april 1991

vrije Universiteit *amsterdam*

On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity
Preliminary version

R.M. de Jong and H.J. Bierens
Free University
Department of Econometrics
De Boelelaan 1105
1081 HV Amsterdam, The Netherlands

April 24, 1991

Abstract

In this paper a consistent model specification test is proposed. Some consistent model specification tests have been discussed in econometrics literature. Those tests are consistent by randomization, display a discontinuity in sample size or have an asymptotic distribution that depends on the datagenerating proces and on the model, whereas our test does not have one of those disadvantages. Our test can be viewed upon as a conditional moment test as proposed by Newey but instead of a fixed number of conditional moments an asymptotically infinite number of moment conditions is employed. The use of an asymptotically infinite number of conditional moments will make it possible to obtain a consistent test. Computation of the test statistic is particularly simple since in finite samples our statistic is equivalent to a chi-square conditional moment test of a finite number of conditional moments.

1 Introduction

Most model specification tests can be put in the framework of Newey's(1985) conditional moment (CM) tests of functional form. Those CM tests of functional form exploit the property that for correctly specified models the conditional expectation of certain functions of the observations should be almost surely equal to zero. A chi-square test can be based on a weighted average of the sample equivalents of these moments. As Bierens(1982,1990) notes, this type of test cannot be consistent against all possible alternatives and the power of the CM tests heavily depends upon the set of weighting functions chosen. It will always be possible to construct a datagenerating process such that the moment conditions hold while the null is false.

Bierens(1982,1984,1987,1988,1990) suggests to remedy this inconsistency by using an infinite number of moment restrictions. We remark that it is also to be expected that as more observations become available we would wish to test a model against a higher-dimensional alternative in order to obtain power against more alternatives. If this is the case, the asymptotic theory for chi-square tests of a fixed number of conditional moments might not necessarily be correct since the number of conditional moment conditions is allowed to grow with n. In this paper we will investigate the behavior of what is known as a type of the chi-square misspecification test if the number of conditional moments tested is allowed to slowly approach infinity as the number of observations increases. The asymptotic distribution of the sequence of statistics obtained this way will be shown to be standardnormal and the statistic will be shown to be consistent under regularity conditions. The research is motivated by the fact that the consistent tests obtained by Bierens(1982,1984,1987,1988,1990) display an undesirable discontinuity in sample size, are consistent by virtue of randomization or have an asymptotic distribution that depends on the model and on the underlying distribution of the datagenerating process. This paper therefore can be considered a natural extension of the results obtained in those papers. Related work has been done by Andrews(1991) who proves asymptotic normality of linear functionals of series estimators. However, our statistic cannot be written as a linear functional of some series estimator.

In section 2 of this paper the theory of consistency of econometric tests will be discussed. In section 3 the asymptotic theory of our result will be explained. In section 4 the consistency of our test in a special case will be proven. Appendix A contains the set of assumptions that are usually maintained in nonlinear regression analysis. It will be referred to as "Assumption A". Appendix B will be devoted to a set of assumptions needed in addition to the standard assumptions of Assumption A. It will be referred to as "Assumption B". A smaller set of assumptions that imply Assumption B is also present in this appendix; it will be referred to as "Assumption B'". Appendix C contains the proofs of the lemmas and theorems.

2 Consistency of tests

In parametric regression analysis it assumed that the regression function $r(x)$ belongs to a family of known real functions $f(x,\theta)$ on $R^p \times \Theta$, where Θ is a subset of R^m. In

Lemma 1 *Let $\epsilon \in R$ and $x \in R^p$ be random variables such that $E\epsilon^2 < \infty$ and $x \in [0, 2\pi]^p$. Suppose x has a continuous density $f(x)$ and cdf $F(x)$ on $[0, 2\pi]^p$. Let Z^p denote the collection of integers in R^p. If $P[E(\epsilon|x) = 0] < 1$ then $\exists t \in Z^p$: $E\epsilon \exp(it'x) \neq 0$.*
Proof: See appendix.

Bierens (1990) proofs that if the conditions of our lemma 1 hold, the set

$$\{t \in R^p : E\epsilon \exp(t'x) = 0\}$$

has Lebesgue measure zero. In that paper, a consistent test is based on a series of functions

$$g_l(x) = \exp(t_l'\Phi(x))$$

for some 1-1 mapping $\Phi(.)$ that maps R^p to a bounded subset C of R^p and a series t_l that grows towards a dense subset of C. The continuity of the density of the explanatory variables is not needed there. However, in this paper it is preferred to use an enumeration of the Fourier series since we expect that this will improve power properties because smooth functions are usually well approximated by a lower-order Fourier expansion.

3 Asymptotic theory of chi-square tests of a growing number of conditional moments

As is well-known a suitably rescaled chi-square distributed random variable converges to a normal distribution as the number of degrees of freedom approaches infinity, i.e.

$$\frac{1}{\sqrt{2k}}(\chi_k^2 - k) \to N(0,1)$$

in distribution. The intuition behind our test is that this might happen also if the number of conditional moments tested in a chi-square misspecification test is allowed to slowly approach infinity as sample size tends to infinity. However, for finite sample size our test will be equivalent to what is known as a chi-square test of a finite number of conditional moments. This property is extremely convenient since this will allow the use of all standard regression software packages in order to calculate the statistic proposed. The ease of computation of this type of consistent test has been an important motivation in the developing of the theory presented here.

In deriving the asymptotic distribution of the test we will not use an enumeration of the Fourier series but instead we will use some abstract series of functions $g_l(x)$. The conditional moments used in the derivation of the limiting distribution of our test will be of the form $E\epsilon_1 g_l(x_1)$, for $l = 1, ..k$, $k = k(n) \to \infty$ as $n \to \infty$. The consistency of the test for a particular series $g_l(x)$, $l = 1, ..$ can then be proven on a case by case basis as is done for the Fourier series in our theorem 3. We note that under the alternative the speed of convergence towards infinity of our test is slower than sample size. This speed of convergence is obtained in Bierens(1990). The

Suppose $k = k(n)$ for some positive integer-valued sequence $k(n)$ such that $k(n) \to \infty$ as $n \to \infty$. The series $k(n)$ will be abbreviated by k in the sequel. Let V_k be the covariance matrix of $\tilde{m}_k$, i.e. for $h, l = 1, ..k$,

$$[V_k]_{hl} = E\epsilon_1^2(g_l(x_1) - [(\partial/\partial\theta)f(x_1,\theta_0)]A^{-1}c_l)(g_h(x_1) - [(\partial/\partial\theta)f(x_1,\theta_0)]A^{-1}c_h)$$

Put $\lambda(k) = \lambda_{min}(V_k)$ and $\hat{\lambda}_k = \lambda_{min}(\hat{V}_k)$. Let $\hat{V}_k$ denote the sample equivalent of V_k, i.e.

$$[\hat{V}_k]_{hl} = (1/n)\sum_{j=1}^{n} e_j^2(g_l(x_j) - [(\partial/\partial\theta)f(x_j,\hat{\theta})]\hat{A}^{-1}\hat{c}_l)(g_h(x_j) - [(\partial/\partial\theta)f(x_j,\hat{\theta})]\hat{A}^{-1}\hat{c}_h)$$

where $e_j = y_j - f(x_j, \hat{\theta})$.

In order to obtain some speeds of convergence in probability used in the derivation of our result it will be necessary to strengthen the standard assumptions for proving consistency and asymptotic normality of nonlinear least squares estimators as listed in Appendix A. The strengthening involves the assumptions listed in Appendix B. In Appendix B ten necessary assumptions are given that are used in the derivation of the result. Those assumptions can be shown to follow from a considerably smaller set of assumptions that is also listed in Appendix B and will be referred to as Assumption B'. The extra assumptions made do not seem to be very restrictive. They involve the existence of moments of functions of the fourth order moment of the regresion residuals and first and second derivatives of the parametric response function $f(x, \theta)$.

Of course, some limitations on the sequences k_n, λ_k and s_k are necessary too in order to obtain the asymptotic normality result. The restrictions on the λ_k sequence can be interpreted as a restriction on the amount of multicollinearity that is asymptotically allowed. Clearly we will have to require that from some point all covariance matrices are nonsingular. This type of condition is also used in Andrews(1991). His results show that if the density of x_1 is bounded away from zero on $[0, 2\pi]$, if no parameters are estimated, if an enumeration of the Fourier series is used for the $g_l(.)$ functions and it is assumed that $E(\epsilon_1^2|x_1) = E\epsilon_1^2 = \sigma^2$ the minimal eigenvalue of the V_k matrices will be bounded away from zero. Our results allow the eigenvalue sequence to slowly converge to zero. In spite of results of this sort it is clear that the restrictions on the eigenvalue sequence are restrictive and in some cases they are likely to be violated. They however appear to be necessary in this type of analysis. See Gallant(1981) and Andrews(1991) for a discussion of this type of assumption.
The assumptions on these sequences that are necessary in this paper are:

Assumption C :

- 1: $\lim_{n\to\infty} k/n = 0$ and $\lim_{n\to\infty} 1/k = 0$
- 2: $\limsup_{n\to\infty} k^{-1/2}s_k < \infty$
- 3: $\lim_{n\to\infty} \lambda_k^{-2}s_k^4k^{3/2}n^{-1/2} = 0$

then U_n is asymptotically $N(0, n^2 EH_n(z_1, z_2)^2/2)$ distributed.

The result of theorem 1 has little practical significance. In practice, of course, the matrix V_k will usually be unknown and will have to be estimated. Also it will be necessary to replace the $\tilde{m}_k$ vector by the $\hat{m}_k$ vector. For this purpose we need the following lemmas of which in particular the first is hard to prove. By means of the next lemma we will prove that under regularity conditions the statistics based on the exact covariance matrix and on the estimated covariance matrix will be asymptotically equivalent.

Lemma 4 *Suppose that Assumptions A, B (or B') and C hold. Then*

$$\text{plim}_{n\to\infty} k^{-1/2} \tilde{m}_k'(V_k^{-1} - \hat{V}_k^{-1})\tilde{m}_k = 0.$$

Proof:
The lemma follows from Assumption C and the following four lemmas:

Lemma 4a:

$$k^{-1/2}(\tilde{m}_k' V^{-1}\tilde{m}_k - \tilde{m}_k'\hat{V}_k\tilde{m}_k) \leq \tilde{m}_k'\tilde{m}_k k^{-1/2}\hat{\lambda}_k\lambda_k[\sum_{h=1}^{k}\sum_{l=1}^{k}(V_k - \hat{V}_k)^2_{(hl)}]^{1/2}$$

almost surely.

Lemma 4b:

$$|\sum_{h=1}^{k}\sum_{l=1}^{k}(V_k - \hat{V}_k)^2_{(hl)}| = O_P(n^{-1}k^2 s_k^4)$$

Lemma 4c:

$$\tilde{m}_k'\tilde{m}_k k^{-1} s_k^{-2} = O_P(1)$$

Lemma 4d:

$$\text{plim}_{n\to\infty}\hat{\lambda}_k/\lambda_k = 1$$

The next lemma will enable us to asymptotically replace the $\tilde{m}_k$-vector by the $\hat{m}_k$-vector which is based on the regression residuals.

Lemma 5 *If Assumptions A, B (or B') and C hold then*

$$\text{plim}_{n\to\infty} k^{-1/2}(\tilde{m}_k'\hat{V}_k^{-1}\tilde{m}_k - \hat{m}_k'\hat{V}_k^{-1}\hat{m}_k) = 0.$$

Proof: See appendix.

After showing that the effect of the estimation of the parameters and the effect of the estimation of the covariance matrix are negligable the following theorem now easily follows:

Theorem 2 *Under assumptions A, B (or B') and C,*

- 1: $(y_1, x_1), ..., (y_n, x_n)$ is a sample from a probability distribution $F(y,x)$ on $R \times R^p$ for which $Ey_1^2 < \infty$.
- 2: The parameter space Θ is a compact and convex subset of R^m and $f(x,\theta)$ is for each $\theta \in \Theta$ a Borel measurable real function on R^p and for each p-vector x a twice continuously differentiable real function on Θ. Moreover, for $a, b = 1, ..m$,
  - 2.1: $E\sup_{\theta\in\Theta} f(x_1,\theta)^2 < \infty$
  - 2.2: $E\sup_{\theta\in\Theta} |(y_1 - f(x_1,\theta))^2((\partial/\partial\theta_a)f(x_1,\theta))((\partial/\partial\theta_b)f(x_1,\theta))| < \infty$
  - 2.3: $E\sup_{\theta\in\Theta} |((\partial/\partial\theta_a)f(x_1,\theta))((\partial/\partial\theta_b)f(x_1,\theta))| < \infty$
  - 2.4: $E\sup_{\theta\in\Theta} |(y_1 - f(x_1,\theta))(\partial/\partial\theta_a)(\partial/\partial\theta_b)f(x_1,\theta)| < \infty$
- 3: $E(y_1 - f(x_1,\theta))^2$ takes a unique minimum on Θ at θ_0. Under H_0 the parameter vector θ_0 is an interior point of Θ.
- 4: The matrix

  $$A = E((\partial/\partial\theta')f(x_1,\theta_0))((\partial/\partial\theta)f(x_1,\theta_0))$$

  is non-singular.

Appendix B

The following assumptions are needed in addition to the standard assumptions of Appendix A in order to prove our results: for $a, b, c = 1, ..m$:

Assumption B :

- 1: $E\sup_{\theta\in\Theta} |[(\partial/\partial\theta_a)(\partial/\partial\theta_b)f(x_1,\theta)][(\partial/\partial\theta_c)f(x_1,\theta)]| < \infty$
- 2: $E|[(\partial/\partial\theta_a)f(x_1,\theta_0)][(\partial/\partial\theta_b)f(x_1,\theta_0)]|^2 < \infty$
- 3: $E\sup_{\theta\in\Theta} |(\partial/\partial\theta_a)(\partial/\partial\theta_b)f(x_1,\theta)| < \infty$
- 4: $E\sup_{\theta\in\Theta} |(y_1 - f(x_1,\theta))^2(\partial/\partial\theta_a)f(x_1,\theta)| < \infty$
- 5: $E\sup_{\theta\in\Theta} |(y_1 - f(x_1,\theta))^2(\partial/\partial\theta_a)(\partial/\partial\theta_b)f(x_1,\theta)| < \infty$
- 6: $E\sup_{\theta\in\Theta} |(y_1 - f(x_1,\theta))^2[(\partial/\partial\theta_a)f(x_1,\theta)][(\partial/\partial\theta_b)(\partial/\partial\theta_c)f(x_1,\theta_0)]| < \infty$
- 7: $E\epsilon_1^4 < \infty$; $E\epsilon_1^4[(\partial/\partial\theta_a)f(x_1,\theta)]^2 < \infty$
- 8: $E\epsilon_1^4[(\partial/\partial\theta_a)f(x_1,\theta_0)]^2[(\partial/\partial\theta_b)f(x_1,\theta_0)]^2 < \infty$
- 9: $E\sup_{\theta\in\Theta} |(y_1 - f(x_1,\theta))[(\partial/\partial\theta_a)f(x_1,\theta)][(\partial/\partial\theta_b)f(x_1,\theta)][(\partial/\partial\theta_c)f(x_1,\theta)]| < \infty$
- 10: $E\sup_{\theta\in\Theta} |(y_1 - f(x_1,\theta))[(\partial/\partial\theta_a)f(x_1,\theta)][(\partial/\partial\theta_b)f(x_1,\theta)]| < \infty$

by Assumption C and some calculation (the inequality $(a'b)^2 \leq a'ab'b$ is used).

Proof of lemma 4 :

In the proof of lemma 4 the k subscripts will be dropped in order to improve readability.

Proof of lemma 4a :

Note that

$$k^{-1/2}\tilde{m}'(V^{-1}-\hat{V}^{-1})\tilde{m} = k^{-1/2}\tilde{m}'V^{-1}(\hat{V}-V)\hat{V}^{-1}\tilde{m}$$

$$\leq k^{-1/2}\tilde{m}'\tilde{m}\lambda_{max}(V^{-1}(\hat{V}-V)\hat{V}^{-1})$$

$$\leq \tilde{m}'\tilde{m}k^{-1/2}(\sup_{c'c=1}|c'V^{-1}(V-\hat{V})\hat{V}^{-1}c|^2)^{1/2}$$

$$\leq \tilde{m}'\tilde{m}k^{-1/2}\lambda_{max}(V^{-1})\lambda_{max}(\hat{V}^{-1})(\lambda_{max}(V-\hat{V})^2)^{1/2}$$

$$\leq \tilde{m}'\tilde{m}k^{-1/2}\sup_{c'c=1}|c'V^{-2}cc'\hat{V}^{-1}(V-\hat{V})^2\hat{V}^{-1}c|$$

$$\leq \tilde{m}'\tilde{m}k^{-1/2}\lambda_k^{-1}\hat{\lambda}_k^{-1}\max(|\lambda_{max}(V-\hat{V})|,|\lambda_{min}(V-\hat{V})|)$$

$$\leq \tilde{m}'\tilde{m}k^{-1/2}\hat{\lambda}_k^{-1}\lambda_k^{-1}\sup_{c'c=1}\sum_{l=1}^{k}\sum_{h=1}^{k}|c_l||c_h||V-\hat{V}|_{(hl)}$$

$$\leq \tilde{m}'\tilde{m}k^{-1/2}\hat{\lambda}_k^{-1}\lambda_k^{-1}(\sum_{h=1}^{k}\sum_{l=1}^{k}(V-\hat{V})^2_{(hl)})^{1/2}$$

by two applications of the Cauchy-Schwartz inequality and we note that the same argument is valid for the $-k^{-1/2}\tilde{m}'(V^{-1}-\hat{V}^{-1})\tilde{m}$ term.

Proof of lemma 4b :

Here we will prove that $(\sum_{h=1}^{k}\sum_{l=1}^{k}(V-\hat{V})^2_{(hl)}) = O_P(n^{-1}k^2s_k^4)$. This will be done in four steps:

- 1: $\sum_{h=1}^{k}\sum_{l=1}^{k}(\hat{V}(\hat{c}_h,\hat{c}_l,\hat{A},\hat{\theta})_{(hl)} - \hat{V}(c_h,c_l,\hat{A},\hat{\theta})_{(hl)})^2 = O_P(n^{-1}k^2s_k^4)$.
- 2: $\sum_{h=1}^{k}\sum_{l=1}^{k}(\hat{V}(c_h,c_l,\hat{A},\hat{\theta})_{(hl)} - \hat{V}(c_h,c_l,A,\hat{\theta})_{(hl)})^2 = O_P(n^{-1}k^2s_k^4)$.
- 3: $\sum_{h=1}^{k}\sum_{l=1}^{k}(\hat{V}(c_h,c_l,A,\hat{\theta})_{(hl)} - \hat{V}(c_h,c_l,A,\theta_0)_{(hl)})^2 = O_P(n^{-1}k^2s_k^4)$.
- 4: $\sum_{h=1}^{k}\sum_{l=1}^{k}(V_{(hl)} - \hat{V}(c_h,c_l,A,\theta_0)_{(hl)})^2 = O_P(n^{-1}k^2s_k^4)$.

where $\hat{V}(.,.,.,.)$ is the sample variance-covariance matrix based on the arguments. Our first effort will be the substitution of $\hat{c}_h$ by c_h in the first formula, i.e. the proof of the fact that

$$\sum_{h=1}^{k}\sum_{l=1}^{k}(\hat{V}(\hat{c}_h,\hat{c}_l,\hat{A},\hat{\theta})_{(hl)} - \hat{V}(c_h,\hat{c}_l,\hat{A},\hat{\theta})_{(hl)})^2 = O_P(n^{-1}k^2s_k^4).$$

and $\hat{A} - A = O_P(n^{-1/2})$ since, for $a, b = 1, ..m$,

$$\hat{A}_{(ab)} - (1/n)\sum_{j=1}^{n}[(\partial/\partial\theta_a)f(x_j,\theta_0)][(\partial/\partial\theta_b)f(x_j,\theta_0)]$$

$$= (1/n)\sum_{j=1}^{n}[(\partial/\partial\theta)(\partial/\partial\theta_a)f(x_j,\tilde{\theta}^{(a,b)})][(\partial/\partial\theta_b)f(x_j,\tilde{\theta}^{(a,b)})](\hat{\theta}-\theta_0)$$

$$+(1/n)\sum_{j=1}^{n}[(\partial/\partial\theta)(\partial/\partial\theta_b)f(x_j,\tilde{\theta}^{(a,b)})][(\partial/\partial\theta_a)f(x_j,\tilde{\theta}^{(a,b)})](\hat{\theta}-\theta_0)$$

and the last two terms are $O_P(n^{-1/2})$ by Assumption C, the ULLN of Jennrich(1969) and consistency of the mean values. The result then follows because, for $a, b = 1, ..m$,

$$E[(1/n)\sum_{j=1}^{n}[(\partial/\partial\theta_a)f(x_j,\theta_0)][(\partial/\partial\theta_b)f(x_j,\theta_0)] - E[(\partial/\partial\theta_a)f(x_j,\theta_0)][(\partial/\partial\theta_b)f(x_j,\theta_0)]]^2$$

$$= O(n^{-1/2})$$

by Assumption C. With this result it is easy to prove that

$$\sum_{h=1}^{k}\sum_{l=1}^{k}(\hat{V}(c_h,c_l,A,\hat{\theta})_{(hl)} - (\hat{V}(c_h,c_l,\hat{A},\hat{\theta})_{(hl)})^2 = O_P(n^{-1}k^2s_k^4).$$

using Assumptions A and C.

The third part is then proven by noting that, by Taylor's theorem,

$$\sup_{h,l=1,..k}|(1/n)\sum_{j=1}^{n}e_j^2(g_l(x_j) - [(\partial/\partial\theta)f(x_j,\hat{\theta})]A^{-1}c_l)(g_h(x_j) - [(\partial/\partial\theta)f(x_j,\hat{\theta})]A^{-1}c_h)$$

$$-(1/n)\sum_{j=1}^{n}\epsilon_j^2(g_l(x_j) - [(\partial/\partial\theta)f(x_j,\theta_0)]A^{-1}c_l)(g_h(x_j) - [(\partial/\partial\theta)f(x_j,\theta_0)]A^{-1}c_h)|$$

$$\leq \sup_{h,l=1,..k} \| [(1/n)\sum_{j=1}^{n}2e_j(\theta)[(\partial/\partial\theta')f(x_j,\theta)](g_l(x_j) - [(\partial/\partial\theta')f(x_j,\theta)]A^{-1}c_l)\times$$

$$(g_h(x_j) - [(\partial/\partial\theta')f(x_j,\theta)]A^{-1}c_h)$$

$$-e_j^2(\theta)[(\partial/\partial\theta')(\partial/\partial\theta)f(x_j,\theta)]A^{-1}(c_lg_h(x_j) + c_hg_l(x_j))$$

$$+e_j^2(\theta)[(\partial/\partial\theta)f(x_j,\theta)]A^{-1}c_l[(\partial/\partial\theta')(\partial/\partial\theta)f(x_j,\theta)]A^{-1}c_h$$

$$+e_j^2(\theta)[(\partial/\partial\theta)f(x_j,\theta)]A^{-1}c_h[(\partial/\partial\theta')(\partial/\partial\theta)f(x_j,\theta)]A^{-1}c_l]_{(\theta^{(a,b)})} \|\|\| (\hat{\theta}-\theta_0) \|$$

$$|\sqrt{n}(\hat{\theta}-\theta_0) - A^{-1}n^{-1/2}\sum_{j=1}^{n}\epsilon_j(\partial/\partial\theta')f(x_j,\theta_0)| = O_P(n^{-1/2})$$

Next we establish the inequality

$$(\tilde{m}_k-\hat{m}_k)'(\tilde{m}_k-\hat{m}_k) = \sum_{h=1}^{k}(\tilde{m}_{kh}-\hat{m}_{kh})^2$$

$$= \sum_{h=1}^{k}[n^{-1/2}\sum_{j=1}^{n}\epsilon_j[(\partial/\partial\theta)f(x_j,\theta)]A^{-1}c_h - (\hat{\theta}-\theta_0)(1/n)\sum_{j=1}^{n}[(\partial/\partial\theta)f(x_j,\tilde{\theta}^{(a)})]g_h(x_j)]^2$$

$$= O_P(n^{-1}s_k^2k) + \sum_{h=1}^{k}[n^{1/2}(\hat{\theta}-\theta_0)'(\tilde{c}_h-c_h)]^2$$

$$\leq \| n^{1/2}(\hat{\theta}-\theta_0) \|^2 \sum_{h=1}^{k}(\tilde{c}_h-c_h)'(\tilde{c}_h-c_h) + O_P(n^{-1}s_k^2k)$$

$$= O_P(kn^{-1}s_k^2)$$

(by the argument in lemma 3) where

$$\tilde{c}_h = (1/n)\sum_{j=1}^{n}[(\partial/\partial\theta)f(x_j,\tilde{\theta}^{(h)})]g_h(x_j)$$

Lemma 4 follows by noting that

$$(k^{-1/2}\tilde{m}_k'\hat{V}_k^{-1}(\tilde{m}_k-\hat{m}_k))^2 = O_P(\lambda_k^{-2}k^{3/2}s^4n^{-1}) = o_P(1)$$

by Assumption B. A similar argument is valid for the term not yet handled.

Proof of theorem 3 :
Since

$$k^{3/2}n^{-1}T = kn^{-1}(\hat{m}_k'\hat{V}_k\hat{m}_k - k)$$

and since the sequence k^2n^{-1} converges to zero we only take into account the expression

$$kn^{-1}\hat{m}_k'\hat{V}_k\hat{m}_k$$

Note that $\lambda_{max}(V_k) < Cks_k^2$ and $\lambda_{max}(\hat{V}_k)/\lambda_{max}(V_k) \to 1$ in probability by an argument similar to that in the proof of lemma 4. Consistency is then seen from

$$kn^{-1}\hat{m}_k'\hat{V}_k\hat{m}_k$$

$$\geq kn^{-1}(\hat{m}_k'\hat{m}_k/\lambda_{max}(\hat{V}_k))$$

$$\geq (n^{-1}\hat{m}_k'\hat{m}_k)(1/C)(\lambda_{max}(V_k)/\lambda_{max}(\hat{V}_k))$$

since by lemma 1 $n^{-1/2}\hat{m}_k$ will eventually contain at least one element that converges in probability to a nonzero constant.