CONSISTENT ESTIMATION WITH A LARGE NUMBER OF WEAK INSTRUMENTS*

John C. Chao
University of Maryland

Norman R. Swanson
Rutgers University

First Draft: June 2001
This Revision: September 2002

Abstract

This paper conducts a general analysis of the conditions under which consistent estimation can be achieved in instrumental variables regression when the available instruments are weak in the local-to-zero sense. More precisely, the approach adopted in this paper combines key features of the local-to-zero framework of Staiger and Stock (1997) and the many-instrument framework of Morimune (1983) and Bekker (1994) and generalizes both of these frameworks in the following ways. First, we consider a general local-to-zero framework which allows for an arbitrary degree of instrument weakness by modeling the first-stage coefficients as shrinking toward zero at an unspecified rate, say b_n^{-1}. Our local-to-zero setup, in fact, reduces to that of Staiger and Stock (1997) in the case where $b_n = \sqrt{n}$. In addition, we examine a broad class of single-equation estimators which extends the well-known k-class to include, amongst others, the Jackknife Instrumental Variables Estimator ($JIVE$) of Angrist, Imbens, and Krueger (1999). Analysis of estimators within this extended class based on a pathwise asymptotic scheme, where the number of instruments K_n is allowed to grow as a function of the sample size, reveals that consistent estimation depends importantly on the relative magnitudes of r_n, the growth rate of the concentration parameter, and K_n. In particular, it is shown that members of the extended class which satisfy certain general condtions, such as $LIML$ and $JIVE$, are consistent provided that $\frac{\sqrt{K_n}}{r_n} \to 0$, as $n \to \infty$. On the other hand, the two-stage least squares ($2SLS$) estimator is shown not to satisfy the needed conditions and is found to be consistent only if $\frac{K_n}{r_n} \to 0$, as $n \to \infty$. A main point of our paper is that the use of many instruments may be beneficial from a point estimation standpoint in empirical applications where the available instruments are weak but abundant, as it provides an extra source, by which the concentration parameter can grow, thus, allowing consistent estimation to be achievable, in certain cases, even in the presence of weak instruments. Our results, thus, add to the findings of Staiger and Stock (1997) who study a local-to-zero framework where K_n is held fixed and the concentration parameter does not diverge as sample size grows; in consequence, no single-equation estimator is found to be consistent under their setup.

JEL classification: C12, C22.
Keywords: instrumental variables, *k*-class estimator, local-to-zero framework, pathwise asymptotics, weak instruments.

*John C. Chao: Department of Economics, University of Maryland, College Park, MD, USA 20742, chao@econ.umd.edu. Norman R. Swanson: Department of Economics, Rutgers University, New Brunswick, NJ, USA 80901, nswanson@econ.rutgers.edu. The authors would like to thank Harry Kelejian, Ingmar Prucha, Graham Elliott, Tao Zha, Eric Zivot, and participants at the 2001 summer econometric society meeting for useful comments and suggestions. In addition, parts of this paper were written while the second author was visiting the University of California, San Diego, and he would like to thank members of the economics department for providing a stimulating research environment within which to work.

1 Introduction

In a well-known recent paper, Staiger and Stock (1997) propose an alternative asymptotic framework for analyzing instrumental variables regression when the available instruments are only weakly correlated with the endogenous explanatory variables. More precisely, Staiger and Stock (1997) model the effects of having weak instruments using a clever device which takes the coefficients of the instruments in the first-stage regression to be in a $T^{-\frac{1}{2}}$ shrinking neighborhood of zero, with T denoting the sample size in their paper. They show that, when such a "local-to-zero" device is employed, the usual single-equation estimators, such as the two-stage least squares ($2SLS$) and the limited information maximum likelihood ($LIML$) estimators, are no longer consistent and instead converge to nonstandard distributions in the limit[1]. An important feature of their framework, as have been noted by Staiger and Stock, is that, in contrast with conventional asymptotic analysis, the concentration parameter under their weak-instrument setup does not diverge but rather, roughly speaking, stays constant in expectation as the sample size grows. This, in turn, explains the inconsistency results they obtained.

This paper conducts a general analysis of the conditions under which consistent estimation can be achieved in instrumental variables regressions even when the available instruments are weak in the local-to-zero sense. In particular, one key difference between our paper and Staiger and Stock (1997) is that, whereas the latter keeps the number of instruments fixed in performing the limiting operation, we in this paper investigate, within a general local-to-zero framework, the case where the number of instruments (or the degree of apparent overidentification) is allowed to approach infinity as a function of the sample size. A main point of our paper is that the use of many instruments, as approximated by taking the number of instruments to infinity as a function of the sample size, often provides an extra source by which the concentration parameter can grow, so that consistent estimation may become achievable even in the presence of weak instruments in this case.

Asymptotic analyses based on taking the number of instruments to infinity have also been undertaken by Morimune (1983), Bekker (1994), and Hahn (1997), Donald and Newey (2001), Hahn, Hausman, and Kuersteiner (2001), amongst others. Hahn and Inoue (2000) have, in fact, referred to this approach as the "many-instrument" asymptotic approach and have presented Monte Carlo evidence showing that this approach often provides very good approximations for the finite sample behavior of the usual single-equation estimators, even when the number of instruments is only moderate. The "many-instrument" papers cited above, however, do not explicitly analyzed the case where the instruments are weak in the local-to-zero sense. Indeed, the relationship between the local-to-zero framework and the many-instrument framework is as yet not fully understood.

An additional objective of this paper is, thus, to provide results which shed light on the connection between these two important frameworks. To this end, we adopt here a very general setup which combines key features of the local-to-zero and the many-instrument asymptotic frameworks, and which generalizes both of these frameworks in a number of ways. First, letting n denote the sample size in our paper, we consider a local-to-zero setup which generalizes that of Staiger and Stock (1997) in the sense that we take the rate of shrinkage toward zero of the coefficients of the instrumental variables in the first-stage equation to be $1/b_n$ for some arbitrary nondecreasing sequence $\{b_n\}$ instead of using the specific rate $1/\sqrt{n}$. Second, we model the rate of information accumulation

[1]Related to this work on the local-to-zero modeling of weak instruments is the research by Phillips (1989) and Choi and Phillips (1992), which addresses the implications for statistical inference when the underlying simultaneous equations model is only partially identified. Indeed, to the best of our knowledge, Phillips (1989) is the first paper which systematically investigates both the finite sample and the asymptotic properties of conventional econometric procedures when the usual rank condition for identification is not formally satisfied.

in the instruments to be some nondecreasing sequence $\{m_{1n}\}$, thus, allowing it to be more general than the usual assumption that $m_{1n} = n$. Third, we consider a very broad class of single-equation estimators, which we call the ω-class. This class of estimators extends the well-known k-class by allowing the value of k to vary across observations. An important reason for this latter generalization is that this larger class of estimators includes the Jackknife Instrumental Variables Estimator ($JIVE$) of Angrist, Imbens, and Krueger (1999), whereas $JIVE$ is not a member of the k-class. Finally, unlike a number of the other papers which take a many-instrument asymptotic approach, we do not make a Gaussian error assumption in this paper. In this sense, the framework adopted here can also be viewed as extending that of Morimune (1983) and Bekker (1994) to the more general case where the disturbances may be non-Gaussian[2].

On the other hand, it should be noted that the scope of our paper is more limited vis-à-vis several of the papers cited above, as we do not consider asymptotic properties of test statistics and interval estimation procedures; nor do we derive the asymptotic distributions of the ω-class estimators. Rather, we focus our attention on establishing the consistency of ω-class estimators under a pathwise asymptotic scheme where both n and the number of instruments, K_n, are allowed to approach infinity, but with K_n going to infinity at a rate no faster than n.

Our results indicate that consistent estimation depends less on individual assumptions about the local-to-zero structure or the rate of information accumulation (as given by specific choices of the sequences b_n and m_{1n}), than it does on the rate of growth of the concentration parameter, which we denote as r_n, and which is some function of b_n and m_{1n}. In particular, consistent estimation is found to depend crucially on the relative magnitudes of r_n and K_n. More specifically, our results show that, even within a local-to-zero framework, consistent estimation is achievable for members of the ω-class which satisfy certain general conditions, provided that $\frac{\sqrt{K_n}}{r_n} \to 0$, as $n \to \infty$. Specializing our results to specific estimators, we show that $LIML$ and $JIVE$ both satisfy our conditions, whereas the $2SLS$ estimator does not. Indeed, it turns out that the $2SLS$ estimator is consistent only if $\frac{K_n}{r_n} \to 0$, as $n \to \infty$[3]. Our results, thus, make precise the sense in which the $2SLS$ estimator is less robust to instrument

[2]In recent years, there have been other papers which extend the many-instrument asymptotic framework to the case with non-Gaussian errors. In particular, Bekker and van der Ploeg (1999) examine the case where the regression errors may be non-Gaussian and even heteroskedastic but where the instruments are restricted to be dummy variables, whereas van Hasselt (2000) studies the IV (or $2SLS$) estimator in the context of a model with non-Gaussian and homoskedastic errors. Our paper can best be viewed as complementing these other papers, as we seek to extend the literature in different directions than that taken in these papers. In particular, note that neither Bekker and van der Ploeg (1999) nor van Hasselt (2000) considers the case of weak instruments in a local-to-zero framework, as we do in this paper. Moreover, we consider general stochastic instruments, which may be either discrete or continuous random variables. Finally, as discussed above, we study a very broad class of estimators which includes, in addition to the $2SLS$ and $LIML$ estimators, the Jackknife IV estimator ($JIVE$) amongst others. On the other hand, neither Bekker and van der Ploeg (1999) nor van Hasselt (2000) considers $JIVE$ in their analyses.

[3]An interesting recent paper by Stock and Yogo (2001) also examines the case where the number of instruments is allowed to approach infinity in a local-to-zero framework and, thus, deserves special note. In particular, it should be pointed out that there are a number of important differences between our paper and theirs. First, their paper is primarily concerned with the development of test procedures for assessing whether instruments are weak. Hence, they do not attempt to characterize general conditions under which consistent estimation may be achieved in the presence of weak instruments, as is done in the current paper. Using our notations, Stock and Yogo (2001) study the case where the concentration parameter diverges at rate K_n, as $K_n \to \infty$ (i.e., the case where $r_n = K_n$). Thus, their analysis does not permit the same degree of generality as this paper in modeling the extent to which instruments may be weak. They do, however, provide interesting results on the asymptotic expansion of the distributions of the $2SLS$ and the $LIML$ estimators under Gaussian errors for the $r_n = K_n$ case. Secondly, Stock and Yogo (2001) do not attempt to show, as we do, that consistent estimation may be possible, with respect to certain estimators, even when instruments are so weak that the rate at which

weakness vis-à-vis $LIML$, $JIVE$, and other ω-class estimators which satisfy our general conditions. Moreover, our analysis gives, to the best of our knowledge, the first formal proof of the consistency of $JIVE$ under a local-to-zero, many-instrument setup[4].

The rest of the paper is organized as follows. Section 2 sets up our model and discusses the assumptions used. Section 3 presents the main results of the paper and comments on the implications of these results. Concluding remarks are given in Section 4, and all proofs are gathered in an appendix. The following notation is used in the remainder of the paper: $Tr(\cdot)$ denotes the trace of a matrix, A^+ denotes the Moore-Penrose inverse of a (possibly singular) matrix, " > 0" denotes positive definiteness when applied to matrices, $\underline{\lim}_{n\to\infty} a_n$ denotes the limit inferior of the sequence $\{a_n\}$, and $\overline{\lim}_{n\to\infty} a_n$ denotes the limit superior of the sequence $\{a_n\}$. In addition, $P_X = X(X'X)^{-1}X'$ denotes the matrix which projects orthogonally onto the range space of X and $Q_X = I - P_X$.

2 Model and Assumptions

Consider the simultaneous equations model (SEM)

$$y_{1n} = Y_{2n}\beta + X_n\gamma + u_n, \quad (1)$$

$$Y_{2n} = Z_n\Pi + X_n\Phi + V_n, \quad (2)$$

where y_{1n} and Y_{2n} are, respectively, an $n \times 1$ vector and an $n \times G$ matrix of observations on the $G+1$ endogenous variables of the system, X_n is an $n \times J$ matrix of observations on the J exogenous variables included in the structural equation (1), Z_n is an $n \times K_n$ matrix of observations on the K_n instrumental variables, or exogenous variables excluded from the structural equation (1), and u_n, V_n are, respectively, an $n \times 1$ vector and an $n \times G$ matrix of random disturbances. Further, let $\eta_i = (u_i, v_i')'$ where u_i and v_i' are, respectively, the $i\underline{th}$ component of the random vector u_n and the $i\underline{th}$ row of the random matrix V_n. The following assumptions are used in the sequel.

Assumption 1: $\Pi = \Pi_n = \frac{C_n}{b_n}$ for some sequence of positive real numbers $\{b_n\}$, nondecreasing in n, and for some sequence of nonrandom, $K_n \times G$ parameter matrices $\{C_n\}$.

Assumption 2: Let $\{\overline{Z}_{n,i} : i = 1, ..., n;\ n \geq 1\}$ be a triangular array of R^{K_n+J}-valued random variables, where $\overline{Z}_{n,i} = (Z_{n,i}',\ X_i')'$ with $Z_{n,i}'$ and X_i' denoting the $i\underline{th}$ row of the matrices Z_n and X_n, respectively. Moreover, suppose that:

(a) $K_n \to \infty$ as $n \to \infty$ such that $\frac{K_n}{n} \to \alpha$ for some constant α satisfying $0 \leq \alpha < 1$.

the concentration parameter diverges is actually slower than K_n (i.e., the case where $\frac{r_n}{K_n} \to 0$, but $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$). Thirdly, Stock and Yogo (2001) do not study as broad a class of estimators as we do, as their analysis does not include the $JIVE$ estimator.

[4]Another related paper is Donald and Newey (2001), which studies mean square error (MSE) properties of various single-equation estimators and proposes procedures for choosing instruments on the basis of MSE criteria. Like Morimune (1983) and Bekker (1994), Donald and Newey (2001) also employ a many-instrument setup, but without a local-to-zero structure. Indeed, the consistency results presented in this paper can be viewed as extending some of the results of Donald and Newey (2001) to the more general case where the asymptotic analysis is conducted using a sequence of (approximating) simultaneous equations models; this setup, in turn, allows for local-to-zero modeling of weak instruments.

(b) There exists a sequence of positive real numbers $\{m_{1n}\}$, nondecreasing in n, and constants D_1 and D_2, with $0 < D_1 \leq D_2 < \infty$, such that

$$D_1 \leq \underline{\lim}_{n\to\infty} \lambda_{\min}\left(\frac{\overline{Z}_n'\overline{Z}_n}{m_{1n}}\right) \quad a.s. \tag{3}$$

and

$$\overline{\lim}_{n\to\infty} \lambda_{\max}\left(\frac{\overline{Z}_n'\overline{Z}_n}{m_{1n}}\right) \leq D_2 \quad a.s., \tag{4}$$

where $\overline{Z}_n = (Z_n \ X_n)$.

(c) There exists a sequence of positive real numbers $\{m_{2n}\}$, nondecreasing in n, and constants D_3 and D_4, with $0 < D_3 \leq D_4 < \infty$, such that

$$D_3 \leq \underline{\lim}_{n\to\infty} \lambda_{\min}\left(\frac{C_n'C_n}{m_{2n}}\right) \tag{5}$$

and

$$\overline{\lim}_{n\to\infty} \lambda_{\max}\left(\frac{C_n'C_n}{m_{2n}}\right) \leq D_4. \tag{6}$$

Assumption 3: $\eta_i \mid \overline{Z}_n \equiv i.i.d.(0,\Sigma)$ almost surely for all n, where $\eta_i \mid \overline{Z}_n$ denotes the conditional distribution of η_i given $\overline{Z}_n$. Further, assume that $\Sigma > 0$, and partition Σ conformably with $(u_i, v_i')'$ as $\Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_{VV} \end{pmatrix}$. Also, define σ_{Vu}^g to be the g^{th} element of σ_{Vu} and $\Sigma_{VV}^{(g,h)}$ to be the $(g,h)^{th}$ element of Σ_{VV}.
Assumption 4: Let $\eta_{i,h}$ be the h^{th} element of η_i with $\eta_{i,j}$, $\eta_{i,k}$, and $\eta_{i,l}$ similarly defined, and suppose that $E|\eta_{i,h}\eta_{i,j}\eta_{i,k}\eta_{i,l}| < \infty$, for $h,j,k,l = 1,...,G+1$. Moreover, for each h, j, k, and l and for all n,

$$E\left([\eta_{i,h}\eta_{i,j}\eta_{i,k}\eta_{i,l}] \mid \overline{Z}_n\right) = \mu_{hjkl} \quad a.s.,$$

where $\mu_{hjkl} = E[\eta_{i,h}\eta_{i,j}\eta_{i,k}\eta_{i,l}]$, the unconditional expectation.
Assumption 5: Define the ratio $r_n = \frac{m_{1n}m_{2n}}{b_n^2}$, and suppose that as $n \to \infty$, $\frac{r_n}{n} \to \kappa$ for some constant κ, such that $0 \leq \kappa < \infty$.

Remark 2.1: (i) Note that the inequality condition (3) in Assumption 2(b) ensures that there exists some (positive) integer N such that for all $n \geq N$, $\left(\overline{Z}_n'\overline{Z}_n\right)/m_{1n}$ is positive definite and, thus, nonsingular with probability one. Moreover, (3) and (4) together imply that the rate of growth of $\overline{Z}_n'\overline{Z}_n$ is the same in all directions of the data (with probability one), so that Assumption 2(b) rules out cases where there may be different rates of information accumulation along different directions, such as the case when one has both trending regressors and non-trending regressors. We do not consider the case where there are multiple rates of information accumulation as this case does not typically arise in empirical situations where there is a weak-instrument problem. In addition, note that Assumption 2(b) is more general than the standard condition in IV regression (with a fixed number of instruments), where $(Z'Z)/n$ is assumed to converge to a positive definite matrix. In particular, this assumption allows us to accommodate cases where the information in the instruments accumulates at a rate different from n.[5]

[5]It should be further noted that the case where $m_{1n} = n$ already accommodates a wide variety of possible instrumental variable designs. In particular, Portnoy (1984, 1985, 1987) has shown that, for the case $m_{1n} = n$, conditions similar to (3) and (4) hold in probability for a wide class of random designs. See also the discussion in Andrews (1991) and Koenker and Machado (1999).

(ii) In order to give an interpretation for r_n, note that under Assumption 2:

$$r_n^{-1}\left(\Sigma_{VV}^{-\frac{1}{2}}\Pi_n Z_n' Q_{X_n} Z_n \Pi_n \Sigma_{VV}^{-\frac{1}{2}}\right) = O_{a.s.}(1),$$

so that the concentration parameter matrix (i.e. $\Sigma_{VV}^{-\frac{1}{2}}\Pi_n Z_n' Q_{X_n} Z_n \Pi_n \Sigma_{VV}^{-\frac{1}{2}}$), when standardized by r_n, is bounded almost surely. Moreover, under Assumptions 2, there exists a positive integer N such that for all $n \geq N$, $(r_n)^{-1}\Pi_n Z_n' Q_{X_n} Z_n \Pi_n$ is nonsingular with probability one, so that the concentration parameter $\Sigma_{VV}^{-\frac{1}{2}}\Pi_n Z_n' Q_{X_n} Z_n \Pi_n \Sigma_{VV}^{-\frac{1}{2}}$ is not of an order less than r_n in any direction, almost surely. Hence, we see that r_n can be interpreted as the rate at which the concentration parameter $\Sigma_{VV}^{-\frac{1}{2}}\Pi_n Z_n' Q_{X_n} Z_n \Pi_n \Sigma_{VV}^{-\frac{1}{2}}$ grows (if it grows) as n increases. In the sequel, we shall pay particular attention to the case where $r_n \to \infty$, as $n \to \infty$. In fact, we shall argue in the next remark that r_n may diverge even in the case where the available instruments are weak in the local-to-zero sense.
(iii) To see the relationship between our framework and that of Staiger and Stock (1997), note that in the Staiger-Stock setup, $b_n = \sqrt{n}$ and $m_{1n} = n$. Additionally, Staiger and Stock (1997) take the number of instruments to be fixed so that in their case the matrix C has a fixed number of columns, say $K_n = \overline{K}$ for all n, so that C does not depend on n. In consequence, $C'C = O(1)$, so we can take $m_{2n} = 1$, $\forall n$. It follows that in their setup

$$r_n = \frac{m_{1n} m_{2n}}{b_n^2} = 1, \quad \text{for all } n.$$

Hence, r_n does not diverge as $n \to \infty$ in their setup; and, as they have shown, none of the usual single-equation estimators consistently estimate β. A main focus of this paper is to add to Staiger and Stock's results by allowing K_n to grow to infinity as a function of n, in which case it is possible for the concentration parameter to diverge (i.e. $r_n \to \infty$, as $n \to \infty$) even if we set $b_n = \sqrt{n}$, and even if we take the rate of information accumulation in the instruments to be n, as is the case in many standard designs. To see this, note that as long as the available instruments are not "too" weak, so that, as $K_n \to \infty$, the elements of $C_n' C_n$ grow in such a way that the sequence m_{2n} which satisfies conditions (5) and (6) tends to infinity as $n \to \infty$; then, even in the standard local-to-zero setup where $b_n = \sqrt{n}$ and $m_{1n} = n$, we have that

$$r_n = \frac{m_{1n} m_{2n}}{b_n^2} = m_{2n} \to \infty \ , \quad \text{as} \quad n \to \infty.$$

Thus, as we will see from the results of the next section, the use of many instruments (as modeled by taking K_n to infinity as $n \to \infty$) has potential benefits for point estimation since it provides an extra source by which the concentration parameter can grow.
(iv) We should also briefly compare and contrast our setup with the many-instrument asymptotic framework of Bekker (1994).[6] To keep this comparison focused on the essential features of our framework via-à-vis that of Bekker (1994), we shall concentrate our discussion on the case where there is no included exogenous variables, so that $J = 0$. Within this setup, the alternative asymptotics considered by Bekker (1994), in our notations, boils down to one where the quantity $(n - G)^{-1}\Pi' Z_n' Z_n \Pi$ is kept fixed, as both K_n and n go to infinity, such that $\frac{K_n}{n} \to \alpha$, for some constant α satisfying $0 \leq \alpha < 1$. However, unlike our setup and that of Staiger and Stock (1997), Bekker (1994) does not model Π as being local-to-zero. Hence, within our framework, the Bekker approach is essentially one of setting $b_n = 1$, $\forall n$ and $r_n = n$. Unlike the Bekker setup, we do not require $(n-G)^{-1}\Pi' Z_n' Z_n \Pi$

[6]The type of asymptotic approximation used by Bekker (1994) dates back to the work of Anderson (1976), Kunitomo (1980), and Morimune (1983), as is pointed out by Bekker in his paper. For further discussion of this type of asymptotics, see Hahn (1997) and Hahn and Inoue (2000).

to stay fixed, or even to converge to a limit as n grows, and we do not make a Gaussian error assumption (see below for further discussion).

(v) As consistent estimation of β turns out to depend crucially on how fast r_n approaches infinity relative to K_n as $n \to \infty$, it is natural to measure the quality (or, conversely, the weakness) of instruments in terms of the relative order of magnitudes of r_n and K_n. In particular, we propose the following taxonomy of instruments in terms of their quality:[7],

(a) *Refer to the set of available instruments as "not weak" if* $\frac{K_n}{r_n} \to 0$, *as* $n \to \infty$.

(b) *Refer to the set of available instruments as "mildly weak" if* $\frac{r_n}{K_n} \to \delta_1$, *for some constant* δ_1 *such that* $0 < \delta_1 < \infty$.

(c) *Refer to the set of available instruments as "moderately weak" if* $\frac{r_n}{K_n} \to 0$, *bur* $\frac{\sqrt{K_n}}{r_n} \to 0$, *as* $n \to \infty$.

(d) *Refer to the set of available instruments as "completely weak" if* $\frac{r_n}{\sqrt{K_n}} \to \delta_2$, *for some constant* δ_2 *such that* $0 \leq \delta_2 < \infty$[8].

By classifying instrument weakness in terms of the rate at which the concentration parameter grows, we are taking note of the fact that in some sense there are two forces at play in determining whether consistent estimation can be attained. To be more specific, note first that it is useful to think of r_n as being the product of two components, m_{1n} and $\frac{m_{2n}}{b_n^2}$. The latter ratio gauges the degree to which Π_n is local-to-zero in the sense that, given Assumptions 1 and 2, there exist real constants $\underline{D}$ and $\overline{D}$with $0 < \underline{D} \leq \overline{D} < \infty$ and some positive integer N^* such that $\forall n \geq N^*$, $\underline{D}\frac{m_{2n}}{b_n^2} \leq ||\Pi_n||^2 \leq \overline{D}\frac{m_{2n}}{b_n^2}$, where $||.||$ denotes the usual Euclidean norm so that $||\Pi_n|| = \sqrt{Tr\left(\Pi_n'\Pi_n\right)}$. Hence, in the case where $\frac{m_{2n}}{b_n^2} \to 0$ as $n \to \infty$, $\frac{m_{2n}}{b_n^2}$ is the rate at which we are shrinking the (squared) distance of Π_n from the origin, in order to obtain an appropriate model for the near identification failure that may be present in a given empirical situation. On the other hand, m_{1n} is the rate at which the information in the instrumental variables accumulates and, thus, reflects the information content of the available instruments. Both m_{1n} and $\frac{m_{2n}}{b_n^2}$ clearly play a role in determining whether consistent estimation is achievable. Indeed, from the viewpoint of point estimation, one is not necessarily better off to be in a situation where the available instruments have low information content but enter into the first-stage equation with relatively larger coefficient values relative to an alternative situation where the instruments have more information content but smaller first-stage coefficient values.

(vi) Assumption 4 requires the disturbances of the model to have finite absolute fourth moments. In addition, we require that the conditional fourth order moments of the disturbances given $\overline{Z}_n$ to be equal to the unconditional fourth moments. This latter condition is implied by an assumption of the independence of η_i and $\overline{Z}_n$ for all i

[7]A potentially interesting future line of research is the exploration of diagnostic procedures which will give empirical investigators a sense of whether the set of available instruments might be "too" weak for useful estimation to take place. Some of the measures of instrument relevance discussed in Shea (1997) and Hall and Peixe (2000) might be useful in this regard.

[8]As we shall see in the next section, all of the usual estimators, including both $2SLS$ and $LIML$, are consistent if the set of available instruments is "not weak" in the sense defined above, which of course includes as a special case the conventional setting, where full (asymptotic) identification is assumed and where the number of available instruments is taken to be fixed, so that $K_n = \overline{K}$, for some fixed constant $\overline{K}$, and $r_n = n \to \infty$. On the other hand, if the instruments are "mildly weak" or "moderately weak", $LIML$, $JIVE$, and other ω-class estimators satisfying certain general conditions are consistent while $2SLS$ is not. Additionally, if instruments are "completely weak" (the case examined by Staiger and Stock (1997) and also Wang and Zivot (1998)), then none of the currently known estimators are likely to be consistent.

and n but is, of course, weaker than such an assumption. Note also that the conditions on the instruments and the errors which we stipulate in Assumptions 2, 3, and 4 are weaker than that of Morimune (1983) and Bekker (1994), which assumed fixed instruments and *i.i.d.* Gaussian errors.
(vii) Assumption 5 stipulates that the rate of growth of the concentration parameter r_n must be no faster than n. This assumption is in accord with our objective of studying the case of weak instruments and weak identification. The case where the concentration parameter grows at a rate faster than n is often a case where the signal from the model is so strong that OLS will be consistent even in the presence of endogeneity. This, however, is not the scenario which we address in this paper.
(viii) It is of further interest to compare our setup with that of an important recent paper by Donald and Newey (2001). Note, in particular, that our setup can be regarded as being more general than that of Donald and Newey (2001) in several respects. First, our conditions do not require the triangular array of exogeneous regressors $\overline{Z}_{n,i}$ to be *i.i.d.* for given n, whereas Donald and Newey (2001) explicitly assume that their vector of exogenous variables is an *i.i.d.* sequence. Second, we also study the case where the number of instruments, K_n, diverges at the same rate as n (i.e. the case where the degree of overidentification is significant relative to the sample size), whereas Donald and Newey (2001) examine the case where $\frac{K_n}{n} \to 0$ as $n \to \infty$. Finally, as mentioned earlier, our asymptotics is based on a sequence of (local-to-zero) models and not a sequence of data generated by the same model, as in the setup considered in Donald and Newey (2001).

3 Asymptotic Behavior of Single-equation Estimators

Since the class of estimators that we examine in this paper is an extension of the well-known k-class, we begin by recalling that the k-class estimator can be written in the form:

$$\widehat{\beta}(k)_n = \left(Y_{2n}'Q_{X_n}Y_{2n} - kY_{2n}'Q_{\overline{Z}_n}Y_{2n}\right)^{+}\left(Y_{2n}'Q_{X_n}y_{1n} - kY_{2n}'Q_{\overline{Z}_n}y_{1n}\right), \tag{7}$$

where k may either be a non-random scalar or may depend on the data and, hence, also on the sample size n.[9] In addition, it is well known that k-class estimators can be viewed as instrumental variables estimators which use as instruments the "adjusted" endogenous regressors:

$$\widehat{Y}_{2n}(k)_i = Y_{2n,i} - k\widehat{v}_{n,i}, \quad i = 1, ..., n, \tag{8}$$

where $\widehat{v}_{n,i}$ is the transpose of the ith row of $\widehat{V}_n = Q_{\overline{Z}_n}Y_{2,n}$, the matrix of OLS residuals from estimating the first-stage equation (see pp. 166 of Schmidt, (1976) for a more detailed discussion of this interpretation of the k-class estimator). Hence, the k-class estimator seeks to remove from the ith observation of the endogenous regressors, $Y_{2n,i}$, that part which is correlated with u_i. It does this by subtracting from $Y_{2n,i}$ a component which is equal to (some scalar multiple) k times an estimate of v_i, since the presence of the disturbance component v_i in $Y_{2n,i}$ is precisely what causes the endogeneity.

Now to motivate our interest in generalizing the k-class, note that by definition, k-class estimators take the scalar multiple k to be invariant with respect to i (i.e. k does not vary with the observation). As a result, the k-class is not a rich enough class to include some interesting, recently proposed estimators, such as the $JIVE$ of Angrist, Imbens, and Krueger (1999). In order to include the $JIVE$ and other potentially interesting estimators,

[9] We define all estimators using the Moore-Penrose generalized inverse so as to allow for the possibility of perfect multicollinearity, for a given n. Asymptotically, however, our conditions rule out perfect multicollinearity.

we examine what we call the ω-class, which extends the k-class by allowing k (or, in our notation, ω) to possibly vary with both i and n. More specifically, in lieu of expression (8) above, the adjustment to the endogenous regressors used in constructing the ω-class estimators takes the more general form:

$$\widehat{Y}_{2n}(\omega_{i,n})_i = Y_{2n,i} - \omega_{i,n}\widehat{v}_{n,i}, \quad i = 1, ..., n. \tag{9}$$

Using $\widehat{Y}_{2n}(\omega_{i,n})_i$, $i = 1, ..., n$ as instruments, it is easy to show, by straightforward algebra, that ω-class estimators can be written as:

$$\widehat{\beta}_{\omega,n} = \left(Y_{2n}'\left[I_n - Q_{\overline{Z}_n}\Omega_n\right]Q_{X_n}Y_{2n}\right)^+\left(Y_{2n}'\left[I_n - Q_{\overline{Z}_n}\Omega_n\right]Q_{X_n}y_{1n}\right), \tag{10}$$

where Ω_n is a diagonal matrix of the form $\Omega_n = diag\left(\omega_{1,n},\ \omega_{2,n}, ..., \omega_{n,n}\right)$. Comparing expressions (7) and (10), it is apparent that every k-class estimator is a special case of the ω-class estimator (10) obtained by setting $\omega_{1,n} = \omega_{2,n} = ... = \omega_{n,n} = k$. In addition, note that it is often convenient to rewrite the estimator given by expression (10) above in the alternative form:

$$\widehat{\beta}_{\omega,n} = \left(Y_{2n}'\left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\widetilde{\Omega}_n Q_{X_n}\right]Y_{2n}\right)^+\left(Y_{2n}'\left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\widetilde{\Omega}_n Q_{X_n}\right]y_{1n}\right), \tag{11}$$

where $\widetilde{\Omega}_n = \Omega_n - I_n$, $\widetilde{\Omega}_n = diag\left(\widetilde{\omega}_{1,n},\ \widetilde{\omega}_{2,n}, ..., \widetilde{\omega}_{n,n}\right)$, and $\widetilde{\omega}_{i,n} = \omega_{i,n} - 1$, for $i = 1, ..., n$. Without further restrictions on $\omega_{i,n}$ $(i = 1, ..., n)$, or alternatively on $\widetilde{\omega}_{i,n}$ $(i = 1, ..., n)$, the ω-class estimator is not a consistent estimator of β under the assumptions of Section 1. Consistent estimation can be obtained, however, given the following restriction:

Assumption 6: Suppose that for each i and n, $\widetilde{\omega}_{i,n}$ can be decomposed into the sum of two components as follows:

$$\widetilde{\omega}_{i,n} = \overline{\omega}_{i,n} + \xi_{i,n}, \tag{12}$$

such that $\overline{\omega}_{i,n}$ is either non-random or depends only on the exogenous variables $\overline{Z}_n$, so that $\overline{\omega}_{i,n} = f_{n,i}(\overline{Z}_n)$. Also, assume that $\overline{\omega}_{i,n}$ and $\xi_{i,n}$ satisfy the following conditions:

(a) $\overline{\lim_{n\to\infty}}\ \bar{l}_n < \infty$ *a.s.*, where $\bar{l}_n = \sup_{1\le i\le n} |\overline{\omega}_{i,n}|$.

(b) $\sum_{i=1}^{n} \overline{\omega}_{i,n}\left(1 - h_{i,n}\right) = K_n$ *a.s.* $\forall n$,
where $h_{i,n}$ is the ith diagonal element of $P_{\overline{Z}_n}$;

(c) $\sum_{i=1}^{n} E\left(\overline{\omega}_{i,n}^2\right) = O(K_n)$;

(d) $\sup_{1\le i\le n} |\xi_{i,n}| = o_p\left(\frac{r_n}{n}\right)$.

Theorem 3.1: *Under Assumptions 1-6, let $\widehat{\beta}_{\omega,n}$ be defined as in equation (11) above. Suppose that $r_n \to \infty$ as $n \to \infty$, such that $\frac{\sqrt{K_n}}{r_n} \to 0$. Then,*

$$\widehat{\beta}_{\omega,n} \xrightarrow{p} \beta_0 \quad as\ n \to \infty.. \tag{13}$$

Remark 3.2: (i) Part (a) of Assumption 6 rules out unreasonable choices of $\widetilde{\omega}_{i,n}$, which may lead to pathological asymptotic behavior of the ω-class estimator. To better understand parts (b)-(d) of Assumption 6, it is helpful to focus discussion on the special case where $J = 0$, so that there are no included exogenous regressors. As mentioned above, an $\omega-$class estimator can be viewed as an IV estimator where the matrix of instrumental variables is given by $\widehat{Y}_{2n}\left(\widetilde{\Omega}_n\right) = \left[P_{Z_n} - \widetilde{\Omega}_n Q_{Z_n}\right] Y_{2n}$, in the case where $J = 0$. Now, let W denote the generic matrix of observations on the instrumental variables. Then, the standard assumption for the validity of instruments can be stated in terms of the moment condition

$$E(W'u) = 0, \tag{14}$$

where u denotes the vector of disturbances in the structural equation of interest. Note, however, that this condition is violated by many members of the $\omega-$class, as this class contains many estimators for which $E\left[\widehat{Y}_{2n}\left(\widetilde{\Omega}_n\right)' u_n\right] \neq 0$. On the other hand, consistency of the ω-class estimators under the asymptotic scheme described in Theorem 3.1 does not require a condition such as (14) to hold exactly; rather, consistency can be attained if an orthogonality condition analogous to (14) holds asymptotically. In particular, consistency can be achieved if:

$$\frac{\widehat{Y}_{2n}\left(\widetilde{\Omega}_n\right)' u_n}{r_n} \xrightarrow{p} 0 \tag{15}$$

under the asymptotic scheme described in Theorem 3.1. Given the other assumptions above, (15) can be shown to hold for those $\omega-$class estimators which satisfy Assumption 6. To understand the role which Assumption 6 plays in establishing (15), note that under this assumption, we can decompose $\widehat{Y}_{2n}\left(\widetilde{\Omega}_n\right)$ into two components as follows: $\widehat{Y}_{2n}\left(\widetilde{\Omega}_n\right) = \widehat{Y}_{2n}^{(1)} - \widehat{Y}_{2n}^{(2)}$, where $\widehat{Y}_{2n}^{(1)} = \left[P_{Z_n} - \overline{\Omega}_n Q_{Z_n}\right] Y_{2n}$ and $\widehat{Y}_{2n}^{(2)} = \Xi_n Q_{Z_n} Y_{2n}$ with $\overline{\Omega}_n = diag\left(\overline{\omega}_{1,n}, \overline{\omega}_{2,n}, ..., \overline{\omega}_{n,n}\right), \Xi_n = diag\left(\xi_{1,n},\ \xi_{2,n}, ..., \xi_{n,n}\right)$, and $\overline{\omega}_{i,n},\ \xi_{i,n}\ (i = 1, ..., n)$ defined as in Assumption 6. Part (b) of Assumption 6 helps to ensure that $\widehat{Y}_{2n}^{(1)}$ satisfies an orthogonality condition of the form (14). To see this, note that

$$\begin{aligned} E\left(\widehat{Y}_{2n}^{(1)\prime} u_n\right) &= E\left(Y_{2n}'\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n\right) \\ &= E\left(\frac{C_n' Z_n' u_n}{b_n}\right) + E\left(V_n'\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n\right) = 0, \end{aligned} \tag{16}$$

given that $E\left(\frac{C_n' Z_n' u_n}{b_n}\right) = 0$ by Assumption 3, and given that

$$\begin{aligned} e_g' E\left(V_n'\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n\right) &= E\left(V_n^{(g)\prime}\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n\right) \\ &= \sigma_{Vu}^g E_{\overline{Z}_n}\left[K_n - \sum_{i=1}^n \overline{\omega}_{i,n}(1 - h_{i,n})\right] = 0, \quad g = 1, ..., G, \end{aligned} \tag{17}$$

by Assumption 6(b), where e_g and $V_n^{(g)}$ denote, respectively, the g^{th} column of the $G \times G$ identity matrix I_G and of V_n and where $E_{\overline{Z}_n}(\cdot)$ denotes the expectation taken with respect to the probability measure of $\overline{Z}_n$.

Part (c) of Assumption 6 puts a restriction on the growth rate of the sum of the second moments of $\overline{\omega}_{i,n}$, $i = 1, ..., n$ (or on the sum of squared $\overline{\omega}_{i,n}$, in the case where the $\overline{\omega}_{i,n}$'s are non-random), as n increases. Without some control on the sum of second moments of $\overline{\omega}_{i,n}$, the variance of $\frac{\widehat{Y}_{2n}^{(1)\prime} u_n}{r_n}$ may not die down asymptotically and $\frac{\widehat{Y}_{2n}^{(1)\prime} u_n}{r_n}$ may not converge in probability to zero as desired.

Finally, part (d) of Assumption 6 ensures that $\frac{\widehat{Y}_{2n}^{(2)\prime} u_n}{r_n}$ converges to zero in probability. To see this, consider the g^{th} element of $\frac{\widehat{Y}_2^{(2)\prime} u_n}{r_n}$, and note that

$$\left|\frac{e_g' \widehat{Y}_{2n}^{(2)\prime} u_n}{r_n}\right| = \left|\frac{V_n^{(g)\prime}\left[Q_{Z_n}\Xi_n\right] u_n}{r_n}\right| \leq \sqrt{\frac{V_n^{(g)\prime} Q_{Z_n} V_n^{(g)}}{n}} \sqrt{\frac{u_n' u_n}{n}} \left[\left(\frac{n}{r_n}\right) \sup_i |\xi_{i,n}|\right], \tag{18}$$

so that part (d) of assumption 6 implies that $\left|\frac{e_g' \widehat{Y}_{2n}^{(2)\prime} u_n}{r_n}\right| \xrightarrow{p} 0$, for $g = 1, ..., G$.

It should also be noted that the set of estimators satisfying assumption 6 is definitely not empty. In fact, as we will verify below, both $LIML$ and $JIVE$ can be shown to satisfy this assumption.

(ii) Note that, given the other assumptions, the consistency result obtained in Theorem 3.1 holds regardless of whether K_n grows at the same rate as n or at a rate slower than n, as long as $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$. Hence, our analysis shows that it is not the relative magnitudes of K_n and n per se, which determines whether consistency can be attained. Rather, what is more important is the speed of divergence of the concentration parameter relative to K_n.

(iii) An important member of the ω-class is the $LIML$ estimator. The $LIML$ estimator is defined as follows:

$$\widehat{\beta}_{LIML,n} = \left(Y_{2n}' Q_{X_n} Y_{2n} - \widehat{\lambda}_{LIML,n} Y_{2n}' Q_{\overline{Z}_n} Y_{2n}\right)^{+} \left(Y_{2n}' Q_{X_n} y_{1n} - \widehat{\lambda}_{LIML,n} Y_{2n}' Q_{\overline{Z}_n} y_{1n}\right), \tag{19}$$

where $\widehat{\lambda}_{LIML,n}$ is the smallest root of the determinantal equation:

$$\det\left\{\begin{pmatrix} y_{1n}' Q_{X_n} y_{1n} & y_{1n}' Q_{X_n} Y_{2n} \\ Y_{2n}' Q_{X_n} y_{1n} & Y_{2n}' Q_{X_n} Y_{2n} \end{pmatrix} - \lambda_n \begin{pmatrix} y_{1n}' Q_{\overline{Z}_n} y_{1n} & y_{1n}' Q_{\overline{Z}_n} Y_{2n} \\ Y_{2n}' Q_{\overline{Z}_n} y_{1n} & Y_{2n}' Q_{\overline{Z}_n} Y_{2n} \end{pmatrix}\right\} = 0 \tag{20}$$

The $LIML$ estimator can be obtained as a specific member of the ω-class by setting $\omega_{1,n} = \omega_{2,n} = ... = \omega_{n,n} = \widehat{\lambda}_{LIML,n}$ (or, alternatively, by setting $\widetilde{\omega}_{1,n} = \widetilde{\omega}_{2,n} = ... = \widetilde{\omega}_{n,n} = \widehat{\lambda}_{LIML,n} - 1$).[10] Moreover, it can be shown that the $LIML$ estimator satisfies Assumption 6, so that it is a consistent estimator of β, in the sense of Theorem 3.1. To verify that this is true, we first provide a result on the limiting behavior of $\widehat{\lambda}_{LIML,n}$ under the asymptotic scheme described in Theorem 3.1.

Theorem 3.3: *Under Assumptions 1-6, let $\widehat{\lambda}_{LIML,n}$ be the smallest root of the determinantal equation given by (20). Suppose that $r_n \to \infty$, as $n \to \infty$, such that $\frac{\sqrt{K_n}}{r_n} \to 0$. Then,*

$$\widehat{\lambda}_{LIML,n} = \frac{n - J}{n - K_n - J} + \xi_n, \tag{21}$$

where $\xi_n = o_p\left(\frac{r_n}{n}\right)$.

Given Theorem 3.3, it is convenient to set $\overline{\omega}_{i,n} = \left(\frac{n-J}{n-K_n-J}\right) - 1 = \frac{K_n}{n-K_n-J}$, for $i = 1, ..., n$. Now, to show that Assumption 6 is satisfied for the $LIML$ estimator, first observe that in this case:

$$\bar{l}_n = \sup_{1 \leq i \leq n} |\overline{\omega}_{i,n}| = \frac{K_n}{n - K_n - J}, \tag{22}$$

from which it follows that since $0 \leq \alpha < 1$,

$$\overline{\lim_{n\to\infty}}\, \bar{l}_n = \lim_{n\to\infty} \bar{l}_n = \frac{\alpha}{1-\alpha} < \infty, \tag{23}$$

[10]Of course, the $LIML$ estimator is also a k-class estimator, with $k = \widehat{\lambda}_{LIML,n}$.

so that Assumption 6(a) is satisfied. Next, observe that in this case:

$$\begin{aligned}\sum_{i=1}^{n} \overline{\omega}_{i,n} (1 - h_{i,n}) &= \left(\frac{K_n}{n - K_n - J} \right) \sum_{i=1}^{n} (1 - h_{i,n}) \\ &= \left(\frac{K_n}{n - K_n - J} \right) (n - K_n - J) = K_n, \end{aligned} \tag{24}$$

so that Assumption 6(b) is satisfied. Furthermore, note that in this case:

$$\sum_{i=1}^{n} E\left(\overline{\omega}_{i,n}^2\right) = K_n \left\{ \frac{K_n n}{(n - K_n - J)^2} \right\} = O(K_n), \tag{25}$$

since we assume that $K_n = O(n)$. Thus, Assumption 6(c) is satisfied. Finally, note that since ω does not vary with the observation in the $LIML$ case, we have that $\xi_{i,n} = \xi_n$, for all i. Hence, given that we have set $\overline{\omega}_{i,n} = \frac{K_n}{n-K_n-J}$, Theorem 3.3 implies that

$$\sup_i |\xi_{i,n}| = |\xi_n| = o_p\left(\frac{r_n}{n}\right), \tag{26}$$

so that Assumption 6(d) is satisfied as well.

(iv) Another important member of the ω-class is the Jackknife Instrumental Variables Estimator (or $JIVE$). $JIVE$, as proposed by Angrist, Imbens, and Krueger (1999) and also derived in Blomquist and Dahlberg (1999), is an estimator whose construction is based on a two-step procedure which can be described as follows:

1. First, construct jackknife fitted values $\widehat{Y}_{2n,i}^{JIVE}$, $i = 1, ..., n$, by running first stage OLS regressions using all but the ith observation (i.e. set $\widehat{Y}_{2n,i}^{JIVE} = \widehat{\Pi}_n(i)' Z_{n,i} + \widehat{\Phi}_n(i)' X_{n,i}$, for $i = 1, ..., n$, where $\left(\widehat{\Pi}_n(i)', \widehat{\Phi}_n(i)'\right)' = \left(\overline{Z}_n(i)' \overline{Z}_n(i)\right)^{-1} \left(\overline{Z}_n(i)' Y_2(i)\right)$ is the OLS estimator of the coefficient matrices of equation (2), obtained by deleting the ith observations, and where $\overline{Z}_n(i)$ denotes a submatrix of $\overline{Z}_n$ obtained by deleting the ith row from the latter).

2. Next, estimate the structural equation (1) by an IV procedure, using the matrix of instruments $\left(\widehat{Y}_{2n}^{JIVE}, X_n\right)$, where $\widehat{Y}_{2n}^{JIVE} = \left(\widehat{Y}_{2n,1}^{JIVE}, \widehat{Y}_{2n,2}^{JIVE}, ..., \widehat{Y}_{2n,n}^{JIVE}\right)'$.

Observe that an important motivation for the $JIVE$ method, as noted by Angrist, Imbens, and Krueger (1999), is that the jackknife fitted value $\widehat{Y}_{2n,i}^{JIVE}$, obtained from the delete-one estimation procedure described in the first step above, is uncorrelated with the structural disturbance u_i, even in finite sample; whereas, in the usual $2SLS$ construction, the fitted values obtained under the first stage regression are only *asymptotically* uncorrelated with the structural disturbance. It can be shown that the estimator constructed on the basis of steps 1 and 2 above can be written in the form:

$$\widehat{\beta}_{JIVE,n} = \left(Y_{2n}' \left[I_n - Q_{\overline{Z}_n} H_n\right] Q_{X_n} Y_{2n}\right)^+ \left(Y_{2n}' \left[I_n - Q_{\overline{Z}_n} H_n\right] Q_{X_n} y_{1n}\right), \tag{27}$$

where $H_n = diag\left(\frac{1}{1-h_{1,n}}, ..., \frac{1}{1-h_{n,n}}\right)$, with $h_{i,n}$ being the ith diagonal element of $P_{\overline{Z}_n}$. In addition, for $JIVE$ to be well-defined, we need to make the following assumption:

Assumption J[11]**:** There exists a constant $\overline{h}$, with $0 < \overline{h} < 1$, such that $0 \leq h_{i,n} \leq \overline{h}$ *a.s.* for $1 \leq i \leq n$ and for all n sufficiently large such that $P_{\overline{Z}_n}$ is well-defined almost surely[12].

Comparing expression (27) with expression (10) or expression (11), we see that $JIVE$ can also be obtained as a special case of the ω-class, by setting $\omega_{i,n} = \left(\frac{1}{1-h_{i,n}}\right)$, for $i = 1, ..., n$ or, alternatively, by setting $\widetilde{\omega}_{i,n} = \left(\frac{h_{i,n}}{1-h_{i,n}}\right)$, for $i = 1, ..., n$. Given Assumption J, it is not difficult to verify that $JIVE$ satisfies Assumption 6 and is thus a consistent estimator of β. To see that this is true, set:

$$\overline{\omega}_{i,n} = \left[h_{i,n} - \frac{J}{n}\right]\left(\frac{1}{1-h_{i,n}}\right), \tag{28}$$

for $i = 1, ..., n$; so that, by construction, $\xi_{i,n} = \frac{J}{n}\left(\frac{1}{1-h_{i,n}}\right)$. Now, observe that, in this case with probability one:

$$\bar{l}_n = \sup_{1\leq i\leq n} |\overline{\omega}_{i,n}| \leq \left[\overline{h} + \frac{J}{n}\right]\left(\frac{1}{1-\overline{h}}\right) \text{ for all } n \text{ sufficiently large,} \tag{29}$$

from which it follows that $\overline{\lim_{n\to\infty}} \bar{l}_n \leq \left(\frac{\overline{h}}{1-\overline{h}}\right) < \infty$ *a.s.*, so that Assumption 6(a) is satisfied. Next, observe that in the $JIVE$ case:

$$\begin{aligned}\sum_{i=1}^{n} \overline{\omega}_{i,n}(1-h_{i,n}) &= \sum_{i=1}^{n}\left[h_{i,n} - \frac{J}{n}\right]\left(\frac{1}{1-h_{i,n}}\right)(1-h_{i,n}) \\ &= \sum_{i=1}^{n}\left[h_{i,n} - \frac{J}{n}\right] = K_n + J - J = K_n,\end{aligned} \tag{30}$$

so that part (b) of Assumption 6 is satisfied. Moreover:

$$\begin{aligned}\sum_{i=1}^{n} E\left(\overline{\omega}_{i,n}^2\right) &= \sum_{i=1}^{n} E\left\{\left[h_{i,n} - \frac{J}{n}\right]^2\left(\frac{1}{1-h_{i,n}}\right)^2\right\} \\ &\leq \left(\frac{1}{1-\overline{h}}\right)^2 E\left\{\sum_{i=1}^{n}\left[h_{i,n}^2 - 2h_{i,n}\frac{J}{n} + \frac{J^2}{n^2}\right]\right\} \\ &\leq \left(\frac{1}{1-\overline{h}}\right)^2\left[K_n + J - 2\frac{(K_n+J)J}{n} + \frac{J^2}{n}\right] = O(K_n),\end{aligned} \tag{31}$$

where the second inequality follows from the fact that, even if we ignore Assumption J, it must be that $0 \leq h_{i,n} \leq 1$; and, hence, $\sum_{i=1}^{n} h_{i,n}^2 \leq \sum_{i=1}^{n} h_{i,n} = K_n + J$. It follows that Assumption 6(c) is also satisfied. Finally, note that:

$$\sup_{1\leq i\leq n} |\xi_{i,n}| = \sup_{1\leq i\leq n} \frac{J}{n}\left(\frac{1}{1-h_{i,n}}\right) \leq \frac{J}{n}\left(\frac{1}{1-\overline{h}}\right) = O(n^{-1}),$$

where the inequality holds by Assumption J, almost surely. Hence, Assumption 6(d) is satisfied as well.

[11] Note that Assumption J does rule out exogenous regressors of the form $e_i = (0, ..., 0, 1, 0, ..., 0)$, where e_i denotes the i^{th} elementary vector, or, alternatively, the i^{th} column of an identity matrix. It is easy to show that $h_{i,n} = 1$ if e_i is a column of $\overline{Z}_n$. The fact that $JIVE$ is not well-defined for this type of dummy regressor has not previously been pointed out in the literature, to the best of our knowledge.

[12] Note also that Assumption 2, part (b) implies that, for n sufficiently large, the matrix $\overline{Z}_n'\overline{Z}_n$ is positive definite almost surely, so that $P_{\overline{Z}_n}$ is well-defined almost surely for large enough n.

(v) Another interesting special case of the ω-class estimator is the $2SLS$ estimator[13], namely:

$$\widehat{\beta}_{2SLS,n} = \left(Y_{2n}'(P_{\overline{Z}_n} - P_{X_n})Y_{2n}\right)^{+} \left(Y_{2n}'(P_{\overline{Z}_n} - P_{X_n})y_{1n}\right). \tag{32}$$

However, unlike $LIML$ and $JIVE$, the $2SLS$ estimator does not satisfy Assumption 6, as can be seen by casual inspection; and so its asymptotic behavior is not covered by Theorem 3.1. To compare the limiting behavior of the $2SLS$ estimator with that of other ω-class estimators which do satisfy Assumption 6, the following result is useful.

Theorem 3.4: *Under Assumptions 1-5, let $\widehat{\beta}_{2SLS,n}$ be defined as in equation (32) above.*

1. *Assume that $\frac{r_n}{K_n} \to 0$ as $n \to \infty$; then*

$$\widehat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0 + \Sigma_{VV}^{-1}\sigma_{Vu} \quad as\ n \to \infty. \tag{33}$$

(b) *Assume instead that $\frac{r_n}{K_n} \to \delta$ as $n \to \infty$, for some constant δ such that $0 < \delta < \infty$; then*

$$\widehat{\beta}_{2SLS,n} - \beta_0 = \left(\delta\Psi_n + \Sigma_{VV}\right)^{-1}\sigma_{Vu} + o_p(1), \tag{34}$$

where $\Psi_n = \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2 r_n}$.

(c) *Assume, on the other hand, $\frac{K_n}{r_n} \to 0$ as $n \to \infty$; then*

$$\widehat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0 \quad as\ n \to \infty. \tag{35}$$

Note that part (a) of Theorem 3.4 shows that under the condition $\frac{r_n}{K_n} \to 0$, as $n \to \infty$ (i.e. when the set of available instruments is either "moderately weak" or "completely weak"), the $2SLS$ estimator, while inconsistent, does not converge to a random variable, as is the case when the number of instruments is held fixed (i.e. see Staiger and Stock (1997)). Rather, the $2SLS$ estimator in our context converges in probability to a nonrandom limit equaling the sum of β_0 and the bias term $\Sigma_{VV}^{-1}\sigma_{Vu}$. Interestingly, this convergence to a nonrandom limit holds even in the case where the concentration parameter does not diverge at all (i.e., even if r_n does not tend to infinity) as long as $K_n \to \infty$, as $n \to \infty$. This result is consistent with the result given in Chao and Swanson (2001) based on sequential asymptotics, where it is shown that the variance of the $2SLS$ estimator tends to zero as the number of instruments goes to infinity; but a non-zero bias remains. Turning to part (b) of Theorem 3.4, note that when r_n goes to infinity at the same rate as K_n (i.e. the case where the instruments are "mildly weak"), then the $2SLS$ estimator is again inconsistent. However, in this case, without further assumption, the $2SLS$ estimator need not converge at all since, under our conditions, we do not require the sequence of matrices $\{\Psi_n\}$ to converge to a limit. On the other hand, a sufficient condition which ensures that the $2SLS$ estimator

[13] Since it is well-known that the $2SLS$ estimator is a k-class estimator, with $k = 1$, it follows that $2SLS$ also belongs to the ω-class with $\omega_{1,n} = \omega_{2,n} = ... = \omega_{n,n} = 1$ (or with $\widetilde{\omega}_{1,n} = \widetilde{\omega}_{2,n} = ... = \widetilde{\omega}_{n,n} = 0$ in the alternative formulation of the ω-class estimator given in expression (11)).

will converge in probability to a non-random limit is the assumption that $\Psi_n \xrightarrow{p} \Psi$, as $n \to \infty$, for some constant positive definite matrix Ψ. Finally, part (c) of Theorem 3.4 shows that consistent estimation based on the $2SLS$ estimator (in the absence of bias correction) is only possible if the rate of growth of the concentration parameter r_n tends to infinity faster than the rate of growth of the instruments, K_n, as n increases (i.e. when the set of available instruments is "not weak"). In the case where the number of instruments is held fixed (so that we can set $m_{2n} = 1$, $\forall n$), this latter requirement is satisfied if the concentration parameter diverges, which in turn may occur even within a local-to-zero setup provided that b_n^2 goes to infinity at a rate slower than m_{1n}.

(vi) It is certainly of interest to compare our results with results obtained by Bekker (1994). Within his framework, Bekker finds that, if $\frac{K_n}{n} \to \alpha \neq 0$ as $n \to \infty$, then $2SLS$ is inconsistent whereas $LIML$ is consistent. On the other hand, if $\frac{K_n}{n} \to 0$, then both $2SLS$ and $LIML$ are consistent. These results can be obtained as special cases of the results provided in Theorem 3.1, Remark 3.2 (iii), and Theorem 3.4. To see this, first note that, since the Bekker framework implicitly takes $r_n = n$, assuming that $\frac{K_n}{n} \to \alpha \neq 0$ within the Bekker framework is equivalent to requiring the condition that $\frac{K_n}{r_n} \to \alpha \neq 0$, so we can deduce directly from Theorem 3.1, Remark 3.2(iii), and part (b) of Theorem 3.4 that $LIML$ is consistent but $2SLS$ is not. Alternatively, assuming that $\frac{K_n}{n} \to 0$ within the Bekker framework is equivalent to the condition that $\frac{K_n}{r_n} \to 0$, so that from Theorem 3.1, Remark 3.2(iii), and part (c) of Theorem 3.4, we deduce that, in agreement with the results of Bekker (1994), both estimators are consistent in this case.

It should be noted, however, that while it is true within the Bekker framework that the $2SLS$ estimator is consistent whenever K_n grows at a slower rate than n, this need not be the case more generally, especially in the presence of weak instruments. This is because whether or not the $2SLS$ estimator is consistent depends more crucially on the relative magnitude of r_n vis-a-vis K_n as $n \to \infty$, and not so much on the relative orders of magnitude of n and K_n, unless of course $r_n = n$, as in the Bekker framework. Hence, it is entirely possible for the instruments to be sufficiently weak (in the sense that r_n is of an order of magnitude much lower than n, for n large) so that we actually end up with a situation where $\frac{K_n}{n} \to 0$, but $\frac{r_n}{K_n} \to \delta$ for some nonnegative $\delta < \infty$, in which case $2SLS$ is actually inconsistent (see Theorem 3.4).

(vii) Interestingly, our analysis shows that ω-class estimators satisfying Assumption 6 may be consistent even if the weakness in the instruments is such that the concentration parameter grows at a rate slower than K_n, so long as r_n grows faster than $\sqrt{K_n}$ as $n \to \infty$[14]. Indeed, we conjecture that, given the other assumptions, the condition that $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$ is not just a sufficient condition but also a necessary condition for consistent estimation to be achievable in the presence of weak instruments. Proof of the necessity of this condition is, however, left to future research.

(viii) Our analysis also makes precise the sense in which ω-class estimators satisfying our Assumption 6 are more robust to instrument weakness than the $2SLS$ estimator. Specializing our general results to some specific estimators, we see that both $LIML$ and $JIVE$ satisfy Assumption 6 and are, thus, consistent even when the instruments are "mildly weak" or "moderately weak" in the terminology of Remark 2.1(v), whereas the $2SLS$

[14]Using the terminology introduced in Remark 2.1(v), we say that the set of available instruments is "moderately weak" if $\frac{r_n}{K_n} \to 0$ but $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$. It should be pointed out that the case with "moderately weak" instruments has not been studied at all by either Bekker (1994) or Staiger and Stock (1997). In fact, since $r_n = n$ in the Bekker framework, the cases studied by Bekker (1994) correspond to scenarios where the instruments are either "not weak" or are only "mildly weak" in our terminology, depending on whether $\frac{K_n}{n} \to 0$ or $\frac{K_n}{n} \to \alpha \neq 0$ as $n \to \infty$. Moreover, as discussed previously, Staiger and Stock (1997) studies the case where the number of instruments are fixed and the concentration parameter does not diverge, so that the ratio $\frac{\sqrt{K_n}}{r_n}$ does not vanish as $n \to \infty$ in their setup. Hence, the case they consider correspond to that where the instruments are "completely weak" in our terminology.

estimator does not satisfy Assumption 6 and is consistent only if the instruments are "not weak"[15]. In the case of $LIML$, our findings are entirely consistent with the numerical results presented in Staiger and Stock (1997), which show $LIML$ to be less biased than $2SLS$ when instruments are weak in the local-to-zero sense.
(ix) For models that are weakly identified, our results suggest that it might be sensible to use many instruments, when available, in constructing ω-class estimators which satisfy our Assumption 6[16]. This is because even if each instrument is only weakly correlated with the endogenous regressors, the combined effect of using a lot of them might nevertheless allow the concentration parameter to be sufficiently large so that reliable point estimation can be achieved.

4 Concluding Remarks

This paper puts forth general conditions under which consistent estimation can be attained in instrumental variables regression in the case where the available instruments are taken to be weak in the local-to-zero sense. In particular, we consider a general class of single-equation estimators, referred to as the ω-class, which extends the well-known k-class to include, amongst others, the Jackknife Instrumental Variables Estimator of Angrist, Imbens, and Krueger (1999). A main conclusion of our paper is that if the number of instruments is allowed to approach infinity as a function of the sample size, then ω-class estimators which satisfy certain general conditions, such as $LIML$ and $JIVE$, will be more robust to instrumental weakness relative to estimators such as the $2SLS$ estimator, which do not. Our results, thus, are useful in identifying point estimators whose performance is less adversely affected by the presence of weak instruments. In addition, our results suggest that the use of a large number of poor quality instruments may actually improve the reliability of point estimation in situations where the available instruments are weak and abundant.

A number of questions remain open. In particular, while we have shown that there exists a class of consistent estimators even when the instruments are modeled to be weak in the local-to-zero sense, we have not derived the asymptotic distributions of these consistent estimators. Obtaining the limiting distributions of these estimators will likely give us insights about the relative efficiency of the different estimators. Related to this question is the problem of finding the optimal estimator amongst the consistent members of the ω-class. Research is ongoing in these areas.

5 Appendix

We begin by providing three lemmas which are used in the subsequent proofs.
Lemma A1: *Under Assumptions 1-6, suppose that $r_n \to \infty$ as $n \to \infty$ such that $\frac{\sqrt{K_n}}{r_n} \to 0$. Then, the following statements are true.*

[15] Although, throughout this paper, we have used $LIML$ and $JIVE$ as specific examples of ω-class estimators which satisfy our Assumption 6, it should be noted that there are other known estimators, not explicitly considered here, which also satisfy this assumption. In particular, it is an easy exercise to verify that the modified $LIML$ estimators studied in Fuller (1977) and Morimune (1983) also satisfy Assumption 6 and are, thus, consistent in the sense of Theorem 3.1.

[16] Note that this prescription only applies to estimators which satisfy our Assumption 6, i.e., estimators such as $LIML$ and $JIVE$. On the other hand, in the absence of additional bias adjustment, increasing the number of instruments used in constructing the $2SLS$ estimator will increase the bias of that estimator as discussed in Chao and Swanson (2001), amongst other places.

(a) *let* $\Psi_n = \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2 r_n}$, *then* $\Psi_n = O_{a.s.}(1)$*; moreover, there exists a positive integer* N *such that* $\forall n \geq N$, Ψ_n *is positive definite almost surely;*

(b) $\frac{V_n' M_n u_n}{r_n} \xrightarrow{p} 0 \quad$ *as* $n \to \infty$;

(c) $\frac{V_n' M_n V_n}{r_n} \xrightarrow{p} 0 \quad$ *as* $n \to \infty$;

(d) $\frac{u_n' M_n u_n}{r_n} \xrightarrow{p} 0 \quad$ *as* $n \to \infty$;

(e) $\frac{C_n' Z_n' Q_{X_n} u_n}{b_n r_n} \xrightarrow{p} \quad$ *as* $n \to \infty$;

(f) $\frac{C_n' Z_n' Q_{X_n} V_n}{b_n r_n} \xrightarrow{p} \quad$ *as* $n \to \infty$;

(g) $\frac{V_n' M_n Z_n C_n}{b_n r_n} \xrightarrow{p} \quad$ *as* $n \to \infty$;

where $M_n = \left[\left(P_{\overline{Z}_n} - P_{X_n} \right) - Q_{\overline{Z}_n} \widetilde{\Omega}_n Q_{X_n} \right]$.

Proof of Lemma A1:

To prove part (a), we first show that Ψ_n is positive definite almost surely for n sufficiently large. To proceed, note that Assumption 2(b) implies that for every ε, where $0 < \varepsilon < D_1$, there exists $N < \infty$ such that for all $n \geq N$

$$\lambda_{\min}\left(\frac{\overline{Z}_n' \overline{Z}_n}{m_{1n}} \right) > D_1 - \varepsilon > 0 \quad a.s. \tag{36}$$

so that $\overline{Z}_n' \overline{Z}_n / m_{1n}$ is invertible almost surely for all $n \geq N$. It follows that for all $n \geq N$

$$\lambda_{\min}\left(\frac{\overline{Z}_n' \overline{Z}_n}{m_{1n}} \right) = \lambda_{\max}\left[\left(\frac{\overline{Z}_n' \overline{Z}_n}{m_{1n}} \right)^{-1} \right] \qquad a.s. \tag{37}$$

Moreover, it follows from the Poincaré Separation Theorem (see Magnus and Neudecker, 1988, pages 209-210) that for all $n \geq N$

$$\lambda_{\max}\left[\left(\frac{Z_n' Q_{X_n} Z_n}{m_{1n}} \right)^{-1} \right] \leq \lambda_{\max}\left[\left(\frac{\overline{Z}_n' \overline{Z}_n}{m_{1n}} \right)^{-1} \right] \quad a.s. \tag{38}$$

or

$$\lambda_{\min}\left(\frac{Z_n' Q_{X_n} Z_n}{m_{1n}} \right) \geq \lambda_{\min}\left(\frac{\overline{Z}_n' \overline{Z}_n}{m_{1n}} \right) > D_1 - \varepsilon > 0 \quad a.s. \tag{39}$$

Furthermore, note that

$$\begin{aligned} \lambda_{\min}(\Psi_n) &= \lambda_{\min}\left(\frac{C_n' Z_n' Q_{X_n} Z_n C_n}{m_{1n} m_{2n}} \right) \\ &\geq \lambda_{\min}\left(\frac{Z_n' Q_{X_n} Z_n}{m_{1n}} \right) \lambda_{\min}\left(\frac{C_n' C_n}{m_{2n}} \right) \\ &> (D_1 - \varepsilon) D_3 > 0 \qquad a.s. \text{ for all } n \geq N, \end{aligned} \tag{40}$$

where the last inequality above follows from Assumption 2(c) (in particular, expression (5)) and from expression (39) above. It follows that for all $n \geq N$, Ψ_n is positive definite almost surely.

Now, to show that $\Psi_n = O_{a.s.}(1)$, we note that Assumption 2(b) also implies that for every ε, where $0 < \varepsilon < \infty$, there exists positive integer $N < \infty$ such that for all $n \geq N$

$$\lambda_{\max}\left(\frac{\overline{Z}_n'\overline{Z}_n}{m_{1n}}\right) < D_2 + \varepsilon < \infty \quad a.s. \tag{41}$$

Moreover, by arguments similar to that given above, we can show that for all $n \geq N$

$$\lambda_{\max}\left(\frac{Z_n'Q_{X_n}Z_n}{m_{1n}}\right) \leq \lambda_{\max}\left(\frac{\overline{Z}_n'\overline{Z}_n}{m_{1n}}\right) < D_2 + \varepsilon < \infty \quad a.s. \tag{42}$$

so that

$$\begin{aligned}\lambda_{\max}(\Psi_n) &= \lambda_{\max}\left(\frac{C_n'Z_n'Q_{X_n}Z_nC_n}{m_{1n}m_{2n}}\right) \\ &\leq \lambda_{\max}\left(\frac{Z_n'Q_{X_n}Z_n}{m_{1n}}\right)\lambda_{\max}\left(\frac{C_n'C_n}{m_{2n}}\right) \\ &< (D_2+\varepsilon)D_4 \quad < \infty \qquad a.s. \text{ for all } n \geq N. \end{aligned} \tag{43}$$

For convenience set $\varepsilon = 1$ and let N^* be that value of N for which the inequality in (43) holds. It follows then that for all $n \geq N^*$, we can bound $\Psi_n^{(i,i)}$, the i^{th} diagonal element of Ψ_n, as follows:

$$\begin{aligned}\Psi_n^{(i,i)} &= \frac{e_i'C_n'Z_n'Q_{X_n}Z_nC_ne_i}{b_n^2r_n} \\ &\leq \lambda_{\max}\left(\frac{C_n'Z_n'Q_{X_n}Z_nC_n}{m_{1n}m_{2n}}\right) \leq (D_2+1)\,D_4 < \infty \qquad a.s., \end{aligned} \tag{44}$$

where e_i denotes the i^{th} column of the $G \times G$ identity matrix I_G. Moreover, for all $n \geq N^*$, we can also bound the absolute value of $\Psi_n^{(i,j)}$ (i.e., the absolute value of the $(i,j)^{th}$ element of Ψ_n for $i \neq j$) as follows:

$$\begin{aligned}\left|\Psi_n^{(i,j)}\right| &= \left|\frac{e_i'C_n'Z_n'Q_{X_n}Z_nC_ne_j}{b_n^2r_n}\right| \\ &\leq \sqrt{\frac{e_i'C_n'Z_n'Q_{X_n}Z_nC_ne_i}{b_n^2r_n}}\sqrt{\frac{e_j'C_n'Z_n'Q_{X_n}Z_nC_ne_j}{b_n^2r_n}} \\ &\leq \lambda_{\max}\left(\frac{C_n'Z_n'Q_{X_n}Z_nC_n}{b_n^2r_n}\right) \leq (D_2+1)\,D_4 < \infty \qquad a.s. \end{aligned} \tag{45}$$

It follows immediately from (44) and (45) that

$$\Psi_n = \frac{C_n'Z_n'Q_{X_n}Z_nC_n}{b_n^2r_n} = O_{a.s.}(1). \tag{46}$$

Before showing parts (b)-(g) of this lemma, we first note that, under Assumption 2, $\overline{Z}_n'\overline{Z}_n$ is nonsingular almost surely for n sufficiently large, as has been shown in the proof of part (a) of the lemma above (see expression (40)). It follows that, for n sufficiently large, the projection matrices $P_{\overline{Z}_n}$, P_{X_n}, $Q_{\overline{Z}_n}$, and Q_{X_n} are well-defined with probability one, and so is the matrix $M_n = \left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - Q_{\overline{Z}_n}\widetilde{\Omega}_nQ_{X_n}\right]$.

Now, to show part (b), it suffices to show that, under the assumptions of the lemma, the g^{th} element of $\frac{V_n'M_nu}{r_n}$ converges in probability to zero, i.e.

$$\frac{V_n^{(g)\prime}M_nu_n}{r_n} \xrightarrow{p} 0, \tag{47}$$

where $V_n^{(g)}$, $g \in \{1,...,G\}$, denotes an arbitrary gth column of V_n. To show (47), note first that, given Assumption 6 and n sufficiently large, we can write

$$\begin{aligned}\frac{V_n^{(g)\prime}M_n u}{r_n} &= \frac{V_n^{(g)\prime}\left[\left(P_{\overline{Z}_n}-P_{X_n}\right)-Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right]u_n}{r_n}-\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}\Xi_n Q_{X_n}u_n}{r_n}\\ &= \frac{V_n^{(g)\prime}\overline{M}_n u_n}{r_n}-\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}\Xi_n Q_{X_n}u_n}{r_n}, \qquad (48)\end{aligned}$$

where $\overline{\Omega}_n = diag\left(\overline{\omega}_{1,n}, \overline{\omega}_{2,n}, ..., \overline{\omega}_{n,n}\right)$ and $\Xi_n = diag\left(\xi_{1,n},\ \xi_{2,n}, ..., \xi_{n,n}\right)$. We will show that both of the terms on the right-hand side of (48) converge in probability to zero. To proceed, note that the expectation of the first term on the right-hand side of (48) can be shown to equal zero as follows:

$$\begin{aligned}E\left(\frac{V_n^{(g)\prime}\overline{M}_n u_n}{r_n}\right) &= E_{\overline{Z}_n}\left(Tr\left\{\frac{\overline{M}_n\left[E_{\eta|\overline{Z}_n}\left(u_n V_n^{(g)\prime}\right)\right]}{r_n}\right\}\right) = \frac{\sigma_{Vu}^{g}}{r_n}E_{\overline{Z}_n}\left(Tr\left\{\left(P_{\overline{Z}_n}-P_{X_n}\right)-Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right\}\right)\\ &= \frac{\sigma_{Vu}^{g}}{r_n}E_{\overline{Z}_n}\left(K_n-\sum_{i=1}^{n}\overline{\omega}_{i,n}\left(1-h_{i,n}\right)\right)=0, \qquad (49)\end{aligned}$$

where $E_{\overline{Z}_n}(\cdot)$ denotes the expectation taken with respect to the probability measure of $\overline{Z}_n$ and $E_{\eta|\overline{Z}_n}$ denotes the expectation taken with respect to the conditional probability measure of $\eta = (u, V)$ given $\overline{Z}_n$ and where the last equality above follows from Assumption 6(b). Next, let $\overline{m}_{ij,n}$ denote the $(i,j)^{th}$ element of $\overline{M}_n$ and let v_{ig} denote the $(i,g)^{th}$ element of V_n, and we can calculate the second moment of $\frac{V_n^{(g)\prime}\overline{M}_n u_n}{r_n}$ as follows:

$$\begin{aligned}\frac{1}{r_n^2}E\left(V_n^{(g)\prime}\overline{M}_n u_n\right)^2 &= \left(\frac{1}{r_n^2}\right)E_{\overline{Z}_n}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\overline{m}_{ij,n}\,\overline{m}_{kl,n}\,E_{\eta|\overline{Z}_n}\left[v_{ig}u_j v_{kg}u_l\right]\right\}\\ &= \left(\frac{1}{r_n^2}\right)E(v_{ig}^2u_i^2)E_{\overline{Z}_n}\left[\sum_{i=1}^{n}\overline{m}_{ii,n}^2\right]+\left(\frac{1}{r_n^2}\right)\Sigma_{VV}^{(g,g)}\sigma_{uu}E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}\overline{m}_{ij,n}^2+\sum_{j=2}^{n}\sum_{i=1}^{j-1}\overline{m}_{ij,n}^2\right]\\ &\quad +2\left(\frac{1}{r_n^2}\right)\left(\sigma_{Vu}^{g}\right)^2E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}+\sum_{i=2}^{n}\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right]\\ &= \mathcal{A}_n+\mathcal{B}_n+\mathcal{C}_n,\ \text{say}, \qquad (50)\end{aligned}$$

where the second equality above follows by noting that the expectation $E_{\eta|\overline{Z}_n}\left[v_{ig}u_j v_{kg}u_l\right]$ equals zero except in the cases where either $(i = j = k = l)$ or $(i = j$ and $k = l)$ or $(i = k$ and $j = l)$ or $(i = l$ and $j = k)$. Dealing with the term $\mathcal{A}_n$ first, observe that with probability one

$$\begin{aligned}
\left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[\sum_{i=1}^{n} \overline{m}_{ii,n}^2\right] &\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2) Tr\left[\overline{M}_n' \overline{M}_n\right] \\
&= \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2) Tr[\left(P_{\overline{Z}_n} - P_{X_n}\right) + Q_{X_n}\overline{\Omega}_n Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}] \\
&\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + \left|Tr\left(\overline{\Omega}_n Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right)\right|\right] \\
&\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + \sqrt{Tr\left(\overline{\Omega}_n Q_{\overline{Z}_n}\overline{\Omega}_n\right)}\sqrt{Tr\left(\overline{\Omega}_n Q_{X_n}\overline{\Omega}_n\right)}\right] \\
&\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + Tr\left(\overline{\Omega}_n^2\right)\right] \\
&= \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + \sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right], \qquad (51)
\end{aligned}$$

where the third inequality above follows from the Cauchy-Schwarz inequality and where the fourth inequality above follows from the fact that $Q_{\overline{Z}_n}$ and Q_{X_n} are symmetric, idempotent matrices. To see the argument behind the fourth inequality, take $Q_{\overline{Z}_n}$ as an example, and note that we can write $Q_{\overline{Z}_n} = B_n \Lambda_n B_n'$, where B_n is an orthogonal matrix (i.e., $B_n B_n' = I_n = B_n' B_n$) whose columns are the orthonormal eigenvectors of $Q_{\overline{Z}_n}$ and Λ_n is a diagonal matrix with $n - K_n - J$ one's and $K_n + J$ zero's along the main diagonal; hence, it follows that

$$Tr\left(\overline{\Omega}_n Q_{\overline{Z}_n}\overline{\Omega}_n\right) = Tr\left(\overline{\Omega}_n B_n \Lambda_n B_n'\overline{\Omega}_n\right) \leq Tr\left(\overline{\Omega}_n B_n B_n'\overline{\Omega}_n\right) = Tr\left(\overline{\Omega}_n^2\right), \qquad (52)$$

and, by a similar argument, $Tr\left(\overline{\Omega}_n Q_{X_n}\overline{\Omega}_n\right) \leq Tr\left(\overline{\Omega}_n^2\right)$. Now, note that since the bound given by (51) holds with probability one, we deduce that

$$\mathcal{A}_n = \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2) E_{\overline{Z}_n}\left[\sum_{i=1}^{n} \overline{m}_{ii,n}^2\right] \leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + E_{\overline{Z}_n}\left(\sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right)\right] = O\left(K_n / r_n^2\right); \qquad (53)$$

hence, $\mathcal{A}_n \to 0$ as $n \to \infty$ if $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$.

Turning our attention next to the term $\mathcal{B}_n$, we note that similar to the argument given for $\mathcal{A}_n$ above, we have that with probability one

$$\begin{aligned}
&\left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} \overline{m}_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1} \overline{m}_{ij,n}^2\right] \\
\leq &\left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu} Tr\left[\overline{M}_n' \overline{M}_n\right] \leq \left(\frac{1}{r_n^2}\right) \sigma_{VV}^{(g,g)} \sigma_{uu}\left[K_n + \sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right]. \qquad (54)
\end{aligned}$$

It follows again that since the bound given in (54) holds with probability one, we deduce that

$$\begin{aligned}
\mathcal{B}_n &= \left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu} E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} \overline{m}_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1} \overline{m}_{ij,n}^2\right] \\
&\leq \left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu}\left[K_n + E_{\overline{Z}_n}\left(\sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right)\right] = O\left(K_n / r_n^2\right), \qquad (55)
\end{aligned}$$

so that $\mathcal{B}_n \to 0$ as $n \to \infty$ if $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$.

Finally, turning to the term $\mathcal{C}_n$, we note that

$$\begin{aligned} & 2\left(\frac{1}{r_n^2}\right)\left(\sigma_{Vu}^g\right)^2\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}+\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right] \\ = & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left\{Tr\left[\overline{M}_n^2\right]+\left(Tr\left[\overline{M}_n\right]\right)^2-2\sum_{i=1}^n\overline{m}_{ii}^2\right\}, \end{aligned} \tag{56}$$

so that with probability one

$$\begin{aligned} & 2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}+\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right]\right| \\ = & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|Tr\left[\overline{M}_n^2\right]+\left(Tr\left[\overline{M}_n\right]\right)^2-2\sum_{i=1}^n\overline{m}_{ii}^2\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|Tr\left[\overline{M}_n^2\right]+\left(Tr\left[\overline{M}_n\right]\right)^2\right|+2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|Tr\left[\overline{M}_n'\overline{M}_n\right]\right| \\ = & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|Tr\left[P_{\overline{Z}_n}-P_{X_n}-Q_{\overline{Z}_n}\overline{\Omega}_n\left(P_{\overline{Z}_n}-P_{X_n}\right)+Q_{\overline{Z}_n}\overline{\Omega}_nQ_{\overline{Z}_n}\overline{\Omega}_nQ_{X_n}\right]\right|+2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|Tr\left[\overline{M}_n'\overline{M}_n\right]\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left[K_n+\left|Tr\left(Q_{\overline{Z}_n}\overline{\Omega}_nQ_{\overline{Z}_n}\overline{\Omega}_nQ_{X_n}\right)\right|\right]+2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|K_n+\sum_{i=1}^n\overline{\omega}_{i,n}^2\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2K_n+\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2Tr\left(\overline{\Omega}_nQ_{\overline{Z}_n}\overline{\Omega}_n\right)+2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|K_n+\sum_{i=1}^n\overline{\omega}_{i,n}^2\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2K_n+\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2Tr\left(\overline{\Omega}_n^2\right)+2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|K_n+\sum_{i=1}^n\overline{\omega}_{i,n}^2\right| \\ = & 3\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left[K_n+\sum_{i=1}^n\overline{\omega}_{i,n}^2\right], \end{aligned} \tag{57}$$

where the second equality from the top follows from the fact that under Assumption 6(b)

$$Tr\left(P_{\overline{Z}_n}-P_{X_n}-Q_{\overline{Z}_n}\overline{\Omega}_nQ_{X_n}\right)=K_n-Tr\left(Q_{\overline{Z}_n}\overline{\Omega}_n\right)=0 \quad a.s. \tag{58}$$

In addition, note that the third inequality in (57) follows from the Cauchy-Schwarz inequality and the fourth inequality follows from the same argument as given in expression (52) above. Since the upper bound given in (57) above holds almost surely, it follows that

$$\begin{aligned} |\mathcal{C}_n| = & 2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|E_{\overline{Z}_n}\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}+\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right]\right| \\ \leq & 2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2E_{\overline{Z}_n}\left|\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}+\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right| \\ \leq & 3\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left[K_n+\sum_{i=1}^nE\left(\overline{\omega}_{i,n}^2\right)\right] \\ = & O\left(K_n/r_n^2\right), \end{aligned} \tag{59}$$

so that $\mathcal{C}_n \to 0$ as $n \to \infty$ if $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$. It follows immediately from (53), (55), and (59) that

$$\frac{1}{r_n^2} E\left(V_n^{(g)\prime}\overline{M}_n u_n\right)^2 \to 0 \tag{60}$$

as $n \to \infty$ under the condition that $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$. Moreover, in view of (49) and (60), it follows as a direct consequence of Chebyshev's inequality that

$$\frac{V_n^{(g)\prime}\overline{M}_n u_n}{r_n} \xrightarrow{p} 0. \tag{61}$$

Next, we show that, under the assumptions of the lemma,

$$\frac{V_n^{(g)\prime} Q_{\overline{Z}_n} \Xi_n Q_{X_n} u_n}{r_n} \xrightarrow{p} 0. \tag{62}$$

To show this, note that

$$\left|\frac{V_n^{(g)\prime} Q_{\overline{Z}_n} \Xi_n Q_{X_n} u_n}{r_n}\right| \leq \sqrt{\frac{V_n^{(g)\prime} Q_{\overline{Z}_n} V_n^{(g)}}{r_n}} \sqrt{\frac{u_n' Q_{X_n} \Xi_n^2 Q_{X_n} u_n}{r_n}} \leq \left[\left(\frac{n}{r_n}\right) \sup_i |\xi_{i,n}|\right] \sqrt{\frac{V_n^{(g)\prime} Q_{\overline{Z}_n} V_n^{(g)}}{n}} \sqrt{\frac{u_n' Q_{X_n} u_n}{n}}, \tag{63}$$

where the first inequality above follows from Cauchy-Schwarz. Next, note that standard arguments yield

$$\frac{u_n' Q_{X_n} u_n}{n} \xrightarrow{p} \sigma_{uu} < \infty, \tag{64}$$

and, from part (e) of lemma A2 given below, we obtain

$$\frac{V_n^{(g)\prime} Q_{\overline{Z}_n} V_n^{(g)}}{n} \xrightarrow{p} \Sigma_{VV}^{(g,g)} (1-\alpha) < \infty \tag{65}$$

where $\Sigma_{VV}^{(g,g)}$ denotes the $(g,g)^{th}$ element of Σ_{VV}. Moreover, it follows from assumption 6(d) that

$$\left(\frac{n}{r_n}\right) \sup_i |\xi_{i,n}| \xrightarrow{p} 0. \tag{66}$$

(63), (64), (65), and (66) immediately imply that

$$\frac{V_n^{(g)\prime} Q_{\overline{Z}_n} \Xi_n Q_{X_n} u_n}{r_n} \xrightarrow{p} 0 \tag{67}$$

as n, K_n, $r_n \to \infty$ such that $\frac{K_n}{n} \to \alpha$ and $\frac{\sqrt{K_n}}{r_n} \to 0$. Now, the desired result for part (a) follows directly from (61) and (67).

To show part (c) of the lemma, note that, similar to part (b) above, it is sufficient to simply show that any arbitrary element of $\frac{V' M_n V}{r_n}$ converges in probability to zero, i.e., it is sufficient to show that, under the assumptions of the lemma,

$$\frac{V_n^{(g)\prime} M_n V_n^{(h)}}{r_n} = \frac{V_n^{(g)\prime}\left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - Q_{\overline{Z}_n}\widetilde{\Omega}_n Q_{X_n}\right] V_n^{(h)}}{r_n} \xrightarrow{p} 0 \tag{68}$$

where $\frac{V_n^{(g)\prime} M_n V_n^{(h)}}{r_n}$ is the $(g,h)^{th}$ element of $\frac{V' M_n V}{r_n}$. Note, however, that (68) can be shown in a manner very similar to the proof of part (b) above, so to avoid redundancy we omit the proof here.

Since the proof of part (d) is also similar to part (b), we again omit it to avoid being repetitive.

To show part (e), we again proceed by showing the mean square convergence of the g^{th} element of $\frac{C_n' Z_n' Q_{X_n} u_n}{b_n r_n}$ to zero, noting, of course, that the matrix Q_{X_n} is well-defined almost surely for n sufficiently large under Assumption 2(b), as argued previously. To proceed, write the g^{th} element of $\frac{C_n' Z_n' Q_{X_n} u_n}{b_n r_n}$ as

$$\frac{e_g' C_n' Z_n' Q_{X_n} u_n}{b_n r_n}, \tag{69}$$

where again e_g denotes that g^{th} column of the $G \times G$ identity matrix I_G. Calculating the first two moments shows that

$$E\left[\frac{e_g' C_n' Z_n' Q_{X_n} u_n}{b_n r_n}\right] = E_{\overline{Z}_n}\left[\frac{e_g' C_n' Z_n' Q_{X_n} E\left(u \mid \overline{Z}_n\right)}{b_n r_n}\right] = 0 \tag{70}$$

by Assumptions 3, and that

$$\begin{aligned} E\left[\frac{e_g' C_n' Z_n' Q_{X_n} u_n}{b_n r_n}\right]^2 &= E_{\overline{Z}_n}\left[\frac{e_g' C_n' Z_n' Q_{X_n} E\left(uu' \mid \overline{Z}_n\right) Q_{X_n} Z_n C_n e_g}{b_n^2 r_n^2}\right] \\ &= \left(\frac{\sigma_{uu}}{r_n}\right) E_{\overline{Z}_n}\left[\frac{e_g' C_n' Z_n' Q_{X_n} Z_n C_n e_g}{b_n^2 r_n}\right], \end{aligned} \tag{71}$$

Note that the expectation $E_{\overline{Z}_n}(\cdot)$ in equation (71) exists for n sufficiently large as a consequence of the proof of part (a) of this lemma, where we show that, under assumption 2, there exists positive integer N and constant $D^* < \infty$ such that for all $n \geq N$,

$$\frac{e_g' C_n' Z_n' Q_{X_n} Z_n C_n e_g}{b_n^2 r_n} < D^* \tag{72}$$

with probability one. It follows from (72) that for all $n \geq N$,

$$E\left[\frac{e_g' C_n' Z_n' Q_{X_n} u_n}{b_n r_n}\right]^2 = \left(\frac{\sigma_{uu}}{r_n}\right)\left\{E_{\overline{Z}_n}\left[\frac{e_g' C_n' Z_n' Q_{X_n} Z_n C_n e_g}{b_n^2 r_n}\right]\right\} \leq \left(\frac{\sigma_{uu}}{r_n}\right) D^* = O(r_n^{-1}). \tag{73}$$

The desired result, thus, follows immediately given (70) and (73).

Part (f) of this lemma can be shown by showing the convergence to zero of the $(g,h)^{th}$ element of $\frac{C_n' Z_n' Q_{X_n} V_n}{b_n r_n}$ in the mean square sense. Since the proof is very similar to that for part (e) above, we omit writing this proof to avoid redundancy.

To show part (g), we again note that a sufficient argument would be to show the convergence (in probability) to zero of the $(g,h)^{th}$ element of $\frac{V_n' M_n Z_n C_n}{b_n r_n}$, where the matrix M_n is well-defined almost surely for n sufficiently large as previously argued. To begin, write the $(g,h)^{th}$ element as

$$\frac{e_g' V_n' M_n Z_n C_n e_h}{b_n r_n} = \frac{V_n^{(g)'}\left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h}{b_n r_n} - \frac{V_n^{(g)'} Q_{\overline{Z}_n} \Xi_n Q_{X_n} Z_n C_n e_h}{b_n r_n}. \tag{74}$$

We will show that both terms on the right-hand side of (74) converge in probability to zero. Starting with the first term, note that

$$\begin{aligned} & E\left(\frac{V_n^{(g)'}\left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h}{b_n r_n}\right) \\ = \;& E_{\overline{Z}_n}\left(\frac{E\left(V_n^{(g)'} \mid \overline{Z}_n\right)\left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h}{b_n r_n}\right) = 0 \end{aligned} \tag{75}$$

since $E\left(V_n^{(g)'} \mid \overline{Z}_n\right) = 0'$ by Assumptions 3. Note further that

$$\begin{aligned}
& E\left(\frac{V_n^{(g)'}\left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h}{b_n r_n}\right)^2 \\
= \; & b_n^{-2} r_n^{-2} E_{\overline{Z}_n}\left(e_h' C_n' Z_n'\left[P_{\overline{Z}_n} - P_{X_n} - Q_{X_n}\overline{\Omega}_n Q_{\overline{Z}_n}\right] E\left(V_n^{(g)} V_n^{(g)'} \mid \overline{Z}_n\right)\left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h\right) \\
= \; & \left(\frac{\Sigma_{VV}^{(g,g)}}{b_n^2 r_n^2}\right) E_{\overline{Z}_n}\left(e_h' C_n' Z_n'\left[P_{\overline{Z}_n} - P_{X_n} + Q_{X_n}\overline{\Omega}_n Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h\right). \qquad (76)
\end{aligned}$$

Now, with probability one

$$\begin{aligned}
& e_h' C_n' Z_n'\left[P_{\overline{Z}_n} - P_{X_n} + Q_{X_n}\overline{\Omega}_n Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h \\
\leq \; & e_h' C_n' Z_n' Q_{X_n} Z_n C_n e_h + e_h' C_n' Z_n' Q_{X_n}\overline{\Omega}_n^2 Q_{X_n} Z_n C_n e_h \\
\leq \; & e_h' C_n' Z_n' Q_{X_n} Z_n C_n e_h + (\overline{l_n})^2 e_h' C_n' Z_n' Q_{X_n} Z_n C_n e_h \qquad (77)
\end{aligned}$$

where $\overline{l_n} = \sup_{1 \leq i \leq n} |\overline{\omega}_{i,n}|$ as defined in Assumption 6(a). Moreover, there exists a positive integer N and a constant $D^* < \infty$ such that for all $n \geq N$

$$\begin{aligned}
& \left(\frac{\Sigma_{VV}^{(g,g)}}{b_n^2 r_n^2}\right) E_{\overline{Z}_n}\left(e_h' C_n' Z_n'\left[P_{\overline{Z}_n} - P_{X_n} + Q_{X_n}\overline{\Omega}_n Q_{\overline{Z}_n}\overline{\Omega}_n Q_{X_n}\right] Z_n C_n e_h\right) \\
\leq \; & \left(\frac{\Sigma_{VV}^{(g,g)}}{r_n}\right) E_{\overline{Z}_n}\left(\left[\frac{e_h' C_n' Z_n' Q_{X_n} Z_n C_n e_h}{b_n^2 r_n}\right] + (\overline{l_n})^2 E_{\overline{Z}_n}\left[\frac{e_h' C_n' Z_n' Q_{X_n} Z_n C_n e_h}{b_n^2 r_n}\right]\right) \\
< \; & \left(\frac{\Sigma_{VV}^{(g,g)}}{r_n}\right)\left[1 + (\overline{l_n})^2\right] D^* = O\left(r_n^{-1}\right), \qquad (78)
\end{aligned}$$

under assumption 6(a), where the last inequality above follows from expression (72). Expressions (75) and (78) and the Chebyshev's inequality then directly imply that the first term on the right-hand side of (74) converges in probability to zero. Next, to show that the second term vanishes as well, write

$$\begin{aligned}
\left|\frac{V_n^{(g)'} Q_{\overline{Z}_n}\Xi_n Q_{X_n} Z_n C_n e_h}{b_n r_n}\right| & \leq \sqrt{\frac{V_n^{(g)'} Q_{\overline{Z}_n} V_n^{(g)}}{b_n r_n}} \sqrt{\frac{e_h' C_n' Z_n' Q_{X_n}\Xi_n^2 Q_{X_n} Z_n C_n e_h}{b_n r_n}} \\
& \leq \left[\sqrt{\frac{n}{r_n}} \sup_i |\xi_{i,n}|\right] \sqrt{\frac{V_n^{(g)'} Q_{\overline{Z}_n} V_n^{(g)}}{n}} \sqrt{\frac{e_h' C_n' Z_n' Q_X Z_n C_n e_h}{b_n^2 r_n}}, \qquad (79)
\end{aligned}$$

where the first inequality above follows from Cauchy-Schwarz. Part (e) of lemma A2 given below, we obtain

$$\frac{V_n^{(g)'} Q_{\overline{Z}_n} V_n^{(g)}}{n} \xrightarrow{p} \Sigma_{VV}^{(g,g)}(1 - \alpha) < \infty. \qquad (80)$$

Moreover, from part (a) of this lemma, we deduce that

$$\frac{e_h' C_n' Z_n' Q_{X_n} Z_n C_n e_h}{b_n^2 r_n} = O_{a.s.}(1). \qquad (81)$$

Furthermore, Assumptions 5 and 6(d) imply that

$$\sqrt{\frac{n}{r_n}} \sup_i |\xi_{i,n}| \xrightarrow{p} 0. \qquad (82)$$

It, thus, follows from (79), (80), (81), and (82) that

$$\frac{V_n^{(g)\prime} Q_{\overline{Z}_n} \Xi_n Q_{X_n} Z_n C_n e_h}{b_n r_n} \xrightarrow{p} 0 \tag{83}$$

under the assumptions of the lemma. The desired result then follows immediately. □

Lemma A2: *Under Assumptions 2-5, suppose that $r_n \to \infty$ as $n \to \infty$ such that $\frac{\sqrt{K_n}}{r_n} \to 0$. Then, the following statements are true as $n \to \infty$.*

(a) $\frac{V_n' M_n^* u_n}{r_n} \xrightarrow{p} 0$;

(b) $\frac{V_n' M_n^* V_n}{r_n} \xrightarrow{p} 0$;

(c) $\frac{u_n' M_n^* u_n}{r_n} \xrightarrow{p} 0$, *where* $M_n^* = \left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - \left(\frac{K_n}{n-K_n-J}\right) Q_{\overline{Z}_n}\right]$;

(d) $\frac{V_n' Q_{\overline{Z}_n} u_n}{n} \xrightarrow{p} \sigma_{Vu}\left(1-\alpha\right)$;

(e) $\frac{V_n' Q_{\overline{Z}_n} V_n}{n} \xrightarrow{p} \Sigma_{VV}\left(1-\alpha\right)$;

(f) $\frac{u_n' Q_{\overline{Z}_n} u_n}{n} \xrightarrow{p} \sigma_{uu}\left(1-\alpha\right)$.

Proof of Lemma A2: To begin, we first note that Assumption 2 implies that, for n sufficiently large, the projection matrices $P_{\overline{Z}_n}$, P_{X_n}, $Q_{\overline{Z}_n}$, and Q_{X_n} are all well-defined with probability one, so that the matrix $M_n = \left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - \left(\frac{K_n}{n-K_n-J}\right) Q_{\overline{Z}_n}\right]$ is well-defined with probability one as well. Now, to show part (a) of the lemma, note first that M_n^* is a special case of the matrix $\overline{M}_n = \left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - Q_{\overline{Z}_n} \overline{\Omega}_n Q_{X_n}\right]$, where we take $\overline{\omega}_{i,n} = \left(\frac{K_n}{n-K_n-J}\right)$ for all i. Hence, given the proof of part (b) of Lemma A1 above, all we need in order to establish part (b) of this lemma is to show that $\left(\frac{K_n}{n-K_n-J}\right)$ satisfies conditions (a)-(c) of Assumption 6. To proceed, observe that, in this case,

$$\bar{l}_n = \sup_{1 \le i \le n} \left|\overline{\omega}_{i,n}\right| = \frac{K_n}{n-K_n-J}, \tag{84}$$

from which it follows that since $0 \le \alpha < 1$,

$$\overline{\lim_{n\to\infty}}\, \bar{l}_n = \lim_{n\to\infty} \bar{l}_n = \frac{\alpha}{1-\alpha} < \infty, \tag{85}$$

so that Assumption 6(a) is satisfied. Moreover, observe that in this case

$$\sum_{i=1}^n \overline{\omega}_{i,n}\left(1-h_{i,n}\right) = \left(\frac{K_n}{n-K_n-J}\right)\sum_{i=1}^n \left(1-h_{i,n}\right) = \left(\frac{K_n}{n-K_n-J}\right)\left(n-K_n-J\right) = K_n, \tag{86}$$

so that Assumption 6(b) is satisfied. Finally, observe that in this case

$$\sum_{i=1}^n E\left(\overline{\omega}_{i,n}^2\right) = K_n \left\{\frac{K_n n}{\left(n-K_n-J\right)^2}\right\} = O\left(K_n\right), \tag{87}$$

since we assume that $K_n = O(n)$, so that assumption 6(c) is satisfied as well. The desired result then follows directly given the proof of Lemma A1 part (b).

The proofs of parts (b) and (c) of this lemma follow in a similar manner from the proofs of parts (c) and (d) of Lemma A1 and, hence, are omitted here.

To show part (d), it suffices to show that, under the assumptions of the lemma, the g^{th} element of $\frac{V_n' Q_{\overline{Z}_n} u_n}{K_n}$ converges in probability to the g^{th} element of the vector $\sigma_{Vu}(1-\alpha)$, i.e.

$$\frac{V_n^{(g)'} Q_{\overline{Z}_n} u_n}{n} \xrightarrow{p} \sigma_{Vu}^g (1-\alpha). \tag{88}$$

To proceed, note first that, for n sufficiently large, so that $(\overline{Z}_n' \overline{Z}_n)^{-1}$ is well-defined almost surely, we can take the expectation of the left-hand sider of (88) to obtain:

$$\begin{aligned} E\left(\frac{V_n^{(g)'} Q_{\overline{Z}_n} u_n}{n}\right) &= E_{\overline{Z}_n}\left(Tr\left\{\frac{Q_{\overline{Z}_n}\left[E_{\eta|\overline{Z}_n}\left(u_n V_n^{(g)'}\right)\right]}{n}\right\}\right) = \frac{\sigma_{Vu}^g}{n} E_{\overline{Z}_n}\left(Tr\left(Q_{\overline{Z}_n}\right)\right) \\ &= \sigma_{Vu}^g\left(\frac{n-K_n-J}{n}\right) \to \sigma_{Vu}^g(1-\alpha), \quad \text{as } n \to \infty. \end{aligned} \tag{89}$$

Next, let $q_{ij,n}$ denote the $(i,j)^{th}$ element of the matrix $Q_{\overline{Z}_n}$, and we can calculate the second moment of $\frac{V_n^{(g)'} Q_{\overline{Z}_n} u_n}{n}$, for n sufficiently large, as follows:

$$\begin{aligned} var\left[\frac{V_n^{(g)'} Q_{\overline{Z}_n} u_n}{n}\right] &= \frac{1}{n^2} E\left[V_n^{(g)'} Q_{\overline{Z}_n} u_n\right]^2 - \frac{1}{n^2}\left(E\left[V_n^{(g)'} Q_{\overline{Z}_n} u_n\right]\right)^2 \\ &= \left(\frac{1}{n^2}\right) E_{\overline{Z}_n}\left\{\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n q_{ij,n}\, q_{kl,n}\, E_{\eta|\overline{Z}_n}\left[v_{ig} u_j v_{kg} u_l\right]\right\} - \left(\sigma_{Vu}^g (n-K_n-J)\right)^2 \\ &= \left(\frac{1}{n^2}\right) E(v_{ig}^2 u_i^2) E_{\overline{Z}_n}\left[\sum_{i=1}^n q_{ii,n}^2\right] + \left(\frac{1}{n^2}\right) \sigma_{VV}^{(g,g)} \sigma_{uu} E_{\overline{Z}_n}\left[\sum_{i=2}^n \sum_{j=1}^{i-1} q_{ij,n}^2 + \sum_{j=2}^n \sum_{i=1}^{j-1} q_{ij,n}^2\right] \\ &\quad +2\left(\frac{1}{n^2}\right)\left(\sigma_{Vu}^g\right)^2 E_{\overline{Z}_n}\left[\sum_{i=2}^n \sum_{j=1}^{i-1} q_{ii,n}\, q_{jj,n} + \sum_{i=2}^n \sum_{j=1}^{i-1} q_{ij,n}\, q_{ji,n} - \left(\frac{1}{2}\right)(n-K_n-J)^2\right] \\ &= \mathcal{D}_n + \mathcal{E}_n + \mathcal{F}_n, \text{ say,} \end{aligned} \tag{90}$$

where the second equality above follows by noting that the expectation $E_{\eta|\overline{Z}_n}[v_{g,i} u_j v_{g,k} u_l]$ equals zero except in the cases where either $(i=j=k=l)$ or $(i=j$ and $k=l)$ or $(i=k$ and $j=l)$ or $(i=l$ and $j=k)$. Dealing with the term $\mathcal{D}_n$ first, observe that with probability one

$$\begin{aligned} \left(\frac{1}{n^2}\right) E(v_{ig}^2 u_i^2)\left[\sum_{i=1}^n q_{ii,n}^2\right] &\leq \left(\frac{1}{n^2}\right) E(v_{ig}^2 u_i^2) Tr\left[Q_{\overline{Z}_n}' Q_{\overline{Z}_n}\right] \\ &= \frac{E(v_{ig}^2 u_i^2)}{n^2}(n-K_n-J). \end{aligned} \tag{91}$$

Moreover, note that since the bound given by (91) holds with probability one, we deduce that

$$\mathcal{D}_n = \left(\frac{1}{n^2}\right) E(v_{ig}^2 u_i^2) E_{\overline{Z}_n}\left[\sum_{i=1}^n q_{ii,n}^2\right] = O\left(\frac{1}{n}\right); \tag{92}$$

hence, $\mathcal{D}_n \to 0$ as $n, K_n \to \infty$.

Turning our attention next to the term $\mathcal{E}_n$, we see that with probability one

$$\begin{aligned}
&\left(\frac{1}{n^2}\right)\Sigma_{VV}^{(g,g)}\sigma_{uu}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1}q_{ij,n}^2\right] \\
\leq &\left(\frac{1}{n^2}\right)\Sigma_{VV}^{(g,g)}\sigma_{uu}Tr\left[Q_{\overline{Z}_n}'Q_{\overline{Z}_n}\right] \\
= &\left(\frac{1}{n^2}\right)\Sigma_{VV}^{(g,g)}\sigma_{uu}\left(n-K_n-J\right) \qquad (93)
\end{aligned}$$

It follows again that since the bound given in (93) holds with probability one, we deduce that

$$\mathcal{E}_n = \left(\frac{1}{n^2}\right)\Sigma_{VV}^{(g,g)}\sigma_{uu}E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1}q_{ij,n}^2\right] = O\left(\frac{1}{n}\right); \qquad (94)$$

so that $\mathcal{E}_n \to 0$ as $n,\ K_n \to \infty$.

Finally, turning to the term $\mathcal{F}_n$, we note that

$$\begin{aligned}
&2\left(\frac{1}{n^2}\right)\left(\sigma_{Vu}^g\right)^2\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ii,n}\,q_{jj,n} + \sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ij,n}\,q_{ji,n} - \frac{1}{2}\left(n-K_n-J\right)^2\right] \\
= &\left(\frac{\sigma_{Vu}^g}{n}\right)^2\left\{Tr\left[Q_{\overline{Z}_n}^2\right] + \left(Tr\left[Q_{\overline{Z}_n}\right]\right)^2 - \left(n-K_n-J\right)^2 - 2\sum_{i=1}^{n}q_{ii}^2\right\} \\
= &\left(\frac{\sigma_{Vu}^g}{n}\right)^2\left\{n-K_n-J-2\sum_{i=1}^{n}q_{ii}^2\right\} \qquad (95)
\end{aligned}$$

so that with probability one

$$\begin{aligned}
&2\left(\frac{\sigma_{Vu}^g}{n}\right)^2\left|\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ii,n}\,q_{jj,n} + \sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ij,n}\,q_{ji,n} - \frac{1}{2}\left(n-K_n-J\right)^2\right]\right| \\
\leq &\left(\frac{\sigma_{Vu}^g}{n}\right)^2\left|n-K_n-J\right| + 2\left(\frac{\sigma_{Vu}^g}{n}\right)^2\left|Tr\left[Q_{\overline{Z}_n}^2\right]\right| = \frac{3\left(\sigma_{Vu}^g\right)^2}{n}\left|n-K_n-J\right| \qquad (96)
\end{aligned}$$

Since the upper bound given in (96) above holds almost surely, it follows that

$$\left|\mathcal{F}_n\right| = 2\left(\frac{\sigma_{Vu}^g}{n}\right)^2\left|E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ii,n}\,q_{jj,n} + \sum_{i=2}^{n}\sum_{j=1}^{i-1}q_{ij,n}\,q_{ji,n} - \frac{1}{2}\left(n-K_n-J\right)^2\right]\right| = O\left(n^{-1}\right), \qquad (97)$$

so that $\mathcal{F}_n \to 0$ as $n,\ K_n \to \infty$. It follows immediately from (92), (94), and (97) that

$$var\left[\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}u_n}{n}\right] \to 0 \qquad (98)$$

as $n,\ K_n \to \infty$. Moreover, on the basis of (89) and (98), we deduce part (d) of the lemma A2 as a direct consequence of Chebyshev's inequality.

Parts (e) and (f) of this lemma can be shown in a manner similar to part (d) above; hence, for brevity, we omit their proofs. □

Lemma A3: *Under Assumptions 2-4, the following statements are true as* $n \to \infty$.

(a) $\frac{V_n'(P_{\overline{Z}_n}-P_{X_n})u_n}{K_n} \xrightarrow{p} \sigma_{Vu}$;

(b) $\frac{V_n'(P_{\overline{Z}_n}-P_{X_n})V_n}{K_n} \xrightarrow{p} \Sigma_{VV}$.

Proof of Lemma A3: Before proceeding, we note that as argued in the proof of lemma 1, Assumption 2 of the paper imply that the projection matrices $P_{\overline{Z}_n}$ and P_{X_n} are well-defined with probability one for n sufficiently large, so that the difference $P_{\overline{Z}_n} - P_{X_n}$ is also well-defined with probability one for large enough n. Now, to show part (a), it suffices to show that, under the assumptions of the lemma, the g^{th} element of $\frac{V_n'(P_{\overline{Z}_n}-P_{X_n})u_n}{K_n}$ converges in probability to the g^{th} element of σ_{Vu}, i.e.

$$\frac{V_n^{(g)'}\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n} \xrightarrow{p} \sigma_{Vu}^g. \tag{99}$$

To proceed, note first that, for n sufficiently large, we can take the expectation of the left-hand side of (99) to obtain:

$$\begin{aligned} E\left(\frac{V_n^{(g)'}\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n}\right) &= E_{\overline{Z}_n}\left(Tr\left\{\frac{\left(P_{\overline{Z}_n} - P_{X_n}\right)\left[E_{\eta|\overline{Z}_n}\left(u_n V_n^{(g)'}\right)\right]}{K_n}\right\}\right) \\ &= \frac{\sigma_{Vu}^g}{K_n} E_{\overline{Z}_n}\left(Tr\left(P_{\overline{Z}_n} - P_{X_n}\right)\right) = \left(\frac{\sigma_{Vu}^g}{K_n}\right) K_n = \sigma_{Vu}^g. \end{aligned} \tag{100}$$

Next, let $p_{ij,n}$ denote the $(i,j)^{th}$ element of the matrix $\left(P_{\overline{Z}_n} - P_{X_n}\right)$, and we can calculate the second moment of $\frac{V_n^{(g)'}(P_{\overline{Z}_n}-P_{X_n})u_n}{K_n}$, for n sufficiently large, as follows:

$$\begin{aligned} var\left[\frac{V_n^{(g)'}\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n}\right] &= \frac{1}{K_n^2} E\left[V_n^{(g)'}\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n\right]^2 - \frac{1}{K_n^2}\left(E\left[V_n^{(g)'}\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n\right]\right)^2 \\ &= \left(\frac{1}{K_n^2}\right) E_{\overline{Z}_n}\left\{\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n p_{ij,n}\, p_{kl,n}\, E_{\eta|\overline{Z}_n}\left[v_{ig}u_j v_{kg}u_l\right]\right\} - \left(\sigma_{Vu}^g\right)^2 \\ &= \left(\frac{1}{K_n^2}\right) E(v_{ig}^2 u_i^2) E_{\overline{Z}_n}\left[\sum_{i=1}^n p_{ii,n}^2\right] \\ &\quad + \left(\frac{1}{K_n^2}\right)\Sigma_{VV}^{(g,g)}\sigma_{uu} E_{\overline{Z}_n}\left[\sum_{i=2}^n\sum_{j=1}^{i-1} p_{ij,n}^2 + \sum_{j=2}^n\sum_{i=1}^{j-1} p_{ij,n}^2\right] \\ &\quad + 2\left(\frac{1}{K_n^2}\right)\left(\sigma_{Vu}^g\right)^2 E_{\overline{Z}_n}\left[\sum_{i=2}^n\sum_{j=1}^{i-1} p_{ii,n}\, p_{jj,n} + \sum_{i=2}^n\sum_{j=1}^{i-1} p_{ij,n}\, p_{ji,n} - \left(\frac{1}{2}\right)K_n^2\right] \\ &= \mathcal{G}_n + \mathcal{H}_n + \mathcal{I}_n, \text{ say,} \end{aligned} \tag{101}$$

where the second equality above follows by noting that the expectation $E_{\eta|\overline{Z}_n}\left[v_{g,i}u_j v_{g,k}u_l\right]$ equals zero except in the cases where either $(i = j = k = l)$ or $(i = j$ and $k = l)$ or $(i = k$ and $j = l)$ or $(i = l$ and $j = k)$. Dealing with the term $\mathcal{G}_n$ first, observe that with probability one

$$\begin{aligned} \left(\frac{1}{K_n^2}\right) E(v_{ig}^2 u_i^2)\left[\sum_{i=1}^n p_{ii,n}^2\right] &\leq \left(\frac{1}{K_n^2}\right) E(v_{ig}^2 u_i^2) Tr\left[\left(P_{\overline{Z}_n} - P_{X_n}\right)'\left(P_{\overline{Z}_n} - P_{X_n}\right)\right] \\ &= \left(\frac{1}{K_n^2}\right) E(v_{ig}^2 u_i^2) Tr\left[\left(P_{\overline{Z}_n} - P_{X_n}\right)\right] = \frac{E(v_{ig}^2 u_i^2)}{K_n}. \end{aligned} \tag{102}$$

Moreover, note that since the bound given by (102) holds with probability one, we deduce that

$$\mathcal{G}_n = \left(\frac{1}{K_n^2}\right) E(v_{ig}^2 u_i^2) E_{\overline{Z}_n}\left[\sum_{i=1}^{n} p_{ii,n}^2\right] = O\left(\frac{1}{K_n}\right); \tag{103}$$

hence, $\mathcal{G}_n \to 0$ as $n, K_n \to \infty$.

Turning our attention next to the term $\mathcal{H}_n$, we note that in this case

$$\begin{aligned} & \left(\frac{1}{K_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu} \left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1} p_{ij,n}^2\right] \\ \leq & \left(\frac{1}{K_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu} Tr\left[\left(P_{\overline{Z}_n} - P_{X_n}\right)'\left(P_{\overline{Z}_n} - P_{X_n}\right)\right] = \left(\frac{1}{K_n}\right) \sigma_{VV}^{(g,g)} \sigma_{uu} \end{aligned} \tag{104}$$

with probability one. Since the bound given in (104) holds with probability one, we deduce that

$$\mathcal{H}_n = \left(\frac{1}{K_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu} E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1} p_{ij,n}^2\right] = O\left(\frac{1}{K_n}\right); \tag{105}$$

so that $\mathcal{H}_n \to 0$ as n, $K_n \to \infty$.

Finally, turning to the term $\mathcal{I}_n$, we note that

$$\begin{aligned} & 2\left(\frac{1}{K_n^2}\right)\left(\sigma_{Vu}^g\right)^2 \left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ii,n}\, p_{jj,n} + \sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ij,n}\, p_{ji,n} - \frac{1}{2}K_n^2\right] \\ = & \left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 \left\{Tr\left[\left(P_{\overline{Z}_n} - P_{X_n}\right)^2\right] + \left(Tr\left[\left(P_{\overline{Z}_n} - P_{X_n}\right)\right]\right)^2 - K_n^2 - 2\sum_{i=1}^{n} p_{ii}^2\right\} \\ = & \left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 \left\{K_n - 2\sum_{i=1}^{n} p_{ii}^2\right\} \end{aligned} \tag{106}$$

so that with probability one

$$\begin{aligned} & 2\left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 \left|\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ii,n}\, p_{jj,n} + \sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ij,n}\, p_{ji,n} - \frac{1}{2}K_n^2\right]\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 \left|Tr\left[\left(P_{\overline{Z}_n} - P_{X_n}\right)\right]\right| + 2\left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 \left|\sum_{i=1}^{n} p_{ii}^2\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 K_n + 2\left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 \left|Tr\left[\left(P_{\overline{Z}_n} - P_{X_n}\right)'\left(P_{\overline{Z}_n} - P_{X_n}\right)\right]\right| = \frac{3\left(\sigma_{Vu}^g\right)^2}{K_n} \end{aligned} \tag{107}$$

Since the upper bound given in (107) above holds almost surely, it follows that

$$|\mathcal{I}_n| = 2\left(\frac{\sigma_{Vu}^g}{K_n}\right)^2 \left|E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ii,n}\, p_{jj,n} + \sum_{i=2}^{n}\sum_{j=1}^{i-1} p_{ij,n}\, p_{ji,n} - \frac{1}{2}K_n^2\right]\right| = O\left(K_n^{-1}\right), \tag{108}$$

so that $\mathcal{I}_n \to 0$ as n, $K_n \to \infty$. It follows immediately from (103), (105), and (108) that

$$var\left[\frac{V_n^{(g)'}\left(P_{\overline{Z}_n} - P_{X_n}\right) u_n}{K_n}\right] \to 0 \tag{109}$$

as n, $K_n \to \infty$. Moreover, on the basis of (100) and (109), we deduce part (a) of the Lemma A3 as a direct consequence of Chebyshev's inequality.

To show part (b) of the lemma, note that, similar to part (a) above, it is sufficient to simply show that any arbitrary element of $\frac{V_n'(P_{\overline{Z}_n}-P_{X_n})V_n}{K_n}$, say the $(g,h)^{th}$ element, converges in probability to the $(g,h)^{th}$ element of Σ_{VV}, i.e., it is sufficient to show that

$$\frac{V_n^{(g)\prime}\left(P_{\overline{Z}_n}-P_{X_n}\right)V_n^{(h)}}{K_n} \xrightarrow{p} \Sigma_{VV}^{(g,h)} \quad \text{as } n\to\infty. \tag{110}$$

Note, however, that (110) can be shown in a manner very similar to the proof of part (a) above, so to avoid redundancy we omit the proof here. □

Proof of Theorem 3.1: To proceed, note that for n sufficiently large so that $M_n=\left[\left(P_{\overline{Z}_n}-P_{X_n}\right)-Q_{\overline{Z}_n}\widetilde{\Omega}_n Q_{X_n}\right]$ is well-defined with probability one, and we can write

$$\frac{Y_{2n}'M_nY_{2n}}{r_n}=\frac{C_n'Z_n'Q_{X_n}Z_nC_n}{b_n^2r_n}+\frac{C_n'Z_n'Q_{X_n}V_n}{b_nr_n}+\frac{V_n'M_nZ_nC_n}{b_nr_n}+\frac{V_n'M_nV_n}{r_n}, \tag{111}$$

Now, it follows from parts (a), (c), (f), and (g) of lemma A1 that

$$\frac{Y_{2n}'M_nY_{2n}}{r_n}=\Psi_n+op(1), \tag{112}$$

where $\Psi_n=b_n^{-2}r_n^{-1}C_n'Z_n'Q_{X_n}Z_nC_n$ is positive definite almost surely for n sufficiently large given the result of lemma A1 part (a). Moreover, for n sufficiently large, we can write

$$\frac{Y_2'M_nu_n}{r_n}=\frac{C_n'Z_n'Q_{X_n}u_n}{b_nr_n}+\frac{V_n'M_nu_n}{r_n}, \tag{113}$$

so that

$$\frac{Y_{2n}'M_nu_n}{r_n}\xrightarrow{p}0,\ \text{as } n\to\infty, \tag{114}$$

given parts (b) and (e) of lemma A1. Next, note that we can write

$$\widehat{\beta}_{\omega,n}-\beta_0=\left(\left[\frac{Y_{2n}'M_nY_{2n}}{r_n}\right]^+\left[\frac{Y_{2n}'M_nY_{2n}}{r_n}\right]-I_G\right)\beta_0+\left[\frac{Y_{2n}'M_nY_{2n}}{r_n}\right]^+\left[\frac{Y_{2n}'M_nu_n}{r_n}\right]. \tag{115}$$

In view of (112) and (114), it follows by Proposition 2.30 of White (1999) and the Slutsky's theorem that

$$\left[\frac{Y_{2n}'M_nY_{2n}}{r_n}\right]^+\left[\frac{Y_{2n}'M_nY_{2n}}{r_n}\right]-I_G\xrightarrow{p}0 \tag{116}$$

and that

$$\left[\frac{Y_{2n}'M_nY_{2n}}{r_n}\right]^+\left[\frac{Y_{2n}'M_nu_n}{r_n}\right]\xrightarrow{p}0, \tag{117}$$

from which we deduce that $\widehat{\beta}_{\omega,n}\xrightarrow{p}\beta_0$ as required. □

Proof of Theorem 3.3 To proceed, note first that $\widehat{\lambda}_{LIML,n}$ is smallest root of the determinantal equation

$$\det\left\{\begin{pmatrix} y_{1n}'Q_{X_n}y_{1n} & y_{1n}'Q_{X_n}Y_{2n}\\ Y_{2n}'Q_{X_n}y_{1n} & Y_{2n}'Q_{X_n}Y_{2n}\end{pmatrix}-\lambda_n\begin{pmatrix} y_{1n}'Q_{\overline{Z}_n}y_{1n} & y_{1n}'Q_{\overline{Z}_n}Y_{2n}\\ Y_{2n}'Q_{\overline{Z}_n}y_{1n} & Y_{2n}'Q_{\overline{Z}_n}Y_{2n}\end{pmatrix}\right\}=0 \tag{118}$$

or, in more succinct notation,

$$\det\left\{Y_n' Q_{X_n} Y_n - \lambda_n Y_n' Q_{\overline{Z}_n} Y_n\right\} = 0, \tag{119}$$

where $Y_n = [y_{1n}, Y_{2n}]$ and where the elements of the determinantal equation given above are all well-defined with probability one for n sufficiently large, as a consequence of Assumption 2. Now, define $\Upsilon = \begin{pmatrix} 1 & 0 \\ -\beta_0 & I_G \end{pmatrix}$ and note that the smallest root of equation (119) is the same as the smallest root of the equation

$$\det\left\{\Upsilon' Y_n' Q_{X_n} Y_n \Upsilon - \lambda_n \Upsilon' Y_n' Q_{\overline{Z}_n} Y_n \Upsilon\right\} = 0, \tag{120}$$

where

$$\begin{aligned} \Upsilon' Y_n' Q_{X_n} Y_n \Upsilon &= \begin{pmatrix} 1 & -\beta_0 \\ 0 & I_G \end{pmatrix} \begin{pmatrix} y_{1n}' Q_{X_n} y_{1n} & y_{1n}' Q_{X_n} Y_{2n} \\ Y_{2n}' Q_{X_n} y_{1n} & Y_{2n}' Q_{X_n} Y_{2n} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\beta_0 & I_G \end{pmatrix} \\ &= \begin{pmatrix} u_n' Q_{X_n} u_n & u_n' Q_{X_n} Y_{2n} \\ Y_{2n}' Q_{X_n} u_n & Y_{2n}' Q_{X_n} Y_{2n} \end{pmatrix}. \end{aligned} \tag{121}$$

and

$$\Upsilon' Y_n' Q_{\overline{Z}_n} Y_n \Upsilon = \begin{pmatrix} u_n' Q_{\overline{Z}_n} u_n & u_n' Q_{\overline{Z}_n} V_n \\ V_n' Q_{\overline{Z}_n} u_n & V_n' Q_{\overline{Z}_n} V_n \end{pmatrix}. \tag{122}$$

Now, let $\lambda_n = \frac{n-J}{n-K_n-J} + \frac{\tau_n r_n}{n}$ and rewrite (120) as

$$\begin{aligned} &\det\left\{\begin{pmatrix} u_n' Q_{X_n} u_n & u_n' Q_{X_n} Y_{2n} \\ Y_{2n}' Q_{X_n} u_n & Y_{2n}' Q_{X_n} Y_{2n} \end{pmatrix} - \left(\frac{n-J}{n-K_n-J}\right) \begin{pmatrix} u_n' Q_{\overline{Z}_n} u_n & u_n' Q_{\overline{Z}_n} V_n \\ V_n' Q_{\overline{Z}_n} u_n & V_n' Q_{\overline{Z}_n} V_n \end{pmatrix}\right. \\ &\left. - \tau_n \begin{pmatrix} \frac{r_n u_n' Q_{\overline{Z}_n} u_n}{n} & \frac{r_n u_n' Q_{\overline{Z}_n} V_n}{n} \\ \frac{r_n V_n' Q_{\overline{Z}_n} u_n}{n} & \frac{r_n V_n' Q_{\overline{Z}_n} V_n}{n} \end{pmatrix}\right\} = 0, \end{aligned} \tag{123}$$

which, in turn, can be shown, by straightforward manipulation, to be equivalent to the determinantal equation

$$\begin{aligned} &\det\left\{\begin{pmatrix} u_n' M_n^* u_n & \frac{u_n' Q_{X_n} Z_n C_n}{b_n} + u_n' M_n^* V_n \\ \frac{C_n' Z_n' Q_{X_n} u_n}{b_n} + V_n' M_n^* u_n & \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2} + \frac{C_n' Z_n' Q_{X_n} V_n}{b_n} + \frac{V_n' Q_{X_n} Z_n C_n}{b_n} + V_n' M_n^* V_n \end{pmatrix}\right. \\ &\left. - \tau_n \begin{pmatrix} \frac{r_n u_n' Q_{\overline{Z}_n} u_n}{n} & \frac{r_n u_n' Q_{\overline{Z}_n} V_n}{n} \\ \frac{r_n V_n' Q_{\overline{Z}_n} u_n}{n} & \frac{r_n V_n' Q_{\overline{Z}_n} V_n}{n} \end{pmatrix}\right\} = 0, \end{aligned} \tag{124}$$

where $M_n^* = \left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - \left(\frac{K_n}{n-K_n-J}\right) Q_{\overline{Z}_n}\right]$. Moreover, it is apparent that $\widehat{\lambda}_{LIML,n}$, the smallest root of equation (118), is related to $\widehat{\tau}_{LIML,n}$, the smallest root of (124), by the equation

$$\widehat{\lambda}_{LIML,n} = \frac{n-J}{n-K_n-J} + \frac{\widehat{\tau}_{LIML,n} r_n}{n}. \tag{125}$$

Furthermore, note that $\widehat{\tau}_{LIML,n}$ is also the smallest root of the determinantal equation

$$\begin{aligned} &\det\left\{\begin{pmatrix} \frac{u_n' M_n^* u_n}{r_n} & \frac{u_n' Q_{X_n} Z_n C_n}{b_n r_n} + \frac{u_n' M_n^* V_n}{r_n} \\ \frac{C_n' Z_n' Q_{X_n} u_n}{b_n r_n} + \frac{V_n' M_n^* u_n}{r_n} & \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2 r_n} + \frac{C_n' Z_n' Q_{X_n} V_n}{b_n r_n} + \frac{V_n' Q_{X_n} Z_n C_n}{b_n r_n} + \frac{V_n' M_n^* V_n}{r_n} \end{pmatrix}\right. \\ &\left. - \tau_n \begin{pmatrix} \frac{u_n' Q_{\overline{Z}_n} u_n}{n} & \frac{u_n' Q_{\overline{Z}_n} V_n}{n} \\ \frac{V_n' Q_{\overline{Z}_n} u_n}{n} & \frac{V_n' Q_{\overline{Z}_n} V_n}{n} \end{pmatrix}\right\} = 0. \end{aligned} \tag{126}$$

It then follows by parts (a), (e), and (f) of lemma A1 and parts (a)-(f) of Lemma A2 and by continuity that as $n \to \infty$, the difference between $\widehat{\tau}_{LIML,n}$ and the smallest root of

$$\det\left\{\begin{pmatrix} 0 & 0 \\ 0 & \Psi_n \end{pmatrix} - \tau_n \begin{pmatrix} \sigma_{uu}(1-\alpha) & \sigma_{Vu}'(1-\alpha) \\ \sigma_{Vu}(1-\alpha) & \Sigma_{VV}(1-\alpha) \end{pmatrix}\right\} = 0 \tag{127}$$

goes to zero in probability as $n \to \infty$. Since the smallest root of (127) is obviously zero, we deduce immediately that $\widehat{\tau}_{LIML,n} = o_p(1)$. It follows that from (125) that

$$\widehat{\lambda}_{LIML,n} = \frac{n-J}{n-K_n-J} + o_p\left(\frac{r_n}{n}\right), \tag{128}$$

as required. □

Proof of Theorem 3.4 To show part (a), note first that, for n sufficiently large, $P_{\overline{Z}_n}$ and P_{X_n} are well-defined with probability one, and we can write

$$\begin{aligned} \frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n} &= \left(\frac{r_n}{K_n}\right)\frac{C_n'Z_n'Q_{X_n}Z_nC_n}{b_n^2 r_n} + \left(\frac{r_n}{K_n}\right)\frac{C_n'Z_n'Q_{X_n}V_n}{b_n r_n} \\ &+ \left(\frac{r_n}{K_n}\right)\frac{V_n'Q_{X_n}Z_nC_n}{b_n r_n} + \frac{V_n'\left(P_{\overline{Z}_n} - P_{X_n}\right)V_n}{K_n}. \end{aligned} \tag{129}$$

Now, since it is assumed in part (a) that $\frac{r_n}{K_n} \to 0$ as $n \to \infty$, it follows from parts (a) and (f) of Lemma A1 and from part (b) of Lemma A3 that

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n} \xrightarrow{p} \Sigma_{VV}, \tag{130}$$

where Σ_{VV} is positive definite by assumption 3 and is, thus, nonsingular. Moreover, for n sufficiently large, we can write

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n} = \left(\frac{r_n}{K_n}\right)\frac{C_n'Z_n'Q_{X_n}u_n}{b_n r_n} + \frac{V_n'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n}, \tag{131}$$

so that

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n} \xrightarrow{p} \sigma_{Vu}, \text{ as } n \to \infty, \tag{132}$$

by part (e) of lemma A1 and part (a) of Lemma A3. Next, write

$$\begin{aligned} \widehat{\beta}_{2SLS,n} - \beta_0 &= \left(\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n}\right]^+ \left[\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n}\right] - I_G\right)\beta_0 \\ &+ \left[\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n}\right]^+ \left[\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n}\right], \end{aligned} \tag{133}$$

Given (130) and (132), it follows immediately by the Slutsky's theorem that

$$\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n}\right]^+ \left[\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n}\right] - I_G \xrightarrow{p} 0 \tag{134}$$

and

$$\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}\right]^+\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{K_n}\right]\xrightarrow{p}\Sigma_{VV}^{-1}\sigma_{Vu}, \tag{135}$$

so that $\widehat{\beta}_{2SLS,n}\xrightarrow{p}\beta_0+\Sigma_{VV}^{-1}\sigma_{Vu}$ as required.

To show part (b) note that since in this case $\frac{r_n}{K_n}\to\delta$, for some $\delta\in(0,\infty)$, as $n\to\infty$, it follows directly from parts (a) and (f) of Lemma A1 and from part (b) of Lemma A3 that

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}=(\delta\Psi_n+\Sigma_{VV})+o_p(1), \tag{136}$$

where $\Psi_n=b_n^{-2}r_n^{-1}C_n'Z_n'Q_{X_n}Z_nC_n$ is positive definite almost surely for n sufficiently large given the result of part (a) of Lemma A1. In addition, from part (e) of Lemma A1 and part (a) of lemma A3, we deduce that

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{K_n}\xrightarrow{p}\sigma_{Vu}. \tag{137}$$

The desired result, thus, follows directly from (136) and (137).

To show part (c), write

$$\begin{aligned}\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n} &= \frac{C_n'Z_n'Q_{X_n}Z_nC_n}{b_n^2r_n}+\frac{C_n'Z_n'Q_{X_n}V_n}{b_nr_n}\\ &\quad+\frac{V_n'Q_{X_n}Z_nC_n}{b_nr_n}+\left(\frac{K_n}{r_n}\right)\frac{V_n'\left(P_{\overline{Z}_n}-P_{X_n}\right)V_n}{K_n},\end{aligned} \tag{138}$$

where, given Assumption 2, both sides of equation (138) are well-defined with probability one for n sufficiently large. Now, since it is assumed in this part that $\frac{K_n}{r_n}\to 0$ as $n\to\infty$, it follows from parts (a) and (f) of lemma A1 and from part (b) of lemma A3 that

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}=\Psi_n+o_p(1), \tag{139}$$

where, again, we note that, given part (a) of Lemma A1, we see that Ψ_n is positive definite almost surely for n sufficiently large. Moreover, note that, for n sufficiently large, we can write

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{r_n}=\frac{C_n'Z_n'Q_{X_n}u_n}{b_nr_n}+\left(\frac{K_n}{r_n}\right)\frac{V_n'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{K_n}, \tag{140}$$

so that

$$\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{r_n}\xrightarrow{p}0,\ \text{as } n\to\infty, \tag{141}$$

given part (e) of lemma A1 and part (a) of Lemma A3 and given the assumption that $\frac{K_n}{r_n}\to 0$ as $n\to\infty$. Next, write

$$\begin{aligned}\widehat{\beta}_{2SLS,n}-\beta_0 &= \left(\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}\right]^+\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}\right]-I_G\right)\beta_0\\ &\quad+\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}\right]^+\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{r_n}\right].\end{aligned} \tag{142}$$

In view of (139) and (141), it follows immediately by the Slutsky's theorem that

$$\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}\right]^{+}\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}\right]-I_G\xrightarrow{p}0 \tag{143}$$

and

$$\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}\right]^{+}\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{r_n}\right]\xrightarrow{p}0, \tag{144}$$

from which we deduce that $\widehat{\beta}_{2SLS,n}\xrightarrow{p}\beta_0$ as required. □

6 References

Anderson, T.W. (1976). "Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simultaneous Equations in Econometrics," *The Journal of the Royal Statistical Society*, Series B, 38, 1-36.

Andrews, D.W.K. (1991). "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica,* 59, 307-346.

Angrist, J.D., G.W. Imbens, and A. Krueger (1999). "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics* 14, 57-67.

Angrist, J. D. and A. B. Krueger (1995). "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225-235.

Bekker, P.A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657-681.

Bekker, P.A. and J. van der Ploeg (1999). "Instrumental Variable Estimation Based on Group Data," Unpublished Working Paper, University of Groningen.

Blomquist, S. and M. Dahlberg (1999). "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics* 14, 69-88.

Bound, J., D.A. Jaeger and R. Baker (1995). "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90, 443-450.

Chao, J.C. and N.R. Swanson (2001). "Bias and MSE Analysis of the IV Estimator under Weak Identification with Application to Bias Correction," unpublished manuscript, University of Maryland.

Choi, I. and P.C.B. Phillips (1992). "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113-150.

Donald, S.G. and W.K.Newey (2001). "Choosing the Number of Instruments," *Econometrica*, 69, 1161-1191.

Fuller, W.A. (1977). "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939-953.

Hahn, J. (1997). "Optimal Inference with Many Instruments," Unpublished Working Paper, University of Pennsylvania.

Hahn, J., J. Hausman, and G. Kuersteiner (2001). "Higher Order MSE of Jacknife 2SLS," Unpublished Working Paper, MIT.

Hahn, J. and A. Inoue (2000). "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators," Brown University Working Paper.

Hall, A.R. and F. P. M. Peixe (2000). "A Consistent Method for the Selection of Relevant Instruments," North Carolina State University Working Paper.

Koenker, R. and J.A.F. Machaso (1999). "GMM Inference When the Number of Moment Conditions Is Large," *Journal of Econometrics*, 93, 327-344.

Kunitomo, N. (1980). "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, 75, 693-700.

Magnus, J.R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (John Wiley & Sons, New York).

Morimune, K. (1983). "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability Is Large Compared with the Sample Size," *Econometrica,* 51, 821-841.

Phillips, P.C.B. (1989). "Partially Identified Econometric Models," *Econometric Theory*, 5, 181-240.

Portnoy, S. (1984). "Asymptotic Behavior of M-estimators of p Regression Parameters When $\frac{p^2}{n}$ Is Large I: Consistency," *Annals of Statistics*, 12, 1209-1298.

Portnoy, S. (1985). "Asymptotic Behavior of M-estimators of p Regression Parameters When $\frac{p^2}{n}$ Is Large II: Normal Approximation," *Annals of Statistics*, 13, 1403-1417.

Portnoy, S. (1987). "A Central Limit Theorem Applicable to Robust Regression Estimators," *Journal of Multivariate Analysis*, 22, 24-50.

Schmidt, P. (1976). *Econometrics* (Marcel Dekker, New York).

Shea, J. (1997). "Instrumental Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics*, 79, 348-352.

Staiger, D. and J.H. Stock (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65, 557-586.

Stock, J.H. and M. Yogo (2001). "Testing for Weak Instruments in Linear IV Regression." Unpublished Working Paper, Harvard University.

Van Hasselt, M. (2000). "The Instrumental Variables Estimator and the Effects of Non-Gaussian Distributions," Unpublished Masters Thesis, University of Groningen.

Zivot, E. and J. Wang (1998). "Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments," *Econometrica*, 66, 1389-1404.

White, H. (1999). *Asymptotic Theory for Econometricians* (Academic Press, San Diego).