Projection-type Score Tests for Subsets of Parameters

Saraswata Chaudhuri

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

2008

Program Authorized to Offer Degree: Economics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Saraswata Chaudhuri

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Co-Chairs of the Supervisory Committee:

Eric Zivot

Thomas Richardson

Reading Committee:

Eric Zivot

Thomas Richardson

Richard Startz

Date: ______________________________

University of Washington

Abstract

Projection-type Score Tests for Subsets of Parameters

Saraswata Chaudhuri

Co-Chairs of the Supervisory Committee:

Professor Eric Zivot
Department of Economics

Professor Thomas Richardson
Department of Statistics

In this thesis we introduce a new projection-type score test for subsets of parameters. Although the new test is based on the projection principle, it is generally less conservative than the usual projection-type tests. In particular, we show that the new test is asymptotically equivalent to the locally optimal (usual) score test. We also show that while the (usual) score test may over-reject the true value of the parameters of interest when the nuisance parameters are not identified; the new test, by virtue of projection, can be useful in guarding against uncontrolled over-rejection. We demonstrate the practical usefulness of the new test in the context of inference on subsets of structural coefficients in linear Instrumental Variables regressions.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank my advisors Eric Zivot and Thomas Richardson for sharing with me the research questions that ultimately led to this thesis. Their mentoring and contribution extend much beyond the completion of this thesis. Academic discussions with Chang-Jin Kim, Charles R. Nelson, Michael Perlman, Jamie Robins, Elaina Rose, Richard Startz, Byron Tsang and Jon Wellner were helpful in the completion of this thesis. The good wishes of my friends and relatives made my life easier.

Most importantly, I wish to express my gratitude to my parents and my wife who chose to make the greatest sacrifices for me. The rest would have been meaningless without them.

DEDICATION

To my late grandparents
Sailendra Nath Chowdhury
Shefalika Chowdhury
Dhirendra Mohan Som
Maya Som
and my late cousin Arkaprava Dutta

Chapter 1
INTRODUCTION

In this thesis we are concerned with the problem of asymptotic inference on subsets of parameters based on the score statistic. We motivate the problem, review the literature and present a non-technical discussion of our contribution in this introductory chapter. A technical discussion along with examples of our proposed method of inference is presented in the subsequent chapters.

Almost always, only specific parameters within a larger model are the objects of primary interest. For example, a common interest in economics is to estimate the return to schooling from a wage equation of the form

$$lnwage = \alpha + \beta \times educ + \gamma \times ControlVars + ModelErrors \tag{1.1}$$

where *educ* and *lnwage* denote respectively years of schooling and logarithm of wages [see, for example, Griliches (1977), Angrist and Krueger (1991), Blackburn and Neumark (1992), Card (1995)]. The coefficient β represents the rate of return to schooling and is often the parameter of primary interest.

1.1 Common problems with the usual Wald-type inference

The usual practice in such cases is to report a $\sqrt{n}$-consistent point estimator $\widehat{\beta}$ along with the standard error. More informative is the practice of reporting the corresponding $(1-\epsilon)$-level Wald-type confidence region given by

$$\left\{ \beta_0 : \frac{n(\widehat{\beta}-\beta_0)^2}{\widehat{V}} \leq \chi_1^2(1-\epsilon) \right\} \tag{1.2}$$

where n denotes the number of observations, $\widehat{V}$ denotes a consistent estimator of the asymptotic variance of $\sqrt{n}(\widehat{\beta} - \beta_0)$ and $\chi^2_1(1-\epsilon)$ denotes the $(1-\epsilon)$-th quantile of a (central) χ^2_1 distribution. Such regions are called the Wald-type confidence regions because they can be obtained by inverting a size-ϵ Wald test of the hypothesis $H : \beta = \beta_0$. The finite-sample properties of such regions depend on the quality of the first-order asymptotic approximation

$$\frac{\sqrt{n}\left(\widehat{\beta} - \beta\right)}{\sqrt{\widehat{V}}} \xrightarrow{d} N(0,1). \tag{1.3}$$

However, it is well known that in finite-sample, the Wald-type confidence regions are not invariant under (real) one-to-one transformations of the parameters [see Gregory and Veall (1985)]. For example, (if $\beta \neq 0$) and if β^3 is the parameter of interest then, in general,

$$\left\{\beta_0^3 : \frac{n(\widehat{\beta} - \beta_0)^2}{\widehat{V}} \leq \chi^2_1(1-\epsilon)\right\} \neq \left\{\beta_0^3 : \frac{n(\widehat{\beta}^3 - \beta_0^3)^2}{9\widehat{\beta}^4\widehat{V}} \leq \chi^2_1(1-\epsilon)\right\}.$$

A more serious drawback of the Wald-type confidence regions was highlighted by Dufour (1997). To recognize the problem it may be helpful to restrict our attention to the inference on β based on instrumental variables (IV). Suppose that the variable *educ* is correlated with the *ModelErrors* in (1.1) and one reports a Wald-type confidence region based on the IV estimator $\widehat{\beta}$ using exogenous instruments that are, unfortunately, also uncorrelated with *educ*. Based on such exogenous but irrelevant instruments, the reduced forms corresponding to (1.1) with any value of the parameter β in the entire real line are observationally equivalent. Also the asymptotic approximation in (1.3) does not hold in this case [see, for example, Phillips (1989) and Staiger and Stock (1997)]. Here the Wald-type confidence region, which is always bounded by construction, is misleading in the sense that it tends to be spuriously precise and, as a result, may not even possess any coverage probability. See Dufour (1997) for an analytical discussion of the phenomenon and Zivot et al. (1998) for simulation results showing the undesirable lack of coverage probability of Wald-type confidence regions.

Dufour (1997) and Nelson and Startz (2007) discussed a broader set of models where inference based on the Wald test can be problematic. While some of these problems are fundamental to the underlying models and not particularly to the Wald test, it has been shown recently that the score test and the likelihood ratio (LR) test can be used to salvage the problem of inference at least in the context of IV regressions.

Since a major part of this thesis deals with IV regressions, we do not focus on the Wald-type inference in the rest of the paper. Inference based on the score test is often computationally more attractive and requires fewer assumptions than inference based on the LR test [see, for example, White (1982)]. Moreover, as will be clear in the sequel, the score principle has certain advantages that allow us to design a new technique of inference for subsets of parameters, which is the main contribution of this thesis. Hence we restrict our attention to inference on subsets of parameters based on the score test.

1.2 Inference based on the usual score test

Moving to a more general model [than (1.1)], let us denote the parameter(s) of interest by θ_1 and the nuisance parameter(s) by θ_2. We follow this notation in the rest of the thesis.

Provided that the method of inference allows for a score function, the usual score test rejects hypotheses of the form $H_1 : \theta_1 = \theta_{*1}$ for large values of the score statistic evaluated under the null hypothesis. For example, consider likelihood based inference of a correctly specified model under standard regularity conditions. Let $l(\theta_1, \theta_2)$ be the sum of the log likelihood function of n observations, $\nabla_i l(\theta_1, \theta_2)$ be the first partial derivative of $l(\theta_1, \theta_2)$ with respect to θ_i, and $\widehat{\mathcal{I}}_{ij}(\theta_1, \theta_2)$ be a consistent estimator of the (i, j)-th block of the partitioned information matrix for $i, j = 1, 2$. Then the score statistic for the hypothesis $H_1 : \theta_1 = \theta_{*1}$ can be defined as

$$R_1(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) = \frac{1}{n}\left[\widehat{\nabla_{1.2}}l(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))\right]' \widehat{\mathcal{I}}_{11.2}^{-1}(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))\left[\widehat{\nabla_{1.2}}l(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))\right] \quad (1.4)$$

where

$$\begin{aligned}\widehat{\nabla_{1.2}}l(\theta_1, \theta_2) &= \nabla_1 l(\theta_1, \theta_2) - \widehat{\mathcal{I}}_{12}(\theta_1, \theta_2)\widehat{\mathcal{I}}_{22}^{-1}(\theta_1, \theta_2)\nabla_2 l(\theta_1, \theta_2) \\ \mathcal{I}_{11.2}(\theta_1, \theta_2) &= \mathcal{I}_{11}(\theta_1, \theta_2) - \mathcal{I}_{12}(\theta_1, \theta_2)\mathcal{I}_{22}^{-1}(\theta_1, \theta_2)\mathcal{I}_{21}(\theta_1, \theta_2).\end{aligned}$$

$\frac{1}{n}\widehat{\nabla_{1.2}}l(\theta_1, \theta_2)$ is an estimator of the population efficient score function for θ_1; i.e., the part of the score for θ_1 orthogonal to the space spanned by the score for θ_2. Following van der Vaart (1998) we refer to $\nabla_{1.2}l(\theta_1, \theta_2)$ (and similar terms to be introduced later) as the efficient score for θ_1.[1] If the true value of θ_2, say θ_{02}, is known *a priori*, one can set $\hat{\theta}_2(\theta_{*1}) = \theta_{02}$; although then it is more efficient to define the score statistic as $\frac{1}{n}\left[\nabla_1 l(\theta_{*1}, \theta_{02})\right]' \widehat{\mathcal{I}}_{11}^{-1}(\theta_{*1}, \theta_{02})\left[\nabla_1 l(\theta_{*1}, \theta_{02})\right]$. However, knowing θ_{02} is unlikely, and in practice one usually sets $\hat{\theta}_2(\theta_{*1})$ to some estimator of θ_2 restricted by the null hypothesis $H_1 : \theta_1 = \theta_{*1}$. Rao's score test uses the restricted maximum likelihood estimator of θ_2 [see Rao (1948)], and Neyman's $C(\alpha)$ test allows for any $\sqrt{n}$-consistent estimator of θ_2 [see Neyman (1959)].

When the null hypothesis is true, $R_1(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))$ converges to a central χ^2 distribution (with degrees of freedom equal to the dimension of θ_1) under stan-

[1]It should be noted that Cox and Hinkley (1974) (page 107) referred to $\nabla_1 l(\theta_1, \theta_2)$ as the efficient score for θ_1 and subsequently in the literature $\nabla_{1.2}l(\theta_1, \theta_2)$ has been often referred to as the effective score for θ_1. We hope that our choice of terminology will not be unduly burdensome to readers.

dard regularity conditions; and hence quantiles of the χ^2 distribution can be used as the critical values for the score test. Variants of (1.4) have been widely used in the literature under more general conditions – for example, Newey and West (1987) introduced the score (Lagrange Multiplier) test in the context of Generalized Method of Moments (GMM), while Boos (1992) introduced the generalized score test in the context of General Estimating Equations (GEE).

In the next chapter we consider the score-principle for methods of inference based on the general extremum estimation framework. This framework is broad in its scope and includes inference based on widely used methods such as M, GEE and GMM estimations as special cases.

For any such method, let $R_1(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))$ denote the score statistic (modified accordingly) for testing $H_1 : \theta_1 = \theta_{*1}$ and let $R_1(\theta_{01}, \theta_{02})$ asymptotically follow $\chi^2_{\nu_1}$ where ν_i and θ_{0i} are respectively the dimension and the true value of the parameters θ_i for $i = 1, 2$. In a correctly specified model and under standard regularity conditions, $\hat{\theta}_2(\theta_{*1})$ is $\sqrt{n}$-consistent for θ_{02} and $R_1(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))$ is asymptotically equivalent to $R_1(\theta_{*1}, \theta_{02})$ for local θ_{*1}. The usual score test rejects the null hypothesis at level ϵ if $R_1(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) > \chi^2_{\nu_1}(1-\epsilon)$. Therefore, a $(1-\epsilon)$-level confidence region obtained by inverting the score test is given by

$$\left\{ \theta_{*1} | R_1(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) \leq \chi^2_{\nu_1}(1-\epsilon) \right\}.$$

While optimality properties of tests are in general complex when $\nu_1 > 1$; under certain assumptions, it can be shown that when $\nu_1 = 1$, the score test described above is a local normal approximation to the *Uniformly Most Powerful Unbiased* test for the hypothesis $H_1 : \theta_1 = \theta_{*1}$ against two-sided alternatives. Similar local optimality statements can be made on the confidence region obtained by inverting the score test.

However, it is important to note that the properties of the usual score test depend crucially on the assumption that $\hat{\theta}_2(\theta_{01})$ is a $\sqrt{n}$-consistent estimator of the nuisance parameter θ_2. When this assumption is not satisfied, the properties of the score test are not well known and, depending on the structure of the model, this may lead to over or under-rejection of the true value of the parameters of interest, i.e. the test for $H_1 : \theta_1 = \theta_{*1}$ can be over-sized or under-sized (in other words, upward or downward size-distorted).

In the paradigm of classical hypothesis testing, while an under-rejection of the true value of the parameters is associated only with the loss in power, over-rejection of the true value is more undesirable. For example, in a series of papers, Dufour and his co-authors [Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005b,a)] strongly recommended the use of projection-type tests that rule out the upward size-distortion at the cost of considerable loss in power.

The USSIV (*unbiased split-sample IV*) test considered by Chaudhuri et al. (2007) is an example of a test that over-rejects the true value of the parameters of interest due to the inconsistent estimation of the nuisance parameters. We briefly discuss their results in the next chapter of this thesis.

Inconsistent estimation of the nuisance parameters can, in principle, affect the size of the score test in different ways and a general discussion along that line is beyond the scope of this thesis. However, as mentioned earlier, if one can ensure that the score statistic $R_1(\theta_{01}, \theta_{02})$ converges to a central $\chi^2_{\nu_1}$ distribution when evaluated at the true value of the parameters, then we show that it is possible to correct for the uncontrolled over rejection of the true value of the parameters under certain conditions.

We achieve this through a new method of projection-type inference based on the score statistic $R_1(\theta_1, \theta_2)$. We also show that the new method of projection introduced in the next section is generally less conservative than the usual method of projection recommended by Dufour and his co-authors.

The main idea behind this new method of projection follows from Robins (2004), and we gratefully acknowledge the help of Jamie Robins in the conception of this thesis.

The asymptotic convergence of $R_1(\theta_{01}, \theta_{02})$ to a known distribution is the single-most important requirement for the validity of the results based the new method of projection. Although we restrict our attention to the usual χ^2 approximation of the limiting distribution, all that is actually necessary is that $R_1(\theta_{01}, \theta_{02})$ is (asymptotically) pivotal. A set of sufficient conditions required for the validity of our results in the context of extremum estimation is listed under Assumption S in the next chapter. Loosely speaking and going back to the context of maximum likelihood estimation, if $\mathcal{I}_{11.2}(\theta_{01}, \theta_{02})$ is positive definite and if for $i, j = 1, 2$, $\widehat{\mathcal{I}_{ij}}(\theta_0) \xrightarrow{P} \mathcal{I}_{ij}(\theta_0)$, then the new method of projection can be applied to solve the problem of uncontrolled over-rejection.[2] Moreover, we also show that if it is possible to obtain a $\sqrt{n}$-consistent confidence region for the nuisance parameters θ_2 then the new method of projection, unlike the usual method, does not entail any loss in asymptotic power.

1.3 *A new method of projection-type score test*

Suppose that $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is a uniform asymptotic $(1-\zeta)$ confidence region for θ_2 when the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ is true. The new method of projection-type score test rejects the null hypothesis if

(i) either $\mathcal{C}_2(1-\zeta, \theta_{*1}) = \varnothing$,

(ii) or $\inf\limits_{\theta_2 \in \mathcal{C}_2(1-\zeta, \theta_{*1})} R_1(\theta_{*1}, \theta_2) > \chi^2_{\nu_1}(1-\epsilon)$.

[2]We are unable to identify the general characteristics of models for which such conditions will be satisfied. This is because under the usual smoothness conditions, often the properties that make $\hat{\theta}_2(\theta_{01})$ inconsistent for θ_{02} also lead to $\widehat{\mathcal{I}}_{ij}(\theta_{01}, \theta_{02})$ being inconsistent for $\mathcal{I}_{ij}(\theta_{01}, \theta_{02})$ for $i, j = 1, 2$. Identifying the general characteristics of such models is a subject of future research. Some progress in that direction has been achieved in the pioneering work of Nelson and Startz (2007).

This can be seen as a two-step procedure: in the first step we construct a (restricted) confidence region for the nuisance parameters such that the region has correct coverage probability $1-\zeta$ under the null hypothesis; and in the second step we reject the null hypothesis if the infimum (with respect to θ_2 inside the confidence region) of the statistic $R_1(\theta_{*1}, \theta_2)$ is larger than the $\chi^2_{\nu_1}(1-\epsilon)$ critical value. As mentioned before, this method is motivated from Theorem 5.1 in Robins (2004). Similar techniques of hypothesis testing in the presence of nuisance parameters were also suggested by Berger and Boos (1994) and Silvapulle (1996).

While an empty confidence region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ may seem counterintuitive; it is possible to obtain a null set when the confidence region is not obtained by inverting a Wald test. Basically this means that the underlying test rejects all possible values of the nuisance parameters θ_2. In Chapter 3 we will see that such confidence regions can occur with distinctly positive probability and hence it is important to take the empty set into account. Notwithstanding, under the null hypothesis $H_1 : \theta_1 = \theta_{*1}$, the region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ contains the true value θ_{02} with asymptotic probability at least $1-\zeta$. Hence it follows from Bonferroni-type arguments that the asymptotic size of the new method of projection-type score test cannot ever exceed $\epsilon + \zeta$.

Furthermore, under standard regularity conditions and conditional on $\mathcal{C}_2(1-\zeta, \theta_{*1}) \neq \varnothing$, any point belonging to $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is $\sqrt{n}$-consistent for θ_2. Hence for $\sqrt{n}$-local θ_{*1}, it can be shown that $\inf_{\theta_2 \in \mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_{*1}, \theta_2) = R_1(\theta_{*1}, \theta_{02}) + o_p(1)$. Thus, conditional on the first-stage confidence region being non-empty, it is possible to show that the new method of projection-type score test is asymptotically equivalent to the usual score test against $\sqrt{n}$-local alternatives.

When the standard regularity conditions are not satisfied, the usual techniques of asymptotic inference applied to finite-samples can be misleading. In particular, as we have noted before, if violation of the regularity conditions leads to inconsistent estimation of the nuisance parameters θ_2, then the usual score test may over-reject the true value of the parameters θ_1. The new method of projection-type score test can be useful in guarding against the uncontrolled over-rejection under some of these non-regular cases; while enjoying the desirable local optimality properties of the usual score test (and hence the Wald and LR tests) whenever the standard regularity conditions are satisfied.

1.4 Weak instruments and weak identification - Violation of regularity conditions and methods of inference for subsets of parameters

In practice, (asymptotic) violation of the standard regularity conditions is not rare; the much studied “weak instrument” problem serves as a common example [see for example Nelson and Startz (1990a,b), Bound et al. (1995), Staiger and Stock (1997)]. If the correlations between an endogenous regressor and the corresponding instruments in a linear IV model are small, and if the level of endogeneity is high, it usually takes an unfeasibly large sample size for the standard asymptotic techniques to be valid; and in finite-samples the Wald, LR and score tests tend to be over-sized. Furthermore, the usual asymptotics degenerate when the actual correlations are zero.

To address this problem, “weak instrument” asymptotics were introduced by Staiger and Stock (1997), in which the correlations between the endogenous regressors and the corresponding instruments approach zero with increasing sample size (at rate $\sqrt{n}$). In this setting, the two-stage least squares (TSLS) and the limited information maximum likelihood (LIML) estimators of the structural coefficients, i.e. $\theta = (\theta_1', \theta_2')'$, are inconsistent and the usual Wald, LR and score tests for testing the parameter vector θ are over-sized [also see Wang and Zivot (1998) and Zivot et al. (1998)].

Alternative techniques for jointly testing all the structural coefficients have now been proposed, which are robust (in terms of size) to the presence of weak instruments. For example, the Anderson-Rubin (AR) test was recommended by Dufour (1997) and Staiger and Stock (1997), the K-test was proposed by Kleibergen (2002) and Moreira (2003) and the Conditional likelihood ratio (CLR) test was proposed by Moreira (2003).

Returning to the problem of testing subsets of structural coefficients, i.e. $H_1 : \theta_1 = \theta_{*1}$, one can always use the projection techniques based on these tests [see Dufour and Taamouti (2005b,a) and Zivot et al. (2006)]. However, such projection-type tests can be conservative. We show that our new method of projection is, in general, less conservative than the usual projection-type tests.

As an alternative to the projection-type tests, Kleibergen (2004) showed that if the instruments are strong for the nuisance parameters θ_2, one can replace θ_2 with their LIML estimators restricted by the null hypothesis, and use the K and CLR tests (with proper adjustment for the degrees of freedom) to test the null hypothesis $H_1 : \theta_1 = \theta_{*1}$. Stock and Wright (2000) and Zivot et al. (2006) suggested similar modifications to the AR test using the restricted TSLS or LIML estimators. In a recent working paper Kleibergen (2007) also showed that when the instruments are weak for the nuisance parameters θ_2, the plug-in principle applied to the AR, K and CLR tests leads to tests that are asymptotically conservative.

Stock and Wright (2000) generalized the idea of weak instruments in a linear IV regression to "weak moment conditions" in a GMM framework. The weak moment conditions lead to "weak identification" of the parameters and, as expected, the GMM estimators of the parameters are inconsistent and the usual Wald, LR and score tests for these parameters are unreliable.

Kleibergen (2005) extended his K-test to the GMM setup and also showed that *when the nuisance parameters θ_2 are strongly identified*, the K-test can be applied for testing $H_1 : \theta_1 = \theta_{*1}$ using the K-statistic evaluated at the null (i.e. plugging in the Continuous Updating estimator of θ_2 restricted by the null hypothesis) and adjusting the degrees of freedom accordingly. Similar procedures can be applied to extend the CLR test to the GMM framework when the nuisance parameters are strongly identified.

The K-statistic is, in principle, a score statistic. Hence the new method of projection can be readily applied to the K-statistic. Under the weak identification framework, we show that our projection-type test statistic (based on the K-principle) is asymptotically equivalent to Kleibergen's K-statistic for $\sqrt{n}$-local alternatives whenever the nuisance parameters θ_2 are strongly identified.

Furthermore, because in this particular setup it is always (i.e. even for unidentified θ_2) possible to construct asymptotically valid confidence sets for the nuisance parameters under the null, we show that the size of the new method of projection-type test can be bounded from above by any pre-specified value.[3]

We recommend the use of the new method of projection-type score tests in cases where inconsistent estimation of the nuisance parameters leads to the standard tests being over-sized. While the usual projection techniques can also be used in these cases to guard against the over-rejection of the true value of the parameters of interest, the new method of projection-type tests can be shown to be more powerful than the usual projection-type tests under quite general conditions.

[3]In an e-mail exchange, Frank Kleibergen recently told us that he and Sophocles Mavroeidis are jointly working on showing that the plug-in technique applied to the S, K and GMM-M (CLR) tests leads to conservative tests whenever the nuisance parameters are weakly identified. This opens a new possibility for improvement when the nuisance parameters are weakly identified: a judicious choice of ζ may increase the asymptotic relative efficiency (ARE) of the new method of projection-type test with respect to the K-test. We plan to pursue this in the future.

1.5 Notations

We use the following notations in the rest of the thesis. If $A = [A_1, \ldots, A_{bc}]$ is an $a \times bc$ matrix, A^+ is its unique Moore-Penrose inverse, $vec(A) := [A_1', \ldots, A_{bc}']'$, $devec_c(A') := \left[(A_1, \ldots, A_c)', \ldots, (A_{(b-1)c+1}, \ldots, A_{bc})'\right]$ and $\|A\| := \sqrt{\text{trace}(A'A)}$. If A is full column rank then $P(A) = A(A'A)^{-1}A'$ and $N(A) = I_a - P(A)$ where I_a is the $a \times a$ identity matrix. If A is a symmetric positive semi-definite matrix then $A^{\frac{1}{2}}$ is the lower-triangular Cholesky factor of A such that $A = A^{\frac{1}{2}}A^{\frac{1}{2}'}$. If $A = ((A_{ij}))_{i,j=1,2}$ is such that the diagonal blocks A_{11} and A_{22} are non-singular then $A_{ii.j} = A_{ii} - A_{ij}A_{jj}^{-1}A_{ji}$ denotes the Schur complement of A_{jj} for $i \neq j = 1, 2$. Lastly, we use the acronym w.p.a.1 for "with probability approaching one".

Chapter 2

THE NEW PROJECTION-TYPE SCORE TEST IN EXTREMUM ESTIMATIONS

Consider a random sample $z = (z_1, \ldots, z_n)' \in \mathcal{S}$ drawn from the unknown distribution $P_{\theta_0} \in \{P_\theta : \theta_{\nu\times 1} \in \Theta\}$. Let the true parameter value $\theta_0 = (\theta_{01}', \theta_{02}')'$ be such that $\theta_{0i} \in$ interior(Θ_i) where Θ_i is a compact (ν_i- dimensional) subset of $\mathbb{R}^{\nu_i}$ for $i = 1, 2$. The parameter space $\Theta = \Theta_1 \times \Theta_2$ is a compact subset of $\mathbb{R}^{\nu}$ where $\nu = \nu_1 + \nu_2$.

Define an extremum estimator of θ based on the random sample z as

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} Q_n(z, \theta) \tag{2.1}$$

where $Q_n(z, \theta) : \mathcal{S} \times \Theta \mapsto \mathbb{R}$ is any criterion function measurable in z for all $\theta \in \Theta$. The definition of the extremum estimator in (2.1) covers the M-estimators such as those obtained by maximum likelihood, quasi-maximum likelihood, least squares, etc., and the Minimum Distance estimators such as the GMM estimator. GMM estimation has received much attention in the last two decades and offers an interesting application of the new method of projection-type score test. Hence we discuss it separately in the next chapter under more specific assumptions. The general theory discussed in this chapter is probably better suited for M-estimators.[1]

For simplicity, we base the discussion of the new projection-type score test

---

[1] Of course any M-estimator, viewed as a solution of a first order condition (i.e. what van der Vaart (1998) calls a Z-estimator), can be interpreted as a GMM estimator [page 2116, Newey and McFadden (1994)].

for the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ on a set of "high-level" assumptions summarized under Assumption A [see chapter 4, Amemiya (1985) for details].

Assumption A:

A1. $Q_n(\theta) \equiv Q_n(z, \theta)$ is twice-continuously differentiable and $n^{-1}Q_n(\theta)$ converges uniformly to a non-stochastic function $Q(\theta)$ which has a unique maximum at θ_0.

A2. $n^{-1/2}\triangledown_i Q_n(\theta_0) \xrightarrow{d} \Psi_i$ where $\triangledown_i Q_n(\theta) := \partial Q_n(Z, \theta)/\partial\theta_i$ for $i = 1, 2$. Also $[\Psi_1', \Psi_2'] \sim N(0_\nu, B(\theta_0))$ for some finite, positive definite matrix $B(\theta_0)$ where $B(\theta) = (B_{ij}(\theta))_{i,j=1,2}$ is continuous in an open neighborhood of θ_0.

A3. $n^{-1}\triangledown_{\theta\theta} Q_n(\theta)$ converges uniformly in probability to a finite, continuous, negative definite matrix $A(\theta) = (A_{ij}(\theta))_{i,j=1,2}$ where the matrix $\triangledown_{\theta\theta} Q_n(\theta) = (\triangledown_{ij} Q_n(\theta))_{i,j=1,2}$ and $\triangledown_{ij} Q_n(\theta) := \partial(\triangledown_i Q_n(\theta))'/\partial\theta_j$.

We use $A \equiv A(\theta_0)$ and $B \equiv B(\theta_0)$ for notational convenience. Following Assumption A3, we use $\widehat{A}(\theta) = n^{-1}\triangledown_{\theta\theta} Q_n(\theta)$ as the estimator of $A(\theta)$. We refer to $\widehat{B}(\theta)$ as some consistent estimator for $B(\theta)$ without explicitly mentioning its functional form.

Before introducing the new projection-type score test in this context, we briefly review the usual score test for $H_1 : \theta_1 = \theta_{*1}$.

Denoting $\tilde{\theta}_* = (\theta_{*1}', \tilde{\theta}_{n2}'(\theta_{*1}))'$, the score statistic for testing $H_1 : \theta_1 = \theta_{*1}$ is defined as

$$R_1(\tilde{\theta}_*) = \left[\frac{1}{\sqrt{n}}\triangledown_{1.2} Q_n(\tilde{\theta}_*)\right]' \left[\widehat{G}_1(\tilde{\theta}_*)\widehat{B}(\tilde{\theta}_*)\widehat{G}_1'(\tilde{\theta}_*)\right]^+ \left[\frac{1}{\sqrt{n}}\triangledown_{1.2} Q_n(\tilde{\theta}_*)\right] \tag{2.2}$$

$$\text{where } \tilde{\theta}_{n2}(\theta_{*1}) = \arg\max_{\theta_2 \in \Theta_2} Q_n(\theta_{*1}, \theta_2), \tag{2.3}$$

$$\triangledown_{1.2} Q_n(\theta) = \triangledown_1 Q_n(\theta) - \widehat{A}_{12}(\theta)\widehat{A}_{22}^{-1}(\theta)\triangledown_2 Q_n(\theta) \text{ and } \widehat{G}_1(\theta) = [I_{\nu_1}, -\widehat{A}_{12}(\theta)\widehat{A}_{22}^{+}(\theta)].$$

The score test rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $R_1(\tilde{\theta}_*) >$

$\chi^2_{\nu_1}(1-\epsilon)$. An alternative form of the score statistic, motivated by the Lagrange Multiplier principle of Aitchison and Silvey (1958) and Silvey (1959), is given by

$$R_1^{\text{alt}}(\tilde{\theta}_*) = \left[\frac{1}{\sqrt{n}}\triangledown_1 Q_n(\tilde{\theta}_*)\right]' \left[\widehat{G}_1(\tilde{\theta}_*)\widehat{B}(\tilde{\theta}_*)\widehat{G}_1'(\tilde{\theta}_*)\right]^+ \left[\frac{1}{\sqrt{n}}\triangledown_1 Q_n(\tilde{\theta}_*)\right].^2 \tag{2.4}$$

While it is obvious that both forms of the score statistic are observationally equivalent when the nuisance parameters are replaced by some θ_2 such that $\partial Q_n(\theta_{*1}, \theta_2)/\partial\theta_2 = 0$; the score statistic in (2.2) has a major practical advantage: it allows θ_2 to be replaced by any $\sqrt{n}$-consistent estimator [see Lemma A.4(iii) in the Appendix]. This particular flexibility of the score statistic $R_1(\tilde{\theta}_*)$ is later utilized in constructing the new projection-type score test. The asymptotic validity of the alternative form of the score statistic in (2.4), however, is only maintained when the nuisance parameters are replaced by some θ_2 such that $\partial Q_n(\theta_{*1}, \theta_2)/\partial\theta_2 = o_p(1)$ [under Assumption A, the estimator $\tilde{\theta}_{n2}(\theta_{*1})$ in (2.3) satisfies this property w.p.a.1] and this can be computationally costly [see example 3.1 in Bera and Bilias (2001)]. In the rest of the paper, by "efficient score statistic", we will refer to a statistic similar to the one considered in (2.2) – i.e. a quadratic form of the estimated efficient score for θ_1, which simultaneously admits interpretation both as Rao's score and Neyman's $C(\alpha)$ statistic.

2.1 The new projection-type score test

Suppose that it is possible to construct a confidence region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ for θ_2 such that under the null hypothesis $H_1 : \theta_1 = \theta_{*1}$, the region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ has uniform asymptotic coverage probability (at least) $1-\zeta$. Denoting $\theta_* = (\theta'_{*1}, \theta'_{*2})'$,

[2]Denoting $\partial Q_n(\theta)/\partial\theta$ by $\triangledown_\theta Q_n(\theta)$, the score statistic in (2.4) can be also alternatively written as $R_1^{\text{alt2}}(\tilde{\theta}_*) = \left[\frac{1}{\sqrt{n}}\triangledown_\theta Q_n(\tilde{\theta}_*)\right]' \left[\widehat{G}(\tilde{\theta}_*)\widehat{B}(\tilde{\theta}_*)\widehat{G}'(\tilde{\theta}_*)\right]^+ \left[\frac{1}{\sqrt{n}}\triangledown_\theta Q_n(\tilde{\theta}_*)\right]$, where $\widehat{G}(\theta) = [\widehat{G}_1'(\theta), G_2']'$ for any arbitrary $\nu_2 \times \nu$ matrix G_2, as long as $\triangledown_2 Q_n(\tilde{\theta}_*) = 0$. We mention this form because, as we will see later, Kleibergen's K-statistic is defined in a similar spirit.

the new projection-type score test rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ if

(i) either $\mathcal{C}_2(1-\zeta, \theta_{*1}) = \varnothing$

(ii) or $\inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_*) > \chi^2_{\nu_1}(1-\epsilon)$.

At this point we do not specify how the confidence region for the nuisance parameters θ_2 is obtained in the first step. In the next chapter we discuss in detail some of the important advantages derived from a judicious choice of the first-stage confidence region.

It is important to note that Assumption A does not include non-regular cases such as the instrumental variables regression with weak instruments. In such a non-regular case while it can still be possible to obtain a valid (though possibly conservative) confidence region for the nuisance parameters, consistent point-estimation of the nuisance parameters (and the parameters of interest) may not be feasible [see, for example, Section 2.3]. Of course, if it is possible to obtain $\sqrt{n}$-consistent estimators for a subset of the nuisance parameters, then constructing the confidence region for those parameters is unnecessary.

Theorem 2.1

(i) Let $\mathcal{C}_2(1-\zeta, \theta_1)$ *be such that* $\lim_{n\to\infty} Pr_{\theta_{01}}\left[\theta_{02} \in \mathcal{C}_2(1-\zeta, \theta_{01})\right] \geq 1-\zeta$. *Then under Assumption A,*

$$\lim_{n\to\infty} Pr_{\theta_{01}}\left[\{\mathcal{C}_2(1-\zeta,\theta_{01}) = \varnothing\} \cup \{\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})} R_1(\theta_{01},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon)\}\right] \leq \zeta + \epsilon$$

(ii) Let $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1$ *where* $d_1 \in \mathbb{R}^{\nu_1}$ *is fixed. Let* $\mathcal{C}_2(1-\zeta,\theta_{*1})$ *be nonempty w.p.a.1 and suppose that for any* $\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{*1})$, $\sqrt{n}\|\theta_{*2}-\theta_{02}\| = O_p(1)$. *Then under Assumption A,*

$$\begin{aligned}&\lim_{n\to\infty} Pr_{\theta_{01}}\left[\{\mathcal{C}_2(1-\zeta,\theta_{*1}) = \varnothing\} \cup \{\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_{*1},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon)\}\right]\\ = \;&\lim_{n\to\infty} Pr_{\theta_{01}}\left[R_1(\theta_{*1},\theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)\right]\end{aligned}$$

Theorem 2.1 discusses the asymptotic properties of the the new projection-type score test and is proved in the Appendix. It shows that the size of the new projection-type score test is always bounded from above by $\epsilon + \zeta$, irrespective of the asymptotic length of the first-stage confidence region. Moreover, if w.p.a.1 the confidence region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is non-empty and the value of θ_2 where the infimum of the test statistic is attained within $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is $\sqrt{n}$-consistent for θ_2, then Theorem 2.1 also shows that this test is asymptotically locally equivalent to the infeasible efficient score test, which rejects $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $R_1(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$. The infeasible efficient score test uses the unknown true value of the nuisance parameters θ_2 and enjoys certain local optimality properties under standard regularity conditions. Often this asymptotic equivalence can be ensured by constructing the confidence region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ such that it belongs to the $\sqrt{n}$-neighborhood of the true value θ_{02}, as is mentioned in the statement of the theorem. We describe the process of the construction of such confidence regions in the next chapter.

Finally we note that, under the conditions of the theorem, the local asymptotic equivalence of the new projection-type score test extends to the usual score test (mentioned before in (2.2)) and the Wald test, which rejects $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $W_1(\theta_{*1}) > \chi^2_{\nu_1}(1-\epsilon)$ where

$$W_1(\theta_{*1}) = n(\widehat{\theta}_{n1} - \theta_{*1})'\widehat{\Omega}^{+}_{11}(\widehat{\theta}_n)(\widehat{\theta}_{n1} - \theta_{*1})$$

and $\widehat{\Omega}_{11}(\theta)$ is the top-left $\nu_1 \times \nu_1$ block of $\widehat{\Omega}(\theta) = ((\widehat{\Omega}_{ij}(\theta)))_{i,j=1,2} = \widehat{A}^{+}(\widehat{\theta}_n)\widehat{B}(\widehat{\theta}_n)\widehat{A}^{+}(\widehat{\theta}_n)$ [see Lemma A.4 in the Appendix].[3]

Readers are referred to Gourieroux et al. (1983) and Newey and McFadden (1994) for other forms of the Wald and score tests. We do not consider the LR

[3]The form of the Wald statistic and the score statistic can be simplified if it is known *a priori* that $A(\theta_0) = -B(\theta_0)$ [see White (1982)]. The usual score test and the Wald test are asymptotically equivalent to the LR test against $\sqrt{n}$-local alternatives when $A(\theta_0) = -B(\theta_0)$.

test here because the asymptotic χ^2 distribution of the test statistic requires an additional assumption that $A = -B$.[4] We also do not consider the Hausman-type tests because these tests are locally optimal for a different hypothesis, $A_{22}^{-1}A_{21}(\theta_1 - \theta_{*1}) = 0$, whose equivalence with the hypothesis under consideration depends on the stringent condition that $\nu_2 \leq \nu_1$ and A_{21} is full row rank [see Holly (1982)].

It should, however, be mentioned that like the usual score test, the new projection-type score test is also not particularly suited for cases where the nuisance parameters are absent under the hypothesis being tested [see Davies (1977, 1987)]. For example, consider a regression model with an additive non-linearity, which takes the form [5]

$$y_t = \theta_1 f(\theta_2, x_t) + u_t, \quad \text{where} \quad u_t \sim (iid)N(0, 1).$$

Under the null hypothesis $H_1 : \theta_1 = 0$, the nuisance parameters θ_2 are absent; meaning that any valid confidence region for θ_2, obtained under the null hypothesis, should be identical to the entire parameter space for θ_2. Hence the new projection-type test reduces to comparing the $\inf_{\theta_2 \in \Theta_2} R_1(0, \theta_2)$ to the $\chi^2_{\nu_1}$ critical value. To apply the usual score test in such cases, it is important to consider the restricted estimator of θ_2 as a stochastic process of θ_1; treating it simply as the $\arg\max_{\theta_2 \in \Theta_2} Q_n(0, \theta_2)$ leaves room for implementational ambiguity. The usual score test and the new projection-type test have poor power at the point of unidentification; in fact, the latter provides a lower bound to the asymptotic power of the usual score test. Of course, in this case one can fol-

[4]Under the null hypothesis the LR statistic, in general, tends to a weighted sum of χ^2 variables with unknown weights [see Foutz and Srivastava (1977) and Kent (1982)].

[5]See Hansen (1996) for examples of the regression form: $y_1 = x_t\gamma + \theta_1 f_t(\theta_2, z_t) + u_t$ and cases where it is of interest to test if the nonlinear term $f(\theta_2, z_t)$ enters the regression or not; i.e., to test $H_1 : \theta_1 = 0$. See Nelson and Startz (2007) for more examples.

low Conniffe (2001) and replace θ_2 by its unrestricted estimator in the score statistic $R_1(0,\theta_2)$ defined in (2.2) and obtain better power as long as the true value $\theta_{01} \neq 0$. Note that the alternative forms of the score statistic do not allow such modification. This technique can be extended to the new projection-type score test by constructing an unrestricted first-stage confidence region for the nuisance parameters θ_2.

2.2 Sufficient conditions for the validity of the new test

Finally, it is important to note that the new method of projection offers wide flexibility in designing a test for the sub-vector θ_1. Provided that it is possible to obtain a valid first-stage confidence region for the nuisance parameters θ_2, the only requirement in designing a valid projection-type (score) test for the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ is a statistic which is (asymptotically) pivotal under the null hypothesis. Although for local optimality properties it is necessary that such a (pivotal) statistic be based on the efficient score for θ_1; a valid (not necessarily optimal) test for $H_1 : \theta_1 = \theta_{*1}$ can be based on a wider choice of (pivotal) statistic. For example, any functional $f_n : \mathcal{S} \times \Theta \mapsto \mathbb{R}^{\nu}$ satisfying the following set of sufficient conditions, summarized under Assumption S, can be used to construct such a (pivotal) statistic that can be used in designing the new projection-type score test.

Assumption S:

S1. $f_n(z,\theta) = H_n(z,\theta)h_n(z,\theta)$ where $H_n = ((H_{n(i,j)}))$ is such that the (i,j)-th element $H_{n(i,j)} : \mathcal{S} \times \Theta \mapsto \mathbb{R}$ for $i = 1,\ldots,\nu$, $j = 1,\ldots,k$ and $h_n : \mathcal{S} \times \Theta \mapsto \mathbb{R}^k$ for some finite positive integer k.

S2. $h_n(z,\theta) \Rightarrow \xi(\theta) \sim N(\mu(\theta),\Xi(\theta))$ such that $\mu(\theta_0) = 0$ and $\Xi(\theta_0)$ is finite.[6]

[6]"$\Rightarrow$" is used to denote weak convergence to the Gaussian stochastic process ξ, indexed by θ [see Andrews (1994)].

$H_n(z,\theta_0)$ and $h_n(z,\theta_0)$ are asymptotically uncorrelated.

S3. $V_n(\theta) = H_n(z,\theta)\Xi(\theta)H_n'(z,\theta)$ is such that $\lim_{n\to\infty} V_n(\theta)$ is finite, continuous and positive definite at θ_0 w.p.a.1. $\widehat{V}_n(\theta) = H_n(z,\theta)\widehat{\Xi}(\theta)H_n'(z,\theta)$ such that $\|\widehat{V}_n(\theta) - V_n(\theta)\| = o_p(1)$ in some open neighborhood of θ_0.

Assumption A implies Assumption S and hence the following discussion based on Assumption S is also applicable to inference based Assumption A.

Using Assumption S, an asymptotically size-ϵ test for the hypothesis $H_* : \theta = \theta_*$ on the full parameter vector can be designed by rejecting the hypothesis when $R(\theta_*) > \chi^2_\nu(1-\epsilon)$ where

$$R(\theta_*) = f_n'(z,\theta_*)\widehat{V}^+(\theta_*)f_n(z,\theta_*).^7 \tag{2.5}$$

If, in addition, $\text{plim}_{n\to\infty} H_n(\theta_*)\mu(\theta_*) \neq 0$ w.p.a.1 for $\theta_* \neq \theta_0$, then this size-ϵ test will be asymptotically unbiased. However, evaluating the loss in ARE with respect to the usual score test needs more specific assumptions.

The same principle can also be used to construct an asymptotically pivotal statistic for testing the hypothesis $H_1 : \theta_1 = \theta_{*1}$ on the sub-vector θ_1. In particular, if we can partition H_n and $\widehat{V}_n$ such that $H_n' = [H_{n[1]}', H_{n[2]}']$ where $H_{n[i]}$ is $\nu_i \times k$ for $i = 1, 2$ and $\widehat{V}_n = ((\widehat{V}_{n[i,j]}))_{i,j=1,2}$, then an asymptotically pivotal statistic for the new projection-type score test can be constructed as

$$R_1(\theta) = f_{n[1.2]}'(z,\theta)\widehat{V}^+_{n[11.2]}(\theta)f_{n[1.2]}(z,\theta) \text{ where} \tag{2.6}$$

$$f_{n[1.2]}(z,\theta) = H_{n[1]}(z,\theta)h_n(z,\theta) - \widehat{V}_{n[12]}(\theta)\widehat{V}^+_{n[22]}(\theta)H_{n[2]}(z,\theta)h_n(z,\theta) \tag{2.7}$$

As before, a size-ϵ test for the hypothesis $H_1 : \theta_1 = \theta_{*1}$ on the sub-vector θ_1 can be designed by rejecting $H_1 : \theta_1 = \theta_{*1}$ when $R_1(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$. This test

[7] The description of the statistic $R(\theta_*)$ fits the LM (K) statistic proposed by Kleibergen (2002) and Moreira (2003).

will be asymptotically unbiased if $\text{plim}_{n\to\infty} H_{n[1]}(\theta_{*1},\theta_{02})\mu(\theta_{*1},\theta_{02}) \neq 0$ w.p.a.1 for $\theta_{*1} \neq \theta_{01}$. However, θ_{02} is unlikely to be known, and the test remains valid if θ_2 is replaced by any $\sqrt{n}$-consistent estimator.[8] When it is only possible to obtain a valid confidence region for θ_2 but not a consistent estimator, then the new method of projection-type score test based on the statistic $R_1(\theta_{*1},\theta_2)$ in (2.6), which is asymptotically pivotal at $\theta = \theta_0$, can be used for testing $H_1 : \theta_1 = \theta_{*1}$.

In the following section we present a simple application of this new projection-type score test in the context of split-sample IV regression. The description is rather elaborate and is hoped to be helpful in explaining the methodology.

2.3 Application to a split-sample linear IV regression

Consider the linear IV model

$$\left.\begin{aligned} y &= X_1\theta_{01} + X_2\theta_{02} + u \\ X_1 &= Z\Pi_1 + \eta_1 \\ X_2 &= Z\Pi_2 + \eta_2 \end{aligned}\right\} \tag{2.11}$$

where y is the dependent variable, $X = [X_1, X_2]$ are the endogenous regressors, u, $\eta = [\eta_1, \eta_2]$ are the unobserved correlated structural errors and Z is the

[8]It is worthwhile to observe that if $\tilde{\theta}_{n2}(\theta_{*1})$ is a root of $H_{n[2]}(\theta_{*1},\theta_2)h_n(\theta_{*1},\theta_2) = 0_{\nu_2\times 1}$, then, using (2.5) and defining $\tilde{\theta}_* = (\theta_{*1}', \tilde{\theta}_{n2}'(\theta_{*1}))'$, other statistics for $H_1 : \theta_1 = \theta_{*1}$ can also be constructed as:

$$R_1^{\text{alt}}(\tilde{\theta}_*) = h_n'(z,\tilde{\theta}_*)H_{n[1]}'(z,\tilde{\theta}_*)\widehat{V}_{n[11.2]}^{+}(\tilde{\theta}_*)H_{n[1]}(z,\tilde{\theta}_*)h_n(z,\tilde{\theta}_*) \tag{2.8}$$

$$R_1^{\text{alt2}}(\tilde{\theta}_*) = R(\tilde{\theta}_*) \tag{2.9}$$

$$\text{where} \quad R_1(\tilde{\theta}_*) \overset{\text{obs}}{\equiv} R_1^{\text{alt}}(\tilde{\theta}_*) \overset{\text{obs}}{\equiv} R_1^{\text{alt2}}(\tilde{\theta}_*). \tag{2.10}$$

The first term in (2.10) is analogous to what we have throughout referred to as the efficient score statistic; the second term is analogous to the usual form of the score statistic derived from the LM principle; and the last term is analogous to the form used by Kleibergen (2004, 2005) to define the K statistic for testing $H_1 : \theta_1 = \theta_{*1}$. Lastly, note that all the three forms of the statistic are robust to the scaling of the term $H_n(z,\theta)$ by nonsingular matrices. This allows for different orders of magnitude (in a possibly probabilistic sense) of $H_{n[1]}(z,\theta)$ and $H_{n[2]}(z,\theta)$.

matrix of non-stochastic instruments.[9] Let the dimensions of θ_i and Π_i be respectively $\nu_i \times 1$ and $k \times \nu_i$ for $i = 1, 2$ where $\nu = \nu_1 + \nu_2$ and ν_1, ν_2 and k are fixed and finite integers. Assume that the order condition $k \geq \nu$ is satisfied. We do not, however, impose the restriction of full column rank on $\Pi = [\Pi_1, \Pi_2]$.

Suppose that there are n i.i.d. observations on y, X and Z, and assume that for $t = 1, \ldots, n$ the structural errors

$$(u_t, \eta_{1t}, \eta_{2t}) \overset{\text{i.i.d.}}{\sim} N(0, \Sigma) \text{ where } \Sigma = \begin{pmatrix} 1 & \rho_{u1} & \rho_{u2} \\ \rho_{1u} & 1 & \rho_{12} \\ \rho_{2u} & \rho_{21} & 1 \end{pmatrix} \text{ is unknown.} \tag{2.12}$$

Split the sample randomly into two sub-samples denoted by a and b – the first one containing n_a observations and the second one containing $n_b = n - n_a$ observations such that $min\{n_a, n_b\} > k$. Let $y_{(i)}$, $X_{(i)}$ and $Z_{(i)}$ represent the matrices containing the n_i observations in the ith sub-sample ($i = a, b$) where the observations are stacked in rows.

Define $\widehat{X}_{(ij)} = Z_{(i)}\widehat{\Pi}_{(j)}$ and $\widehat{\Pi}_{(j)} = (Z'_{(j)}Z_{(j)})^{-1}Z'_{(j)}X_{(j)}$ for $i, j = a, b$. We point out that the USSIV estimator of θ, proposed by Angrist and Krueger (1995), can alternatively be defined as an extremum estimator that maximizes the objective function

$$Q_n(\theta) = -\frac{1}{2}\left(y_{(a)} - X_{(a)}\theta\right)' \widehat{X}_{(ab)} \left(X'_{(a)}\widehat{X}_{(ab)}\right)^{-1} \widehat{X}'_{(ab)} \left(y_{(a)} - X_{(a)}\theta\right) \tag{2.13}$$

with respect to θ provided $X'_{(a)}\widehat{X}_{(ab)}$ and is nonsingular. The objective function focuses just on the structural parameters θ_1 and θ_2 by partialling out the other parameters Π and Σ. In this model, Π can be estimated consistently as long

[9] We impose this rather restrictive assumption on the instruments just for the purpose of exposition. The results in this section are valid with instruments that are independent, exogenous or predetermined with respect to the structural errors [see Chaudhuri et al. (2007)].

as $\lim n^{-1}Z'Z$ is positive definite and Σ can be estimated consistently in any $\sqrt{n}$-neighborhood of Π and θ_0. The objective function is twice continuously differentiable with respect to θ. The gradients and the Hessian are respectively

$$\begin{aligned} \nabla_\theta Q_n(\theta) &= \widehat{X}'_{(ab)}\left(y_{(a)} - X_{(a)}\theta\right) \text{ giving } \nabla_i Q_n(\theta) = \widehat{X}'_{(ab)i}\left(y_{(a)} - X_{(a)}\theta\right) \text{ for } i = 1,2; \\ \nabla_{\theta\theta} Q_n(\theta) &= -\widehat{X}'_{(ab)}X_{(a)} \text{ giving } \nabla_{ij} Q_n(\theta) = -\widehat{X}'_{(ab)i}X_{(a)j} \text{ for } i,j = 1,2. \end{aligned}$$

Assuming that $n_a/n \to c$ where c is a constant $\in (0,1)$, it is instructive to see how Assumption A works in this example. Let us do it in reverse order starting with Assumption A3, which describes the curvature of the objective function.

$$\frac{1}{n}\nabla_{\theta\theta}Q_n(\theta) \xrightarrow{P} -\text{plim}\frac{n_a}{n}\left[\widehat{\Pi}'_{(b)}\frac{Z'_{(a)}Z_{(a)}}{n_a}\Pi + \widehat{\Pi}'_{(b)}\frac{Z'_{(a)}\eta_{(a)}}{n_a}\right] = -c\Pi'\Upsilon\Pi \equiv A, \quad (2.14)$$

if $\lim n_i^{-1}Z'_{(i)}Z_{(i)} = \Upsilon$ for $i = a, b$. The probability limit does not depend on θ and hence is uniform in θ and, further if Υ is positive definite and Π is full column rank, then the probability limit is a negative definite matrix as required in A3.

Now consider Assumption A2. Again, the same set of sufficient conditions, i.e. the convergence of $n_i^{-1}Z'_{(i)}Z_{(i)}$ to a positive definite matrix Υ (for $i = a, b$) and the full column rank of Π ensures that A2 holds because

$$\frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_0) = \sqrt{\frac{n_a}{n}}\widehat{\Pi}'_{(b)}\frac{Z'_{(a)}u_{(a)}}{\sqrt{n_a}} \xrightarrow{d} N(0, c\Pi'\Upsilon\Pi \equiv B). \quad (2.15)$$

Finally, consider Assumption A1. The same set of conditions ensures that

$$\begin{aligned} \frac{1}{n}Q_n(\theta) &= -\frac{n_a}{2n}(\theta_0-\theta)'\frac{\widehat{X}'_{(ab)}X_{(a)}}{n_a}(\theta_0-\theta) - \frac{n_a}{2n}\frac{u'_{(a)}\widehat{X}_{(ab)}}{n_a}\left(\frac{X'_{(a)}\widehat{X}_{(ab)}}{n_a}\right)^{-1}\frac{\widehat{X}'_{(ab)}u_{(a)}}{n_a} \\ &\quad -\frac{n_a}{2n}(\theta_0-\theta)'\frac{\widehat{X}'_{(ab)}u_{(a)}}{n_a} - \frac{n_a}{2n}\frac{u'_{(a)}\widehat{X}_{(ab)}}{n_a}\left(\frac{X'_{(a)}\widehat{X}_{(ab)}}{n_a}\right)^{-1}\frac{\widehat{X}'_{(ab)}X_{(a)}}{n_a}(\theta_0-\theta) \\ &\xrightarrow{P} -\frac{c}{2}(\theta_0-\theta)'\Pi'\Upsilon\Pi(\theta_0-\theta) \equiv Q(\theta) \end{aligned} \quad (2.16)$$

where $Q(\theta)$ is a non-stochastic function of θ attaining a unique maximum at $\theta = \theta_0$. Hence under the conditions that $n_i^{-1}Z'_{(i)}Z_{(i)}$ converges to a positive definite matrix Υ (for $i = a, b$) and Π is full column rank; the USSIV estimator

$$\widehat{\theta}_{\text{USSIV}} = \left(\widehat{X}'_{(ab)}X_{(a)}\right)^{-1}\widehat{X}'_{(ab)}y_{(a)}, \tag{2.17}$$

obtained by minimizing (2.13), satisfies Assumption A and all the results discussed before in this chapter hold. Moreover, since (2.14) and (2.15) imply that $A = -B$ one can also construct a LR test for testing the parameter vector θ or the sub-vector θ_1. However, it is obvious that under such standard circumstances, the USSIV-framework is not useful because of the lost precision (information) due to sample-splitting. Inference based on the standard TSLS obtained by maximizing the objective function $Q_n^{\text{TSLS}}(\theta) = (y - X\theta)'P(Z)(y - X\theta)$ is more efficient when Assumption A is satisfied.

2.3.1 Why consider the USSIV-framework then?

At the risk of repetition, note that the convergence results in (2.14) – (2.16) depend crucially on two conditions – the positive definiteness of $\lim n^{-1}Z'Z (= \Upsilon)$ and the full column rank of Π. Combined together, they imply the requirement of positive definiteness of $\Pi'\Upsilon\Pi$, the probability limit of ($1/n$ times) the inner product of the projection of X on to the space spanned by the columns of Z, and hence is related to the relevance of the instruments Z for the endogenous regressors X. Intuitively this means that the space spanned by the projected endogenous regressors is ν-dimensional where ν is the number of endogenous regressors, i.e., there is no multicollinearity among the columns of projected X.

The condition related to the relevance of the instruments has received a lot of attention recently and is helpful in recognizing the usefulness of the USSIV-framework. When $\Pi'\Upsilon\Pi$ is rank deficient, the standard asymptotics degenerate and it is not possible to find consistent estimators for every element of θ [see Phillips (1989)]. Because of the lack of uniform convergence in (2.14) – (2.16), when $\Pi'\Upsilon\Pi$ is nearly rank deficient, it may take an unfeasibly large number of observations for the standard asymptotics (under Assumption A) to provide a good approximation of the finite-sample properties of the inference [see Bound et al. (1995)].

The weak instrument asymptotics characterizes the near rank deficiency of $\Pi'\Upsilon\Pi$ using a local-to-zero approximation of Π. Staiger and Stock (1997) also showed that such approximations are more representative of the finite-sample properties of the inference when Π (and hence $\Pi'\Upsilon\Pi$) is nearly rank deficient. Assumption WI below describes a simplified form of the weak instrument characterization of Staiger and Stock (1997).

Assumption WI:

For $i = 1, 2$, let $\Pi_i = 0_{k\times\nu_i} 1_{[\delta_i=0]} + n^{-1/2}\mathbb{C}_i 1_{[\delta_i=1/2]} + \mathbb{C}_i 1_{[\delta_i=1]}$ where $\mathbb{C}_i$ is a $k \times \nu_i$ matrix of fixed and bounded elements such that $\mathbb{C} = [\mathbb{C}_1, \mathbb{C}_2]$ is full column rank and δ_i are constants such that $1_{[\delta_i=0]} + 1_{[\delta_i=1/2]} + 1_{[\delta_i=1]} = 1$.

The case with $\delta_i = 1$ for $i = 1, 2$ is the regular case where Assumption A holds and hence standard techniques such as TSLS are valid, while the other eight cases refer to partial or complete unidentification (weak identification) of θ. We present this notion more formally in the next chapter where we also show that similar to the other tests in the literature, the new projection-type score test does not work when the inference is based on TSLS (interpreted as a GMM). Furthermore, Chaudhuri et al. (2007) showed that the USSIV estimator of θ is inconsistent under the weak instrument asymptotics and hence the standard Wald or score tests (in this case the USSIV score test) for $H : \theta_1 = \theta_{*1}$ can be over-sized.

However, the USSIV-framework is useful in the sense that Assumption S is satisfied under Assumption WI. To see this first note the properties of the score function that follow from [see (2.14) and (2.15)]. Let θ_* be such that $\sqrt{n}(\theta_* - \theta_0) = d_\theta$ for some finite real number $d_\theta = (d_1', d_2')'$ [where d_i is a $\nu_i \times 1$ vector for $i = 1, 2$]. Observe that conditional on sub-sample b,

(i) $\frac{1}{n}(y_{(a)} - X_{(a)}\theta_*)'N(Z_{(a)})(y_{(a)} - X_{(a)}\theta_*) \xrightarrow{P} c,$

(ii) $\frac{1}{n}\widehat{X}_{(ab)}'\widehat{X}_{(ab)} \xrightarrow{P} c\widehat{\Pi}_{(b)}'\Upsilon\widehat{\Pi}_{(b)}$ is positive definite [Okamoto (1973)],

(iii) $\frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_*) = \frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_0) + \left[\frac{1}{n}\nabla_{\theta\theta}Q_n(\bar{\theta})\right] d_\theta \xrightarrow{d} N\left(\sqrt{c}\widehat{\Pi}_{(b)}'\Upsilon\Pi d_\theta, c\widehat{\Pi}_{(b)}'\Upsilon\widehat{\Pi}_{(b)}\right)$
where $\bar{\theta}$ is such that $\sqrt{n}\|\bar{\theta} - \theta_0\| \leq \sqrt{n}\|\theta_* - \theta_0\| = \|d_\theta\|$ and hence

(iv) $\frac{1}{\sqrt{n}}\left[\nabla_1 Q_n(\theta_*) - \left(\widehat{X}_{(ab)1}'\widehat{X}_{(ab)2}\right)\left(\widehat{X}_{(ab)2}'\widehat{X}_{(ab)2}\right)^{-1}\nabla_2 Q_n(\theta_*)\right]$
$\xrightarrow{d} N\left(\sqrt{c}\widehat{\Pi}_{(b)1}'\Upsilon^{1/2}N\left(\Upsilon^{1/2'}\widehat{\Pi}_{(b)2}\right)\Upsilon^{1/2'}\Pi d_\theta, c\widehat{\Pi}_{(b)1}'\Upsilon^{1/2}N\left(\Upsilon^{1/2'}\widehat{\Pi}_{(b)2}\right)\Upsilon^{1/2'}\widehat{\Pi}_{(b)1}\right).$

The above observations suggest that Assumption A holds in the USSIV-framework conditional on sub-sample b and hence so does Assumption S. Therefore, the usual score statistic for jointly testing θ_1 and θ_2 takes the form

$$\mathcal{R}^{SS}(\theta) = \frac{(y_{(a)} - X_{(a)}\theta)'P(\widehat{X}_{(ab)})(y_{(a)} - X_{(a)}\theta)}{\frac{1}{n_a - k}(y_{(a)} - X_{(a)}\theta)'N(Z_{(a)})(y_{(a)} - X_{(a)}\theta)}. \tag{2.18}$$

This is same as the split-sample statistic considered by Staiger and Stock (1997) and Dufour and Jasiak (2001). From *(i)*, *(ii)* and *(iii)* it follows that $\mathcal{R}^{SS}(\theta_0) \xrightarrow{d} \chi^2_\nu$.[10] Hence the usual method of projection based on the score statistic rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ at level at most ϵ if

$$\inf_{\theta_2 \in \mathbb{R}^{\nu_2}} \mathcal{R}^{SS}(\theta_{*1}, \theta_2) > \chi^2_\nu(1 - \epsilon). \tag{2.19}$$

[10] The asymptotic test "based on generated regressors", proposed by Dufour and Jasiak (2001), rejects the null hypothesis $\theta = \theta_*$ at level ϵ if $\mathcal{R}^{SS}(\theta_{*1}, \theta_2) > \chi^2_\nu(1 - \epsilon)$.

On the other hand, from (2.2), *(i)*, *(ii)* and *(iv)*, it is clear that the efficient score statistic for θ_1, which takes the form

$$\mathcal{R}_1^{SS}(\theta) = \frac{(y_{(a)} - X_{(a)}\theta)' P\left(N(\widehat{X}_{(ab)2})\widehat{X}_{(ab)1}\right)(y_{(a)} - X_{(a)}\theta)}{\frac{1}{n_a - k}(y_{(a)} - X_{(a)}\theta)' N(Z_{(a)})(y_{(a)} - X_{(a)}\theta)}, \tag{2.20}$$

satisfies the property $\mathcal{R^{SS}}_1(\theta_0) \xrightarrow{d} \chi^2_{\nu_1}$. But, as shown by Chaudhuri et al. (2007), this result is not of much use in designing the usual score test (in this context, the USSIV score test) when $\delta_2 \neq 1$. This is because when θ_2 is not identified, it is not possible to obtain a consistent estimator of θ_2. In particular the estimator $\tilde{\theta}_{n2}(\theta_{01}) = \left(\widehat{X}'_{(ab)2} X_{(a)2}\right)^{-1} \widehat{X}'_{(ab)2}\left(y_{(a)} - X_{(a)}\theta_{01}\right)$, obtained by maximizing $Q_n(\theta_{01}, \theta_2)$ with respect to θ_2, is inconsistent. This is known to cause the upward size-distortion of the (USSIV) score test for $H_1 : \theta_1 = \theta_{*1}$, based on the USSIV-framework, when the unknown nuisance parameters θ_2 are not identified; the distortion is more severe when the corresponding regressors X_2 are highly endogenous.

However, Chaudhuri et al. (2007) also showed that the confidence region

$$\begin{aligned} \mathcal{C}_2(1-\zeta, \theta_{*1}) &= \left\{\theta_{*2} | \mathcal{R}_2^{SS\text{alt}}(\theta_*) \leq \chi^2_{\nu_2}(1-\zeta)\right\}, \text{ where} \\ \mathcal{R}_2^{SS\text{alt}}(\theta) &= \frac{(y_{(a)} - X_{(a)}\theta)' P\left(\widehat{X}_{(ab)2}\right)(y_{(a)} - X_{(a)}\theta)}{\frac{1}{n_a - k}(y_{(a)} - X_{(a)}\theta)' N(Z_{(a)})(y_{(a)} - X_{(a)}\theta)}, \end{aligned} \tag{2.21}$$

is such that – (i) it always has the correct coverage probability $1 - \zeta$, (ii) it is never empty, and (iii) only θ_{*2} in the $\sqrt{n}$-neighborhood of θ_{02} has non-zero probability of being contained in this region when $\delta_2 = 1$, i.e., when θ_2 is identified. Therefore, the new projection-type score test can be applied to test $H_1 : \theta_1 = \theta_{*1}$. The new test, as described before, rejects the null hypothesis when

$$\inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta, \theta_{*1})} \mathcal{R}_1^{SS}(\theta_*) > \chi^2_{\nu_1}(1-\epsilon). \tag{2.22}$$

In the next sub-section we present a Monte Carlo study of the finite sample behavior of the new projection-type score test described in (2.22) and show its superiority over the usual projection-type score test described in (2.19).

2.3.2 Finite-sample properties: Simulation study

The simulation study shows that - (i) the new projection-type score test is not as conservative as the usual projection-type score test, and (ii) when θ is identified, the finite-sample power of the new projection-type split-sample score test "almost" attains the "infeasible power envelope" provided by the finite-sample power of the infeasible score test, which rejects $H_1 : \theta_1 = \theta_{*1}$ if $\mathcal{R}_1^{SS}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$.

The data generating process (DGP) from the model in (2.11) is similar to that in Dufour and Taamouti (2005a). Our results are based on 10,000 replications and are also supported by a more extensive simulation study conducted by Chaudhuri et al. (2007). The DGP is described below.

(a) $\Sigma = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.3 \\ 0.8 & 0.3 & 1 \end{pmatrix}$, $\theta_{01} = 0.5$, $\theta_{02} = 1$, $n = 100$, $n_a = 75$ and $n_b = 25$.

(b) The first column of Z is an $n \times 1$ column of ones and the elements in the other $k-1$ columns are generated as i.i.d. $N(0,1)$ variables but are kept fixed over simulations. We report the results for $k = 4$ and $k = 10$ (for size-comparison only). The results are similar for other choices of k (not reported) that are not too large as compared to n_a and n_b.

(c) Π is constructed such that $\Pi = \mathbb{C}/\sqrt{n}$ where $\mathbb{C} = [\mathbb{C}_1, \mathbb{C}_2]$ and the elements of $\mathbb{C}_i$ are set at 0, 1.1547 and 20 when $\delta_i = 0$, 1/2 and 1 respectively for $i = 1, 2$. This satisfies the classification of "unidentification", "weak identification" and "strong identification" by Dufour and Taamouti (2005a).

The usual and new projection-type (split-sample) score tests never over-reject the true value of θ_1 even in finite-sample [see Table 2.1]. The results are similar even if we consider sample size as large as 10,000 with $n_1 = 7,500$ and $n_2 = 2,500$ [see Table 2.2].

Now let us turn to the finite-sample rejection rate of the false values of θ_1. This discussion is based on Figure 2.1.

It is clearly evident that the new method of projection is considerably less conservative than the usual method; for example, the rejection rate of the new test with $\zeta = 1\%, \epsilon = 5\%$ uniformly dominates the rejection rate of the usual projection-type split-sample score test with $\epsilon = 10\%$.

Regarding the choice of ζ and ϵ: the conservativeness of the new test decreases more rapidly when ϵ increases. Moreover, when θ_2 is strongly identified, the effect of the choice of ζ on the over all conservativeness of the new test seems to be negligible.

Validating the analytical discussion in this chapter, the simulations also provide ample evidence of the (local) asymptotic equivalence between the new projection-type (split-sample) score test and the infeasible split-sample score test when θ_1 is weakly identified or strongly identified.

Table 2.1: Rejection rates for $H: \theta_1 = \theta_{01}$ when $n_a = 75, n_b = 25$.

Identi-fication	for θ_2	$\delta_2 = 0$		$\delta_2 = \frac{1}{2}$		$\delta_2 = 1$	
for θ_1	Instruments	k=4	k=10	k=4	k=10	k=4	k=10
$\delta_1 = 0$	usual: 5%	0.0	0.0	0.7	0.7	1.6	1.4
	usual: 10%	0.0	0.0	2.1	1.7	3.5	3.5
	new: (4%+1%)	0.0	0.0	0.0	0.0	0.9	0.8
	new: (1%+4%)	0.0	0.0	0.0	0.0	3.4	3.3
	new: (1%+5%)	0.0	0.0	0.2	0.1	4.2	4.2
	new: (5%+5%)	0.0	0.0	0.5	0.4	4.4	4.5
	infeasible (5%)	5.4	5.5	5.3	5.4	5.5	5.4
$\delta_1 = \frac{1}{2}$	usual: 5%	0.0	0.0	0.8	0.7	2.0	1.5
	usual: 10%	0.2	0.1	2.1	1.8	3.9	3.4
	new: (4%+1%)	0.0	0.0	0.2	0.0	1.1	0.8
	new: (1%+4%)	0.0	0.0	0.2	0.0	3.8	3.2
	new: (1%+5%)	0.0	0.0	0.2	0.1	4.6	4.1
	new: (5%+5%)	0.0	0.0	1.0	0.4	4.8	4.3
	infeasible (5%)	5.4	5.5	5.4	5.4	5.7	5.2
$\delta_1 = 1$	usual: 5%	0.0	0.1	1.3	0.6	1.7	1.8
	usual: 10%	0.2	0.2	3.0	1.6	3.4	3.5
	new: (4%+1%)	0.0	0.0	0.2	0.1	1.0	1.1
	new: (1%+4%)	0.0	0.0	0.2	0.1	3.4	3.7
	new: (1%+5%)	0.0	0.0	0.2	0.1	4.2	4.5
	new: (5%+5%)	0.0	0.0	0.8	0.4	4.4	4.7
	infeasible (5%)	5.3	5.2	5.3	5.4	5.2	5.3

Table 2.2: Rejection rates for $H : \theta_1 = \theta_{01}$ when $n_a = 7500, n_b = 2500$.

Identi-fication	for θ_2	$\delta_2 = 0$		$\delta_2 = \frac{1}{2}$		$\delta_2 = 1$	
for θ_1	Instruments	k=4	k=10	k=4	k=10	k=4	k=10
$\delta_1 = 0$	usual: 5%	0.1	0.1	0.7	0.9	1.6	1.5
	usual: 10%	0.3	0.2	1.6	2.2	3.2	3.3
	new: (4%+1%)	0.0	0.0	0.0	0.1	0.9	0.8
	new: (1%+4%)	0.0	0.0	0.1	0.1	3.3	3.5
	new: (1%+5%)	0.0	0.0	0.1	0.1	4.0	4.5
	new: (5%+5%)	0.0	0.0	0.4	0.5	4.3	4.6
	infeasible (5%)	5.0	5.6	4.8	5.1	5.1	5.2
$\delta_1 = \frac{1}{2}$	usual: 5%	0.1	0.0	0.7	0.9	1.4	1.4
	usual: 10%	0.2	0.2	1.5	2.1	3.2	3.2
	new: (4%+1%)	0.0	0.0	0.0	0.0	0.9	0.8
	new: (1%+4%)	0.0	0.0	0.0	0.1	3.0	3.4
	new: (1%+5%)	0.0	0.0	0.0	0.1	3.8	4.2
	new: (5%+5%)	0.1	0.0	0.4	0.6	4.2	4.4
	infeasible (5%)	5.4	5.1	4.8	5.2	5.2	5.0
$\delta_1 = 1$	usual: 5%	0.1	0.1	0.6	1.1	1.6	1.5
	usual: 10%	0.2	0.1	1.5	2.8	3.3	3.2
	new: (4%+1%)	0.0	0.0	0.0	0.2	1.0	1.0
	new: (1%+4%)	0.0	0.0	0.1	0.1	3.3	3.4
	new: (1%+5%)	0.0	0.0	0.1	0.1	4.0	4.3
	new: (5%+5%)	0.0	0.0	0.3	0.6	4.2	4.5
	infeasible (5%)	5.1	4.8	4.7	4.9	5.0	5.0

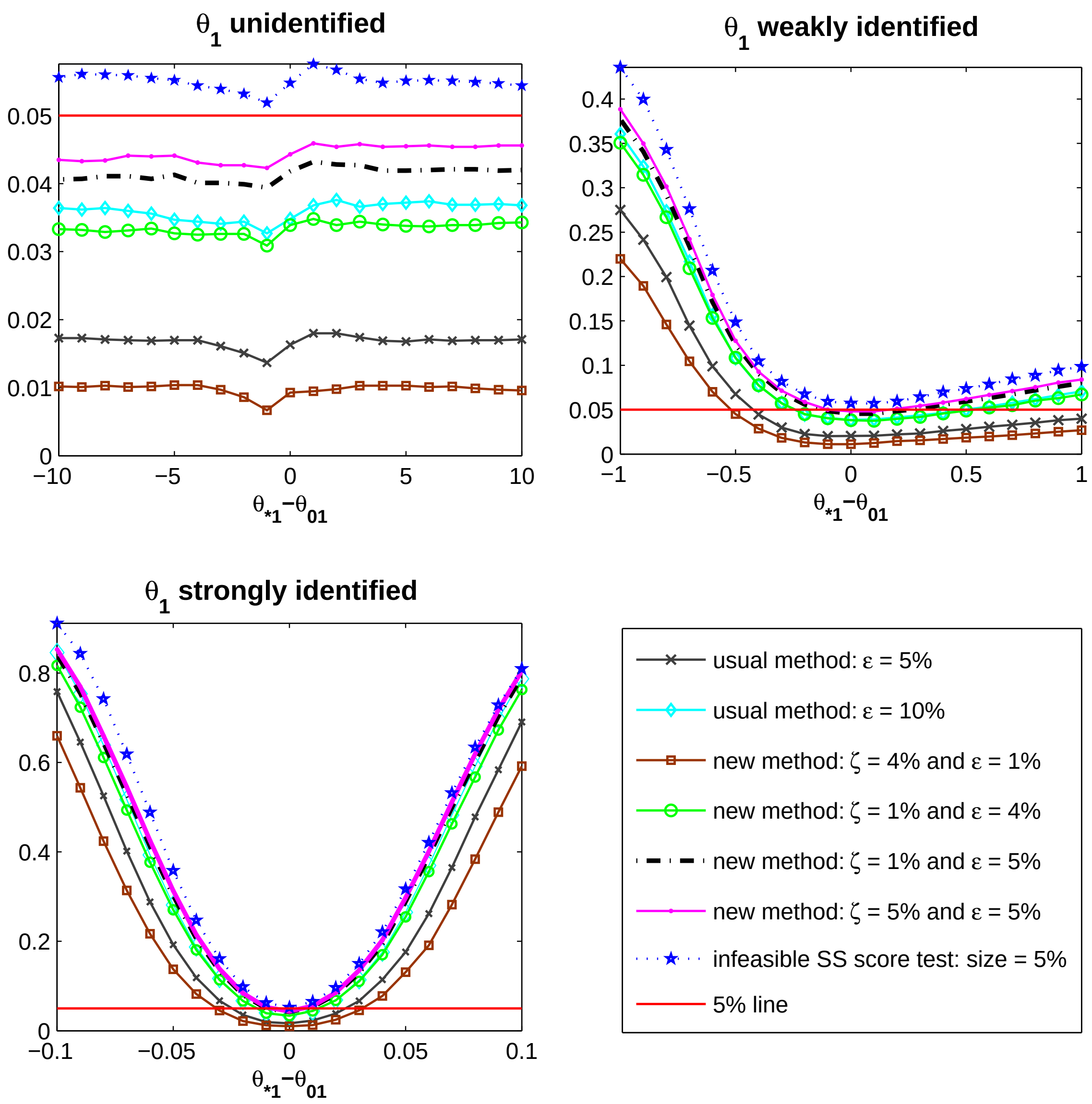

Figure 2.1: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when θ_2 is strongly identified, $n_a = 75, n_b = 25$ and $k = 4$.

Chapter 3

THE NEW PROJECTION-TYPE SCORE TEST IN GMM

In this chapter we describe the application of the new projection-type score test to a widely used sub-class of extremum estimators – the GMM estimators. The assumptions made here are more specific to the GMM framework than those of the last chapter.

The moment restrictions defined below can (but need not) be viewed as obtained from the first order condition of some optimization problem. Let $g : \Theta \times \mathcal{S} \mapsto \mathbb{R}^k$ be a measurable and twice-continuously differentiable function such that

$$\begin{aligned} Eg(z_t, \theta) &= 0 \text{ if } \theta = \theta_0, \\ &\neq 0 \text{ if } \theta \neq \theta_0. \end{aligned} \tag{3.1}$$

Equation (3.1) gives $k \geq \nu$ moment restrictions for inference on ν unknown elements of θ and is often referred to as the *global* identification condition. Let ν and k be fixed and finite numbers. We suppress the explicit dependence of the functionals on the observations in the rest of the discussion to avoid notational clutter; for example, $g_t(\theta)$ should be read as $g(z_t, \theta)$.

For simplicity, let $\theta_0 = (\theta_{01}', \theta_{02}')'$ be such that $\theta_{0i} \in$ interior(Θ_i) where Θ_i is a compact (ν_i-dimensional) subset of $\mathbb{R}^{\nu_i}$ for $i = 1, 2$. The parameter space $\Theta = \Theta_1 \times \Theta_2$ is a compact subset of $\mathbb{R}^{\nu}$ where $\nu = \nu_1 + \nu_2$.

A GMM estimator of θ is defined as $\widehat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\theta)$ where $Q_n(\theta)$ is some objective function taking the form:[1]

$$Q_n(\theta) = \frac{1}{2n}\left[\sum_{t=1}^{n} g_t(\theta)\right]' W_n(\theta_n)\left[\sum_{t=1}^{n} g_t(\theta)\right] \tag{3.2}$$

and $W_n(\theta_n)$ is some positive (semi)definite weighting matrix. In the usual two-step GMM, θ_n is some initial (possibly inefficient) estimator such that $\theta_n \xrightarrow{P} \theta_0$. The Continuous Updating GMM (CU-GMM) uses $W_n(\theta_n) = W_n(\theta)$ such that $W_n(\theta)$ converges uniformly to a positive definite matrix for $\theta \in \Theta$. Efficient estimation in both types results when, for all $\theta_n \xrightarrow{P} \theta_0$ and the weighting matrix $W_n(\theta_n) \xrightarrow{P} Var^{-1}\left[\lim_{n\to\infty} n^{-1/2}\sum_{t=1}^{n} g_t(\theta_0)\right]$ (provided the limit is positive definite).

In this chapter we show that under weak identification, in the sense of Stock and Wright (2000), the new method of projection based on Kleibergen's K-statistic offers a valid way of testing hypotheses of the form $H_1 : \theta_1 = \theta_{*1}$. Before describing the operational details, it will be helpful to introduce the weak identification framework and the related assumptions on the moment restrictions following Stock and Wright (2000), Guggenberger and Smith (2005) and Kleibergen (2005).

A word on our notations: for any random variable X such that $E\|X\| \leq \infty$, let $\overline{X} = X - EX$. Since all the matrices considered here are of finite dimension, we tend to be less scrupulous with mixing the notations like $\overline{X} = o_p(1)$ and $\|\overline{X}\| = o_p(1)$; both implying that every element of $\overline{X}$ converges in probability to zero.

---

[1] In continuation to the previous chapter, one could define $\widehat{\theta}_n = \arg\max_{\theta\in\Theta} -Q_n(\theta)$ where $Q_n(\theta)$ is as defined in (3.2). The multiplier 1/2 is for convenience.

3.1 Weak identification and other assumptions

High level assumptions on the moment restrictions and their first derivatives are summarized under Assumption M. These assumptions are more specific to the GMM setup than the generic ones mentioned under Assumption A in the last section.

Assumption M:

M1. There exists an open neighborhood $\mathcal{T} \subset \Theta$ containing θ_0 where

(i) $n^{-1}\sum_{t=1}^{n} g_t(\theta)$ is continuously differentiable almost surely

(ii) $\sup_{\theta\in\mathcal{T}} \|\nabla_\theta n^{-1}\sum_{t=1}^{n} g_t(\theta)\|$ is integrable.

M2. $\sup_{\theta\in\Theta} n^{-1}\sum_{t=1}^{n} \overline{g_t(\theta)} = o_p(1)$, $n^{-1}\sum_{t=1}^{n} \overline{\nabla_\theta g_t(\theta)} = o_p(1)$, $n^{-1}\sum_{t=1}^{n} \overline{\nabla_{\theta\theta} g_t(\theta)} = o_p(1)$ where $\nabla_\theta g_t(\theta) := \partial g_t(\theta)/\partial\theta'$ and $\nabla_{\theta\theta} g_t(\theta) := \partial vec\nabla_\theta g_t(\theta)/\partial\theta'$. $En^{-1}\sum_{t=1}^{n} \nabla_{\theta\theta} g_t(\theta)$ converges to some continuous and bounded function $L(\theta)$ for $\theta \in \Theta$.

M3. $n^{-1/2}\sum_{t=1}^{n} \begin{bmatrix} \overline{g_t(\theta_0)} \\ \overline{vec\nabla_\theta g_t(\theta_0)} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g \\ \Psi_\nabla \end{bmatrix} \sim N(0, V(\theta_0))$ and the asymptotic variance-covariance matrix $V(\theta) = ((V_{ab}(\theta)))_{a,b=g,\nabla}$ is bounded, continuous, symmetric and positive semi-definite. $V_{gg}(\theta)$ is symmetric, positive definite and differentiable with respect to $\theta \in \Theta$.

M4. There exist $\widehat{V}_{\nabla g}(\theta)$ and a symmetric positive definite matrix $\widehat{V}_{gg}(\theta)$ such that
$\widehat{V}_{\nabla g}(\theta) - V_{\nabla g}(\theta) = o_p(1)$ and $\partial vec\widehat{V}_{gg}(\theta)/\partial\theta' - \partial vec V_{gg}(\theta)/\partial\theta' = o_p(1)$ for $\theta \in \Theta$ and $\sup_{\theta\in\Theta} \widehat{V}_{gg}(\theta) - V_{gg}(\theta) = o_p(1)$.

The characterization of weak identification due to Stock and Wright (2000) is summarized under Assumption W.

Assumption W:
$E\frac{1}{n}\sum_{t=1}^{n} g_t(\theta_1,\theta_2) = \sum_{i=1}^{2}\left[1_{[\delta_i=1]}m_i(\theta_i) + 1_{[\delta_i=\frac{1}{2}]}n^{-1/2}\tilde{m}_{ni}(\theta_1,\theta_2)\right]$ where for $i = 1, 2$,

1. $m_i(\theta_{0i}) = 0, m_i(\theta_i) \neq 0$ for $\theta_i \neq \theta_{0i}$, $M_i(\theta_i) := \partial m_i(\theta_i)/\partial\theta_i'$ is continuous and $M_i(\theta_{0i})$ has full column rank.

2. $\tilde{m}_{ni}(\theta) \to \tilde{m}_i(\theta)$ uniformly in $\theta \in \Theta$, $\tilde{m}_i(\theta_0) = 0$ and $\tilde{m}_i(\theta)$ is continuous in θ and bounded on Θ. For $i, j = 1, 2$, $\tilde{M}_{n(i,j)}(\theta) := \partial\tilde{m}_{ni}(\theta)/\partial\theta_j'$ converges to some function $\tilde{M}_{(i,j)}(\theta)$.

Discussions on Assumptions M and W can be found in Stock and Wright (2000), Guggenberger and Smith (2005) and Kleibergen (2005).

The non-random indicator functions involving the δ's in Assumption W distinguish between the four cases of weak partial identification summarized in Table 3.1.[2]

Now note that Assumptions M1, M2 and W imply that for $\theta \in \mathcal{T}$ and for $i = 1, 2$,

$$\begin{aligned} J_{ni}(\theta) := E\frac{1}{n^{\delta_i}}\nabla_i\sum_{t=1}^{n} g_t(\theta) &= \frac{\partial}{\partial\theta_i'}E\frac{1}{n^{\delta_i}}\sum_{t=1}^{n} g_t(\theta) \\ &= 1_{[\delta_i=1]}M_i(\theta_i) + \sum_{j=1}^{2} 1_{[\delta_j=\frac{1}{2}]}\frac{n^{\delta_j}}{n^{\delta_i}}\tilde{M}_{n(j,i)}(\theta) \end{aligned} \quad (3.3)$$

where $\nabla_1 g_t(\theta)$ and $\nabla_2 g_t(\theta)$ are, respectively, the first ν_1 columns and last ν_2 columns of $\nabla_\theta g_t(\theta)$. Assumption W further implies that for $i, j = 1, 2$ (and $i \neq j$), $J_{ni}(\theta)$ is continuous in θ and

$$J_{ni}(\theta) \xrightarrow{P} J_i(\theta) := 1_{[\delta_i=1]}M_i(\theta_i) + 1_{[\delta_i=\frac{1}{2}]}\left[\tilde{M}_{(i,i)}(\theta) + 1_{[\delta_j=\frac{1}{2}]}\tilde{M}_{(j,i)}(\theta)\right],$$

[2] The values of the δ_i's are assigned the values $\frac{1}{2}$ and 1 because, as it can be seen from (3.3), n^{δ_i} will often be used as a suitable scaling factor.

which has full column rank for $\theta \in \theta_{0i} \times \Theta_j$ when θ_i is strongly identified. When both θ_1 and θ_2 are strongly identified, this implies the so-called *local* identification condition; i.e., the expected Jacobian is full column rank at $\theta = \theta_0$ [see equation (13) in Kleibergen (2005)].

Table 3.1: Four Cases of Weak-Partial-Identification.

	$\delta_2 = \frac{1}{2}$	$\delta_2 = 1$
$\delta_1 = \frac{1}{2}$	**WI-Case I** θ_1 : weakly identified θ_2 : weakly identified	**WI-Case II** θ_1 : weakly identified θ_2 : strongly identified
$\delta_1 = 1$	**WI-Case III** θ_1 : strongly identified θ_2 : weakly identified	**WI-Case IV** θ_1 : strongly identified θ_2 : strongly identified

In the context of nonlinear IV regression, WI-Case IV is the standard case where the usual Wald, LR (J) and score tests can be used to test $H_1 : \theta_1 = \theta_{*1}$; in WI-Case II, the K-test and the Generalized Empirical Likelihood based test [due to Guggenberger and Smith (2005)] can be used to test $H_1 : \theta_1 = \theta_{*1}$; finally in WI-Cases I and III, based on available research, only the projection-type tests have been shown to be valid for testing $H_1 : \theta_1 = \theta_{*1}$. We show that the new projection-type (score) test based on the K statistic test can be validly used under all four cases of partial identification. In addition, under WI-Cases II and IV, this test is shown to be asymptotically equivalent to the K-test against $\sqrt{n}$-local alternatives.

It is possible to relax some of our assumptions. For example, since we are only concerned with the asymptotic behavior of the tests against $\sqrt{n}$-local alternatives, it is sufficient if all the assumptions specified for $\theta \in \Theta$ hold in

$\mathcal{T}_1 \times \Theta_2$ where $\mathcal{T}_1 \subset \Theta_1$ is an open neighborhood containing θ_{01}.[3] Nevertheless, we made these simplifying assumptions to avoid the secondary details in the exposition which can obscure the basic idea behind the new projection-type test. Assumption M is stated in a somewhat unconventional form so that it can be directly applied to prove our results. However, these assumptions are not different in nature from those in Stock and Wright (2000), Guggenberger and Smith (2005) and Kleibergen (2005). For example, Assumption B in Stock and Wright (2000) states that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\overline{g_t(\theta)} \Rightarrow \xi(\theta)$ for $\theta \in \Theta$ where $\xi(\theta)$ is a mean-zero Gaussian stochastic process. By definition of weak convergence, this implies that $\sup_{\theta\in\Theta}\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\overline{g_t(\theta)}\| \xrightarrow{d} \sup_{\theta\in\Theta}\|\xi(\theta)\|$ and thus implies that $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\overline{g_t(\theta)} = o_p(1)$ [see Andrews (1994)]. Under this assumption, in order to show the consistency of the CU estimator of a strongly identified θ_2, restricted by a hypothesized $\sqrt{n}$-local θ_1, it is also required to assume that $\sup_{\theta\in\mathcal{T}_1\times\Theta_2}\|\xi(\theta)\| = O_p(1)$. Instead, we directly assume that $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\overline{g_t(\theta)} = o_p(1)$ (in Assumption M2) and that $V_{gg}(\theta)$ is bounded in Θ (in Assumption M3). Similarly, instead of specifying the form of $\widehat{V}_{gg}(\theta)$ and making assumptions such as Assumption M(ii) in Guggenberger and Smith (2005) to ensure its non-singularity for $\theta \in \mathcal{T}_1 \times \Theta_2$, we directly assume that there exists a matrix $\widehat{V}_{gg}(\theta)$ which is positive-definite along with its convergence result (in Assumption M4). Finally, unlike Assumptions M_θ (iii) and (vii) in Guggenberger and Smith (2005), our distributional assumption (in Assumption M3) is local in nature [similar to Kleibergen (2005)] and hence we need to specify assumptions (in Assumption M2) on the second derivative of the moment vector for studying the properties of the new projection-type test against $\sqrt{n}$-local alternatives. Assumption M1, which allows us to interchange the order of differentiation and

[3] In fact, to prove the validity of the new projection-type test and to show that under WI-Cases II and IV it is asymptotically equivalent to the infeasible efficient score test (described after Lemma 3.2 and before Theorem 3.3) against $\sqrt{n}$-local alternatives, it is sufficient if the above assumptions hold in $\mathcal{T} \subset \Theta$, an open neighborhood containing θ_0.

integration, is made for simplicity.

3.2 Problem with the usual GMM-score test

Kleibergen (2005) pointed out that the failure of the usual score test, based on the (efficient) two-step GMM, is due to the (asymptotic) non-zero correlation between the estimator of the expected Jacobian and the moment vector (both under suitable scaling). To see this, assume for now that there are no nuisance parameters θ_2 and hence $\theta = \theta_1$. Therefore, the only two relevant cases are WI-Case IV – where the parameters are strongly identified, and WI-Case II – where the parameters are weakly identified. The gradient of the two-step GMM objective function with respect to θ is given by

$$\nabla_\theta Q_n(\theta) := \frac{\partial Q_n(\theta)}{\partial \theta'} = \frac{1}{n}\left[\sum_{t=1}^{n} g_t(\theta)\right]' W_n(\theta_n)\left[\sum_{t=1}^{n} \nabla_\theta g_t(\theta)\right] \tag{3.4}$$

where $\theta_n \xrightarrow{P} \theta_0$ is some initial first-stage estimator and $W_n(\theta)$ converges in probability to some positive definite matrix $W(\theta)$. In WI-Case IV, using (3.3), Assumptions M2 and M3, and scaling (3.4) by $n^{-1/2}$ give

$$\frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_0) = \left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_t(\theta_0)\right]' W_n(\theta_n)\left[\frac{1}{n}\sum_{t=1}^{n} \nabla_\theta g_t(\theta_0)\right] \xrightarrow{d} \Psi_g' W(\theta_0) M_1(\theta_0).$$

Thus there is no problem in constructing a score statistic based on $n^{-1/2}\nabla_\theta Q_n(\theta_0)$ (compare with the sufficient conditions stated under Assumption S in the last section). However, in WI-Case II the above scaling leads to degenerate results; only scaling by 1 results in tractable asymptotic quantities. Such a scaling of (3.4) gives

$$\nabla_\theta Q_n(\theta_0) = \left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_t(\theta_0)\right]' W_n(\theta_n)\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \nabla_\theta g_t(\theta_0)\right] \xrightarrow{d} \Psi_g' W(\theta_0) \Psi_\nabla$$

and using Assumption M3, it follows that the usual techniques of constructing a score statistic are not going to work unless $V_{\nabla g} = 0$.

Kleibergen (2005) noted that the CU-GMM offers a natural way of constructing the estimator of the expected Jacobian such that even under weak identification, the (scaled) estimator of the expected Jacobian and the moment vector are asymptotically uncorrelated (and under Assumption M3, asymptotically independent). Kleibergen's K statistic is a quadratic form in the gradient of the CU-GMM objective function; and hence our discussion of the new projection-type test is based on the CU-GMM objective function.

3.3 The new projection-type test based on the K-statistic

Using Assumptions M3 and M4, the general form in (3.2) can be modified to construct the CU-GMM objective function as

$$Q_n(\theta) = \frac{1}{2n}\left[\sum_{t=1}^{n} g_t(\theta)\right]' \widehat{V}_{gg}^{-1}(\theta)\left[\sum_{t=1}^{n} g_t(\theta)\right].$$

The gradient of the CU-GMM objective function with respect to θ is given by [4]

$$\nabla_\theta Q_n(\theta) := \frac{\partial Q_n(\theta)}{\partial \theta'} = \frac{1}{n} g_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_T(\theta) \quad (3.5)$$

$$\text{where } \; g_T(\theta) = \sum_{t=1}^{n} g_t(\theta),\; \widehat{D}_T(\theta) = \sum_{t=1}^{n} \widehat{D}_t(\theta),$$

$$\text{and } \; \widehat{D}_t(\theta) = devec_k\left[vec\nabla_\theta g_t(\theta) - \widehat{V}_{\nabla g}(\theta)\widehat{V}_{gg}^{-1}(\theta)g_t(\theta)\right].$$

The CUE (CU-GMM estimator) $\widehat{\theta}_n$ of $\theta \in \Theta$ satisfies the first-order condition

$$\nabla_\theta Q_n(\widehat{\theta}_n) = \left[\nabla_1 Q_n(\widehat{\theta}_n), \nabla_2 Q_n(\widehat{\theta}_n)\right] = \frac{1}{n} g_T'(\widehat{\theta}_n)\widehat{V}_{gg}^{-1}(\widehat{\theta}_n)\widehat{D}_T(\widehat{\theta}_n) = 0$$

[4] See Kleibergen (2005) for details.

w.p.a.1. Similarly, under the restriction that $\theta_1 = \theta_{*1}$, the CUE $\tilde{\theta}_{n2}(\theta_{*1})$ of θ_2 minimizes $Q_n(\theta_{*1}, \theta_2)$ with respect to $\theta_2 \in \Theta_2$ and hence, w.p.a.1, $\tilde{\theta}_* = (\theta'_{*1}, \tilde{\theta}'_{n2}(\theta_{*1}))'$ satisfies the first-order condition

$$\nabla_2 Q_n(\tilde{\theta}_*) = \frac{1}{n} g'_T(\tilde{\theta}_*) \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \widehat{D}_{T2}(\tilde{\theta}_*) = 0 \tag{3.6}$$

where $\widehat{D}_{Ti}(\theta) = \sum_{t=1}^{n} devec_k \left[vec \nabla_i g_t(\theta) - \widehat{V}_{ig}(\theta) \widehat{V}_{gg}^{-1}(\theta) g_t(\theta) \right]$ for $i = 1, 2$. The above expression uses the partition $\widehat{D}_T(\theta) = \left[\widehat{D}_{T1}(\theta), \widehat{D}_{T2}(\theta) \right]$ and $\widehat{V}_{\nabla g}(\theta) = [\widehat{V}'_{1g}(\theta), \widehat{V}'_{2g}(\theta)]'$ with respect to θ_1 and θ_2.

Lemma 3.1 *Let $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta$ where $d_1 \in \mathbb{R}^{\nu_1}$. Then under Assumptions M and W, $\sqrt{n}(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}) = O_p(1)$ in WI-Cases II and IV.*

Lemma 3.1 follows directly from Lemma A1 in Stock and Wright (2000) which shows the $\sqrt{n}$-consistency of the unconstrained estimator of θ_2 under WI-Case II.

Kleibergen's K-test rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $K_n(\tilde{\theta}_*) > \chi^2_{\nu_1}(1-\epsilon)$ where the K-statistic is defined by Kleibergen (2005) as

$$K_n(\theta) = n \left(\nabla_\theta Q_n(\theta) \right) \left[\widehat{D}'_T(\theta) \widehat{V}_{gg}^{-1}(\theta) \widehat{D}_T(\theta) \right]^{-1} \left(\nabla_\theta Q_n(\theta) \right)'. \tag{3.7}$$

In WI-Cases II and IV and under Assumptions M and W, $K_n(\theta_{01}, \tilde{\theta}_{n2}(\theta_{01})) \xrightarrow{d} \chi^2_{\nu_1}$. See Theorem 2 in Kleibergen (2005) for the proof under presumably weaker conditions. The limiting $\chi^2_{\nu_1}$ distribution of the K-statistic $K_n(\theta_{01}, \tilde{\theta}_{n2}(\theta_{01}))$ in Kleibergen's proof crucially depends on the $\sqrt{n}$-consistency of $\tilde{\theta}_{n2}(\theta_{01})$.

However, under WI-Cases I and III, $\tilde{\theta}_{n2}(\theta_{01})$ is inconsistent and the properties of the K-statistic (and hence the K-test) are an area of current research. It is in these two cases where the literature recommends the use of projection techniques. Hence our test, which is a projection-type test based on the K-statistic [and described in Theorem 3.3], is likely to be most useful in these two cases because it is generally less conservative than the usual projection-type tests and at the same time it avoids an uncontrolled over-rejection of the true parameter values.

Notice that the form of the K-statistic in (3.7) satisfies the sufficient conditions mentioned under Assumption S (in Section 2). Hence we can follow the same principle, as described in the last section, to construct the new projection-type test based on the K-statistic. To do that, define the estimated efficient score version for θ_1 as

$$\nabla_{1.2}Q_n(\theta) = \frac{1}{n} g_T'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta).$$

Since $\nabla_2 Q_n(\tilde{\theta}_*) = 0$ w.p.a.1 [see (3.6)], it becomes obvious that $K(\tilde{\theta}_*)$ is a (normalized) quadratic form of the estimated efficient score of θ_1 evaluated at $\tilde{\theta}_*$, once we note that the top-left $\nu_1 \times \nu_1$ block of $\left[\widehat{D}_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_T(\theta)\right]^{-1}$ is $\left[\widehat{D}_{T1}'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta)\right]^{-1}$. Finally, similar to the efficient score statistic $R_1(\theta)$ in the previous section, define the efficient score version of the K-statistic (or the efficient K-statistic) as

$$K_{n1}(\theta) =$$
$$n\left(\nabla_{1.2}Q_n(\theta)\right)\left(\widehat{D}_{T1}'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta)\right)^{-1}\left(\nabla_{1.2}Q_n(\theta)\right)'.$$

Lemma 3.2 *Let $\theta_{ni} = \theta_{0i} + d_i/\sqrt{n} \in \Theta$ such that for $i = 1, 2$, $d_i \in \mathbb{R}^{\nu_i}$ is fixed; and let $\theta_n = (\theta_{n1}', \theta_{n2}')'$ and $d_\theta := (d_1', d_2')'$. Let $L(\theta)$ and $\Psi_{\nabla.g}$ be partitioned with respect to θ_1 and θ_2 such that $L(\theta) = [L_1'(\theta), L_2'(\theta)]'$ and $\Psi_{\nabla.g} = \left[\Psi_{1.g}', \Psi_{2.g}'\right]'$. Then under Assumptions M and W,*

(i) $K_{n1}(\theta_n) \xrightarrow{d} \mathbb{B}'P\left(N\left(\mathbb{A}_2\right)\mathbb{A}_1\right)\mathbb{B}$

(ii) $\tilde{K}_{n2}(\theta_n) := n\left(\nabla_2 Q_n(\theta_n)\right)\left(\widehat{D}_{T2}'(\theta_n)\widehat{V}_{gg}^{-1}(\theta_n)\widehat{D}_{T2}(\theta_n)\right)^{-1}\left(\nabla_2 Q_n(\theta_n)\right)' \xrightarrow{d} \mathbb{B}'P\left(\mathbb{A}_2\right)\mathbb{B}$

(iii) $S_n(\theta_n) := 2Q_n(\theta_n) \xrightarrow{d} \mathbb{B}'\mathbb{B}$

where $\mathbb{A}_i = V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[J_i(\theta_0) + (1 - 1_{[\delta_i=1]})devec_k\left[\Psi_{i.g} + L_i(\theta_0)d_\theta\right]\right]$ *for* $i = 1, 2$ *and* $\mathbb{B} = V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[\Psi_g + \sum_{i=1}^{2} 1_{[\delta_i=1]}M_i(\theta_{0i})d_i\right]$.

Lemma 3.2 is proved in the Appendix. Several remarks are in order here.

(i) It follows from Lemma 3.2 that $K_{n1}(\theta_0) \xrightarrow{d} \chi^2_{\nu_1}$ in WI-Cases I-IV. Further using Lemma 3.1, it follows that $K_n(\theta_{n1}, \tilde{\theta}_{n2}(\theta_{n1})) = K_{n1}(\theta_n) + o_p(1) = K_{n1}(\theta_{n1}, \theta_{02}) + o_p(1)$ in WI-Cases II and IV.

(ii) $\tilde{K}_{n2}(\theta_{*1}, \theta_{*2})$ is the K-statistic for testing $H_2 : \theta_2 = \theta_{*2}$ when θ_1 is assumed to be equal to θ_{*1} (and hence no longer considered an unknown parameter). Note that $\tilde{K}_{n2}(\theta_0) \xrightarrow{d} \chi^2_{\nu_2}$, i.e. if the true value of θ_1 is known *a priori* then the test which rejects $H_2 : \theta_2 = \theta_{*2}$ if $\tilde{K}_{n2}(\theta_{01}, \theta_{*2}) > \chi^2_{\nu_2}(1 - \zeta)$ has asymptotic size $1 - \zeta$. In WI-Cases II and IV, $\tilde{K}_{n2}(\theta_n)$ converges to a non-central $\chi^2_{\nu_2}$ distribution with non-centrality parameter

$$\left[\sum_{i=1}^{2} 1_{[\delta_i=1]}M_i(\theta_{0i})d_i\right]' V_{gg}^{-\frac{1}{2}}(\theta_0)P\left(V_{gg}^{-\frac{1}{2}'}(\theta_0)J_2(\theta_0)\right)V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[\sum_{i=1}^{2} 1_{[\delta_i=1]}M_i(\theta_{0i})d_i\right]$$

which, under (3.1), can be finite only in the $\sqrt{n}$-neighborhood of θ_{02}.

(iii) $S_n(\theta_n)$ is the S-statistic proposed by Stock and Wright (2000). In WI-Cases II and IV, $S_n(\theta_n)$ converges to a non-central χ^2_k distribution with non-centrality parameter

$$\left[\sum_{i=1}^{2} 1_{[\delta_i=1]}M_i(\theta_{0i})d_i\right]' V_{gg}^{-1}(\theta_0)\left[\sum_{i=1}^{2} 1_{[\delta_i=1]}M_i(\theta_{0i})d_i\right]$$

which, under (3.1), can be finite only in the $\sqrt{n}$-neighborhood of θ_{02}. When θ_{01} is known *a priori*, then the test which rejects $H_2 : \theta_2 = \theta_{*2}$ if $S_n(\theta_{01}, \theta_{*2}) > \chi^2_k(1 - \zeta)$ has asymptotic size $1 - \zeta$.

From the above discussion it should be clear that in WI-Cases II and IV, and against $\sqrt{n}$-local alternatives, the level-ϵ K-test for $H_1 : \theta_1 = \theta_{*1}$ is asymptotically equivalent to the infeasible efficient K-test which rejects $H_1 : \theta_1 = \theta_{*1}$ at level ϵ when $K_{n1}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$. The latter test uses the unknown true value of the nuisance parameter θ_2 and hence is infeasible. Based on these observations, we define and describe the new projection-type score (K) test for the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ in Theorem 3.3.

Theorem 3.3 *Let* $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta$ *where* $d_1 \in \mathbb{R}^{\nu_1}$. *Define* $\mathcal{C}_2(1-\zeta, \theta_{*1}) := \{\theta_{*2} : S_n(\theta_{*1}, \theta_{*2}) \leq \chi^2_k(1-\zeta)\}$. *Define the random variable* $\phi_n(\theta_{*1}) := \phi_n\left(\theta_{*1}, \{z_t\}_{t=1}^n\right)$ *as*

$$\phi_n(\theta_{*1}) = \begin{cases} 1 \text{ \textit{if} } \mathcal{C}_2(1-\zeta,\theta_{*1}) = \varnothing \text{ \textit{or} } \inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} K_{n1}(\theta_{*1},\theta_{*2}) > \chi^2_{n_1}(1-\epsilon) \\ 0 \text{ \textit{otherwise}} \end{cases}$$

The new projection-type test based on the K-statistic rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ *if* $\phi_n(\theta_{*1}) = 1$. *Under Assumptions M and W,*

(i) $\lim_{n\to\infty} E_{\theta_{01}}\phi_n(\theta_{01}) \leq \epsilon + \zeta$*;*

(ii) if in WI-Cases II and IV, $\mathcal{C}_2(1-\zeta, \theta_{*1})$ *is nonempty w.p.a.1, then*

$$\lim_{n\to\infty}\left[E_{\theta_{01}}\phi_n(\theta_{*1}) - Pr_{\theta_{01}}\left[K_{n1}(\theta_{*1},\theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)\right]\right] = 0.$$

Theorem 3.3 describes the new projection-type test based on the K-statistic. The size of this test is bounded from above by $\epsilon + \zeta$. Under WI-Cases I and III, ϵ and ζ can be chosen such that the desired level of the test is not exceeded. Under WI-Cases II and IV, the choice of ζ becomes asymptotically irrelevant if the first-stage confidence set for θ_{02} is nonempty w.p.a.1; then the new projection-type test against $\sqrt{n}$-local alternatives is asymptotically equivalent to the infeasible efficient K-test rejecting $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $K_{n1}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$.

It is straightforward to see that the new projection-type test can be inverted to obtain a confidence region for θ_1 as

$$\begin{aligned} &\{\theta_{*1} : \phi_n(\theta_{*1}) = 0\} \\ &= \left\{\theta_{*1} : \mathcal{C}_2(1-\zeta, \theta_{*1}) \neq \varnothing, \inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2}) \leq \chi^2_{\nu_1}(1-\epsilon)\right\}. \end{aligned}$$

This is a conservative $(1-\epsilon-\zeta)$-level uniform asymptotic confidence region for θ_1; the uniformity follows from the fact that we are considering the infimum of the efficient K-statistic. In WI-Cases II and IV, the region's coverage (and length) is asymptotically equivalent to the asymptotic coverage (and length) of the infeasible region $\left\{\theta_{*1} : K_{n1}(\theta_{*1}, \theta_{02}) \leq \chi^2_{\nu_1}(1-\epsilon)\right\}$ if $\mathcal{C}_2(1-\zeta, \theta_1) \neq \varnothing$ w.p.a.1 for $\theta_1 = \theta_{01}$ (for $\theta_1 \in \Theta_1$). The asymptotic equivalence under WI-Cases II and IV naturally extends to the K-test for subsets of parameters (see the discussion preceding Theorem 3.3).

Another choice for the first-stage confidence set $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is the region $\mathcal{C}_2^K(1-\zeta, \theta_{*1}) = \left\{\theta_{*2} : \tilde{K}_{n2}(\theta_{*1}, \theta_{*2}) \leq \chi^2_{\nu_2}(1-\zeta)\right\}$ [see Lemma 3.2(ii)]. By definition, $\tilde{K}_{n2}(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) = 0$ and such a first-stage confidence set for θ_2 is nonempty w.p.a.1. This further implies that

$$K_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) \geq \inf_{\theta_{*2} \in \mathcal{C}_2^K(1-\zeta,\theta_{*1})} K_{n1}(\theta_*)$$

and hence the power (size) of the K-test for $H_1 : \theta_1 = \theta_{*1}$ dominates the power (size) of the new projection-type test when this region is used in the first step. However, using $\mathcal{C}_2(1-\epsilon, \theta_{*1})$ in the statement of Theorem 3.3 has a major advantage. The underlying S-test concurrently tests the model-specification (i.e. $Eg_t(\theta_{*1}, \theta_{*2}) = 0$) and thus, under equation (3.1), rules out the spurious decline of power at all local-minima and saddle points (unless present in the $\sqrt{n}$-neighborhood of θ_0) of the objective function which is typical to tests based on the K-statistic. As a consequence, unlike Kleibergen (2005), a separate pre-testing for misspecification is not required to avoid the spurious decline in power of the new projection-type test.

It is important to note that since $\mathcal{C}_2(1-\zeta,\theta_{*1})$, in the statement of Theorem 3.3, can be empty with positive probability, the asymptotic equivalence in WI-Cases II and IV may not hold – the new projection-type test can be more powerful than the infeasible efficient K-test and the K-test and the asymptotic size of the new test may belong to the interval $(\epsilon, \epsilon+\zeta)$. The Monte-Carlo experiment in the next section reveals this fact.

The usual projection-type test for $H_1 : \theta_1 = \theta_{*1}$ based on the K-statistic rejects the null hypothesis at level ϵ if $\inf_{\theta_{*2}\in\Theta_2} K_n(\theta_{*1},\theta_{*2}) > \chi^2_\nu(1-\epsilon)$. The K-test for $H_1 : \theta_1 = \theta_{*1}$ is at least as powerful as the usual projection-type test [see Lemma A.6 in the Appendix]. Similar conclusions can be expected when comparing the new projection-type and the usual projection-type tests based on the K-statistic in WI-Cases II and IV. All the tests discussed in this section can be validly applied even under complete unidentification [see Phillips (1989) and Kleibergen (2005)]; of course none of them will have any power against the alternatives.

Finally it should be noted that, if it is possible to obtain a $\sqrt{n}$-consistent point estimator for a subset of the nuisance parameters, the computational cost of the new projection-type test can be reduced significantly by using this estimator and restricting the search for the infimum of the efficient K-statistic to the remaining nuisance parameters only.

3.4 *Application to a linear IV regression*

In this section we do a Monte Carlo study and show that the asymptotic theory of the last section provides a good approximation to the finite-sample behavior of the new projection-type test in a linear Simultaneous Equations model. We briefly discuss how the general theory simplifies in this setup before describing our simulation results.

3.4.1 *Simplifications under this framework:*

Our framework is similar to that of Kleibergen (2004) and Zivot et al. (2006). Consider the linear IV model (2.11) in Section 2.3, of course without splitting the sample. Then the different quantities from the general exposition in the GMM setup translate into

(i) $g_t(\theta) = Z_t'(y_t - X_t\theta) = Z_t'Z_t\Pi(\theta_0 - \theta) + Z_t'(u_t + \eta_t(\theta_0 - \theta))$ and $E[g_t(\theta)] = Z_t'Z_t\Pi(\theta_0 - \theta)$,

(ii) $\triangledown_\theta g_t(\theta) = -Z_t'Z_t\Pi - Z_t'\eta_t$, $E[\triangledown_\theta g_t(\theta)] = -Z_t'Z_t\Pi$,

(iii) $\triangledown_{\theta\theta} g_t(\theta) = 0$ and $L(\theta) = 0$,

(iv) $V(\theta) = \begin{bmatrix} V_{gg}(\theta) & V_{g\triangledown}(\theta) \\ V_{\triangledown g}(\theta) & V_{\triangledown\triangledown}(\theta) \end{bmatrix} = (\Omega'\Sigma\Omega) \otimes \Upsilon$ where $\Omega = \begin{bmatrix} 1 & 0_{1\times\nu} \\ \theta - \theta_0 & I_\nu \end{bmatrix}$ and $\Upsilon = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} Z_t'Z_t$ is assumed to be a finite, symmetric, positive definite matrix.

The conditions for Assumption M can be enforced simply by assuming that (a) $\theta_0 \in$ interior(Θ) where Θ is compact, (b) the sample moments involving u_t, η_{1t} and η_{2t} converge in probability to their expectations, and (c) $n^{-1/2}Z'(u, \eta_1, \eta_2) \xrightarrow{d} \Upsilon^{\frac{1}{2}'}(\xi_{Zu}, \xi_{Z1}, \xi_{Z2})$ where $vec(\xi_{Zu}, \xi_{Z1}, \xi_{Z2}) \sim N(0, \Sigma \otimes I_k)$ [see pages 1070–1071 in Stock and Wright (2000)].

Thus the other relevant quantities translate into

(v) $\Psi_g = \Upsilon^{\frac{1}{2}'}[\xi_{Zu} + \xi_{Z1}(\theta_{01} - \theta_1) + \xi_{Z2}(\theta_{02} - \theta_2)]$, $\Psi_\nabla = -\Upsilon^{\frac{1}{2}'}[\Psi_{Z1}, \Psi_{Z2}]$,

(vi) $\widehat{V}(\theta) = n^{-1}[y - X\theta, -X]' N_Z [y - X\theta, -X] \otimes \frac{1}{n} Z'Z$ and

(vii) $\widehat{D}_{Ti} = -Z'X_i + Z'(y - X\theta)\frac{(y-X\theta)N_Z X_i}{(y-X\theta)'N_Z(y-X\theta)}$ for $i = 1, 2$.

Finally, note from (i) and (iv) that the moment restrictions in (3.1) are valid as long as Π is full column rank. Hence we can recast Assumption W in this scenario by modeling weak identification using the weak instrument framework of Staiger and Stock (1997). In particular, we assume that $\Pi_i = 1_{[\delta_i=1]}\mathbb{C}_i + 1_{[\delta_i=\frac{1}{2}]}\frac{\mathbb{C}_i}{\sqrt{n}}$ where for $i = 1, 2$, $\mathbb{C}_i$ is a $k \times \nu_i$ matrix of fixed and bounded elements such that $\mathbb{C} = [\mathbb{C}_1, \mathbb{C}_2]$ is full column rank [similar to Section 2.3].

Therefore, the expectation of the average moment vector and its first derivative are

(viii) $E[n^{-1}g_T(\theta)] = n^{-1}Z'Z \sum_{i=1}^{2}[1_{[\delta_i=1]}\mathbb{C}_i + 1_{[\delta_i=\frac{1}{2}]}\frac{\mathbb{C}_i}{\sqrt{n}}](\theta_{0i} - \theta_i)$

(ix) $E[n^{-1}\nabla_\theta g_T(\theta)] = -n^{-1}Z'Z \sum_{i=1}^{2}[1_{[\delta_i=1]}\mathbb{C}_i + 1_{[\delta_i=\frac{1}{2}]}\frac{\mathbb{C}_i}{\sqrt{n}}]$ implying that

(x) $J_{ni} := E[n^{-\delta_i}\nabla_i g_T(\theta)] = -n^{-1}Z'Z\mathbb{C}_i$ and $J_i(\theta) = -\Upsilon\mathbb{C}_i$ for $i = 1, 2$.

It is apparent that the specific structure of the moment restrictions allows for substantial simplification in this section. Another simplification comes when finding the CUE $\tilde{\theta}_{n2}(\theta_{*1})$ of the nuisance parameters θ_2; because of the block-diagonal covariance matrix of the structural errors, the minimization of $Q_n(\theta_{*1}, \theta_2)$ with respect to θ_2 boils down to an eigen-value problem and this significantly reduces the computational cost of Kleibergen's K-test.[5] To highlight these simplifications, it is helpful to restate the forms of the relevant test

[5]The simplification assumes that $\Theta_2 = \mathbb{R}^{\nu_2}$ and hence relaxes the compactness assumption. However, for all practical purposes relaxing the boundedness assumption is not going to alter the results in this linear model. Alternatively one can also impose a condition similar to Assumption D in chapter 4 of Amemiya (1985).

statistics using the construction of Moreira (2003). This directly fits into the framework discussed previously under Assumption S [Chapter 2].

The CUE of θ minimizes the objective function:

$$Q_n(\theta) = \frac{1}{2n} g_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)g_T(\theta) = \frac{n}{2}\frac{(y - X\theta)'P(Z)(y - X\theta)}{(y - X\theta)'N(Z)(y - X\theta)}.$$

After suitable standardization with respect to $\widehat{V}_{gg}(\theta)$ and hence defining

$$h_n(\theta_1, \theta_2) = n^{-1/2}V_{gg}^{-\frac{1}{2}'}(\theta)g_T(\theta) = \frac{(Z'Z)^{-\frac{1}{2}'}Z'(y - X_1\theta_1 - X_2\theta_2)}{\sqrt{\frac{1}{n}(y - X\theta)'N(Z)(y - X\theta)}} \text{ and}$$

$H_n'(\theta) = n^{-1/2}V_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_T(\theta) = [H_{n1}'(\theta), H_{n2}'(\theta)]'$ where for $i = 1, 2$,

$$H_{ni}'(\theta_1, \theta_2) = n^{-1/2}V_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{Ti}(\theta)$$
$$= \frac{(Z'Z)^{-\frac{1}{2}'}Z'}{\sqrt{\frac{1}{n}(y - X\theta)'N(Z)(y - X\theta)}}\left[X_i - (y - X_1\theta_1 - X_2\theta_2)\frac{(y - X\theta)'N(Z)X_i}{(y - X\theta)'N(Z)(y - X\theta)}\right],$$

we note that $Q_n(\theta) = \frac{1}{2}h_n'(\theta)h_n(\theta)$ and $\nabla_\theta Q_n(\theta) = [\nabla_1 Q_n(\theta), \nabla_2 Q_n(\theta)] = H_n(\theta)h_n(\theta)$.

We compare the finite-sample performance of the new projection-type score (K) test against that of the projection-type test based on the AR(S) statistic and the original K-test.

The projection-type test based on the AR statistic rejects $H_1 : \theta_1 = \theta_{*1}$ at level at most ϵ if $\inf_{\theta_{*2}\in\Theta_2} AR(\theta_*) > \chi_k^2(1 - \epsilon)$ where $AR(\theta) \equiv S(\theta) := 2Q_n(\theta) = h_n'(\theta)h_n(\theta)$. See Dufour (1997), Staiger and Stock (1997), Stock and Wright (2000) and Dufour and Taamouti (2005b,a) for details.

The K-test rejects $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $K_n(\tilde{\theta}_*) > \chi_{\nu_1}^2(1 - \epsilon)$ where $\tilde{\theta}_* = (\theta_{*1}', \tilde{\theta}_{n2}'(\theta_{*1}))'$ and $K_n(\theta) := h_n'(\theta)P_{H_n(\theta)}h_n(\theta)$. See Kleibergen (2004, 2005) for details.

The new projection-type test based on the K-statistic rejects $H_1 : \theta_1 = \theta_{*1}$ at level at most $\epsilon + \zeta$ if

$$\mathcal{C}_2(1-\zeta, \theta_{*1}) = \varnothing \text{ or } \inf_{\theta_2 \in \mathcal{C}_2(1-\zeta,\theta_{*1})} K_{n1}(\theta_*) > \chi^2_{\nu_1}(1-\epsilon)$$

where $\mathcal{C}_2(1-\zeta, \theta_{*1}) := \left\{\theta_{*2} : AR(\theta_*) > \chi^2_k(1-\zeta)\right\}$

and $K_{n1}(\theta) := h_n(\theta) N_{H_{n2}(\theta)} H_{n1}(\theta) \left[H_{n1}(\theta) N_{H_{n2}(\theta)} H_{n1}(\theta)\right]^{-1} H_{n1}(\theta) N_{H_{n2}(\theta)} h_n(\theta).$

3.4.2 Finite-sample properties: Simulation study

The true values of the structural coefficients are arbitrarily taken as $\theta_{01} = 1$ and $\theta_{02} = 10$. We take the sample size $n = 100$.

The structural errors $[u, \eta_1, \eta_2]$ are generated by drawing n independent random samples from $N_3(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 1 & \rho_{u1} & \rho_{u2} \\ \rho_{1u} & 1 & 0 \\ \rho_{2u} & 0 & 1 \end{pmatrix}. \tag{3.8}$$

If η_1 and η_2 are correlated, the level of endogeneity of the regressor X_1 depends on the correlations between [η_1 and u], [η_1 and η_2] and [η_2 and u]. Our choice of Σ in (3.8) simplifies the set-up by ensuring that the level of endogeneity of X_1 depends only on the correlation between η_1 and u and similarly the level of endogeneity of X_2 depends only on the correlation between η_2 and u. We make three different choices for the pair (ρ_{u1}, ρ_{u2}): $(\rho_{u1}, \rho_{u2}) = (0.5, 0.5), (0.1, 0.99), (0.99, 0.1)$. X_1 and X_2 are moderately (and equally) endogenous in the first case, X_1 is highly endogenous and X_2 is mildly endogenous in the second case, X_1 is mildly endogenous and X_2 is highly endogenous in the third case. We refer to the corresponding covariance matrix of the structural errors as Σ_1, Σ_2 and Σ_3 respectively.

The instruments Z are generated by drawing n independent random samples from $N(0, I_{k-1})$ independently of the structural errors and appending the matrix with an $n \times 1$ column vector 1_n. We consider three different values of k: $k = 2, 4, 20$. The first choice gives a just identified model and the latter two give over-identified models. The large value $k = 20$ is taken in accordance with the result that when the nuisance parameters are completely unidentified, the limiting null distribution of the K-statistic converges to the $\chi^2_{\nu_1}$ distribution from below as $k \to \infty$ and $n \to \infty$ such that $k/n \to 0$ [see Kleibergen (2007)].

To our knowledge, there does not exist a universally accepted measure of instrumental relevance for individual structural coefficients in a linear IV model with more than one endogenous regressor. However, for a model with a single endogenous regressor, the instruments are considered weak for the structural coefficient if the concentration parameter is less than 10 [see Staiger and Stock (1997)]. We follow Zivot et al. (2006) and generate the matrix Π such that the concentration matrix μ, as defined by Stock and Yogo (2005), is diagonal where for $i = 1, 2$, the i-th diagonal element μ_i corresponds to the concentration parameter for θ_i. The weak instrument setup for the experiment is summarized in Table 3.2.

Table 3.2: Four Cases of Weak Instruments.

	$\mu_2 = 1$	$\mu_2 = 10$
$\mu_1 = 1$	**WI-Case I** θ_1 : weak instrument θ_2 : weak instrument	**WI-Case II** θ_1 : weak instrument θ_2 :strong instrument
$\mu_1 = 10$	**WI-Case III** θ_1 : strong instrument θ_2 : weak instrument	**WI-Case IV** θ_1 : strong instrument θ_2 :strong instrument

The results reported below are based on 10,000 Monte-Carlo trials. The instrument matrix Z is kept fixed over the 10,000 trials.

We choose $\epsilon = 0.05$ and consider two different levels for the first-stage confidence set by choosing $\zeta = 0.01$ and $\zeta = 0.05$. The finite sample rejection rate of the tests for $H_1 : \theta_1 = \theta_{*1}$ are plotted in Figures 3.1 - 3.9. The four different tests considered are: (i) AR: the projection-type based on the AR statistic, (ii) K: the K-test, (iii) the new projection-type score (K) test with $\zeta = .01$, and once again, (iv) the same new projection-type score (K) test, but with $\zeta = .05$.

The figures plotting the finite-sample rejection rates show that the asymptotic assertions regarding the new projection-type score (K) test are equally true in finite-sample. Similar results for the K-test and the AR test are already well known [see for example, Kleibergen (2004) and Zivot et al. (2006)].

Simulations show that the projection-type test based on the AR-statistic is extremely conservative. As mentioned in Kleibergen (2007), the simulations also show that the K-test is conservative when the instruments are weak for θ_2; the estimated size approaches $\epsilon = 0.05$ when the number of instruments gets large (i.e. when $k = 20$).

The estimated size of the new projection-type test based on a 95% first-stage confidence region for θ_2, exceeds 5% ($= \epsilon$) in cases where this region is more likely to be empty. This happens mainly when X_2 is highly endogenous and/or the number of instruments is large [see Table 3.3], and to a large extent explains the difference in size between the new projection-type tests based on a 1% and a 5% first-stage confidence region. However, the estimated size of these two tests never exceeds 6% and 10% ($\zeta + \epsilon$) respectively as was suggested in Theorem 3.3. This means that if even slight over-rejection of the true value of the parameters is of serious consequence, one should choose ϵ and ζ such that $\epsilon + \zeta$ does not exceed the desired level of Type-I error; for any given $\epsilon + \zeta$, a smaller value of ζ with respect to that of ϵ may result in better power.

Tables 3.3–3.6 summarize the likelihood of occurrence of different structures of the first-stage confidence region based on the AR-statistic. To see how the pattern of the confidence regions vary with the sample size, we report the result for sample size $n = 100, 1000$ and 10000. While the empty set becomes less likely as the sample size increases from 100 to 1000, the same does not hold true when it further increases to 10000. This further shows that it is important to take into account the empty first-stage confidence region while designing the new projection-type score (K) test.

Although our analytical results assume that the parameter space is compact, following the convention in linear IV regression the confidence regions in Tables 3.4 and 3.5 are expressed as unbounded. Even so, it is evident that whenever the instruments are strong for θ_2, i.e. in WI-Cases II and IV, a bounded confidence region is more likely to occur, thus significantly reducing the computational cost of the new test. The bounded regions are also more likely when the number of instruments is relatively large and/or when the sample size is large. If the compactness assumption of the parameter space is relaxed (of course by compensating for it with other assumptions) the confidence region should not be bounded with high probability whenever the instruments are weak for θ_2 [see Dufour (1997)]. [6]

The simulations in this section provide a comparative study of the finite-sample behavior of the new projection-type tests with the existing tests in the literature. They corroborate the preceding analytical discussion of the usefulness of the new test under both regular and non-regular conditions of inference.

---

[6]Our experience suggests that plotting the efficient K-statistic against different values of θ_2 can be helpful in finding its minimum value. For the cases considered in this section, the efficient K-statistic seems to stabilize for distant values of the nuisance parameter θ_2 thus helping the search for the global infimum (with respect to the first-stage confidence region). Of course, it should be noted that finding the exact minimum is not absolutely necessary because the new test rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ if the efficient K-statistic exceeds the χ^2 critical value for any θ_2 belonging to the first-stage confidence region.

Table 3.3: % of times $\mathcal{C}_2(100-\zeta, \theta_{01}) = \varnothing$.

n	k	Σ	WI-Case I		WI-Case II		WI-Case III		WI-Case IV	
			$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$
10^2	2	Σ_1	0	0.09	0.28	1.27	0.01	0.18	0.23	1.28
10^2	2	Σ_2	0.27	1.43	0.35	1.48	0.27	1.51	0.30	1.58
10^2	2	Σ_3	0	0.12	0.27	1.22	0.01	0.11	0.21	1.11
10^2	4	Σ_1	0.01	0.46	0.47	2.40	0.03	0.45	0.56	2.70
10^2	4	Σ_2	0.56	2.75	0.61	2.72	0.62	2.91	0.66	3.04
10^2	4	Σ_3	0.01	0.27	0.44	2.26	0.03	0.39	0.53	2.58
10^2	20	Σ_1	0.78	3.44	1.58	5.91	0.87	3.76	1.88	6.05
10^2	20	Σ_2	1.84	6.35	1.63	6.05	1.98	6.22	1.92	6.30
10^2	20	Σ_3	0.42	2.80	1.55	5.80	0.57	3.03	1.86	5.89
10^3	2	Σ_1	0.03	0.26	0.18	1.27	0.02	0.26	0.18	1.22
10^3	2	Σ_2	0.19	1.33	0.27	1.40	0.31	1.63	0.22	1.41
10^3	2	Σ_3	0.01	0.21	0.15	1.19	0.02	0.26	0.16	1.13
10^3	4	Σ_1	0.01	0.50	0.42	2.33	0.03	0.56	0.22	2.10
10^3	4	Σ_2	0.35	2.46	0.49	2.44	0.44	2.16	0.32	2.25
10^3	4	Σ_3	0.03	0.41	0.41	2.24	0.01	0.34	0.25	2.00
10^3	20	Σ_1	0.27	2.45	0.65	3.69	0.32	2.24	0.80	3.77
10^3	20	Σ_2	0.72	3.82	0.68	3.92	0.76	3.58	0.86	3.91
10^3	20	Σ_3	0.19	1.94	0.65	3.63	0.21	1.69	0.80	3.78
10^4	2	Σ_1	0	0.17	0.19	1.27	0.01	0.18	0.15	1.31
10^4	2	Σ_2	0.27	1.49	0.22	1.50	0.29	1.45	0.23	1.52
10^4	2	Σ_3	0	0.14	0.18	1.28	0	0.09	0.14	1.28
10^4	4	Σ_1	0.04	0.35	0.32	2.08	0.02	0.50	0.22	2.01
10^4	4	Σ_2	0.41	2.15	0.36	2.41	0.28	2.29	0.25	2.18
10^4	4	Σ_3	0.03	0.31	0.31	2.05	0.03	0.46	0.22	1.88
10^4	20	Σ_1	0.27	2.11	0.59	3.41	0.22	2.11	0.58	3.74
10^4	20	Σ_2	0.69	3.54	0.65	3.56	0.65	3.40	0.63	3.83
10^4	20	Σ_3	0.17	1.68	0.59	3.36	0.19	1.71	0.56	3.69

Table 3.4: % of times $\mathcal{C}_2(100-\zeta,\theta_{01})=(-\infty,\infty)$.

n	k	Σ	WI-Case I		WI-Case II		WI-Case III		WI-Case IV	
			$\zeta=1\%$	$\zeta=5\%$	$\zeta=1\%$	$\zeta=5\%$	$\zeta=1\%$	$\zeta=5\%$	$\zeta=1\%$	$\zeta=5\%$
10^2	2	Σ_1	82.81	61.50	2.31	0.52	82.15	60.38	2.52	0.52
10^2	2	Σ_2	0	0	0	0	0	0	0	0
10^2	2	Σ_3	86.03	66.04	8.05	2.16	85.88	65.95	8.57	2.09
10^2	4	Σ_1	74.35	50.20	0.06	0.01	74.71	49.51	0.02	0
10^2	4	Σ_2	0	0	0	0	0	0	0	0
10^2	4	Σ_3	80.58	58.46	0.65	0.07	80.21	56.97	0.59	0.09
10^2	20	Σ_1	20.00	7.84	0	0	19.30	7.81	0	0
10^2	20	Σ_2	0	0	0	0	0	0	0	0
10^2	20	Σ_3	34.34	15.82	0	0	33.71	15.42	0	0
10^3	2	Σ_1	80.69	57.62	0.71	0.11	80.83	57.14	0.57	0.10
10^3	2	Σ_2	0	0	0	0	0	0	0	0
10^3	2	Σ_3	85.07	63.92	3.56	0.73	85.50	64.13	3.04	0.60
10^3	4	Σ_1	73.78	48.18	0.02	0	73.38	48.28	0	0
10^3	4	Σ_2	0	0	0	0	0	0	0	0
10^3	4	Σ_3	80.46	56.55	0.26	0.01	80.04	56.52	0.20	0.01
10^3	20	Σ_1	16.64	5.25	0	0	16.64	5.49	0	0
10^3	20	Σ_2	0	0	0	0	0	0	0	0
10^3	20	Σ_3	33.16	13.53	0	0	33.51	14.24	0	0
10^4	2	Σ_1	81.25	58.77	0.86	0.10	81.44	59.17	0.92	0.10
10^4	2	Σ_2	0	0	0	0	0	0	0	0
10^4	2	Σ_3	85.58	64.86	4.09	0.83	85.51	64.95	3.96	0.99
10^4	4	Σ_1	71.17	45.33	0	0	70.35	44.92	0	0
10^4	4	Σ_2	0	0	0	0	0	0	0	0
10^4	4	Σ_3	79.05	54.42	0.06	0	78.11	53.80	0.08	0.01
10^4	20	Σ_1	17.20	5.35	0	0	16.87	5.39	0	0
10^4	20	Σ_2	0	0	0	0	0	0	0	0
10^4	20	Σ_3	34.46	14.54	0	0	34.49	14.42	0	0

Table 3.5: % of times $\mathcal{C}_2(100-\zeta,\theta_{01}) = (-\infty, a] \cup [b, +\infty)$ for some $-\infty < a < b < +\infty$.

n	k	Σ	WI-Case I		WI-Case II		WI-Case III		WI-Case IV	
			$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$
10^2	2	Σ_1	9.66	18.95	9.20	3.16	9.79	19.28	8.79	2.90
10^2	2	Σ_2	92.22	80.11	11.85	3.77	92.43	79.56	11.11	3.73
10^2	2	Σ_3	6.40	13.63	3.45	1.34	6.06	13.72	3.09	1.51
10^2	4	Σ_1	13.87	22.81	1.01	0.11	13.63	23.41	0.94	0.15
10^2	4	Σ_2	88.72	73.07	1.24	0.22	88.26	72.55	1.14	0.22
10^2	4	Σ_3	7.79	14.55	0.34	0.08	8.37	15.32	0.37	0.08
10^2	20	Σ_1	25.63	18.02	0	0	25.52	17.63	0	0
10^2	20	Σ_2	45.04	24.81	0	0	44.98	25.16	0	0
10^2	20	Σ_3	11.00	9.93	0	0	11.10	9.56	0	0
10^3	2	Σ_1	10.37	19.23	4.2	1.06	10.12	19.32	4.23	1.25
10^3	2	Σ_2	91.13	76.70	5.08	1.26	91.13	77.23	4.93	1.26
10^3	2	Σ_3	6.01	12.43	1.42	0.57	5.68	13.26	1.61	0.48
10^3	4	Σ_1	13.49	22.28	0.30	0.07	13.73	21.37	0.23	0.06
10^3	4	Σ_2	86.81	70.00	0.34	0.03	87.51	70.74	0.27	0.03
10^3	4	Σ_3	7.49	13.70	0.16	0.01	7.11	13.52	0.13	0.02
10^3	20	Σ_1	24.79	15.36	0	0	25.33	15.65	0	0
10^3	20	Σ_2	41.60	20.69	0	0	42.14	20.65	0	0
10^3	20	Σ_3	8.55	7.27	0	0	8.34	6.75	0	0
10^4	2	Σ_1	9.87	18.83	5.09	1.46	10.19	19.14	4.93	1.28
10^4	2	Σ_2	91.47	77.49	5.89	1.53	91.75	77.36	5.87	1.58
10^4	2	Σ_3	6.05	12.94	1.86	0.81	6.09	12.91	1.61	0.57
10^4	4	Σ_1	15.13	23.07	0.16	0.02	14.40	22.12	0.14	0.01
10^4	4	Σ_2	86.70	68.83	0.11	0.02	85.27	67.04	0.12	0
10^4	4	Σ_3	7.10	13.50	0.03	0	7.09	13.56	0	0
10^4	20	Σ_1	25.98	16.17	0	0	26.31	16.12	0	0
10^4	20	Σ_2	43.34	21.40	0	0	43.86	21.71	0	0
10^4	20	Σ_3	8.59	7.27	0	0	7.96	7.20	0	0

Table 3.6: % of times $\mathcal{C}_2(100-\zeta, \theta_{01}) = [a, b]$ for some $-\infty < a < b < +\infty$.

n	k	Σ	WI-Case I		WI-Case II		WI-Case III		WI-Case IV	
			$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$
10^2	2	Σ_1	7.53	19.91	88.21	95.05	8.05	20.16	88.46	95.30
10^2	2	Σ_2	7.51	18.46	87.80	94.75	7.30	18.93	88.59	94.69
10^2	2	Σ_3	7.57	20.21	88.23	95.28	8.05	20.22	88.13	95.29
10^2	4	Σ_1	11.77	26.53	98.46	97.48	11.59	26.63	98.48	97.09
10^2	4	Σ_2	10.72	24.18	98.15	97.06	11.12	24.54	98.20	96.74
10^2	4	Σ_3	11.62	26.72	98.57	97.59	11.39	27.32	98.51	97.25
10^2	20	Σ_1	53.59	70.70	98.42	94.09	54.31	70.80	98.12	93.95
10^2	20	Σ_2	53.12	68.84	98.37	93.95	53.04	68.62	98.08	93.70
10^2	20	Σ_3	54.24	71.45	98.45	94.20	54.62	71.99	98.14	94.11
10^3	2	Σ_1	8.91	22.89	94.91	97.56	9.03	23.28	95.02	97.43
10^3	2	Σ_2	8.68	21.97	94.65	97.34	8.56	21.14	94.85	97.33
10^3	2	Σ_3	8.91	23.44	94.87	97.51	8.80	22.35	95.19	97.79
10^3	4	Σ_1	12.72	29.04	99.26	97.60	12.86	29.79	99.55	97.84
10^3	4	Σ_2	12.84	27.54	99.17	97.53	12.05	27.10	99.41	97.72
10^3	4	Σ_3	12.02	29.34	99.17	97.74	12.84	29.62	99.42	97.97
10^3	20	Σ_1	58.30	76.94	99.35	96.31	57.71	76.62	99.20	96.23
10^3	20	Σ_2	57.68	75.49	99.32	96.08	57.10	75.77	99.14	96.09
10^3	20	Σ_3	58.10	77.26	99.35	96.37	57.94	77.34	99.20	96.22
10^4	2	Σ_1	8.88	22.23	93.86	97.17	8.36	21.51	94.00	97.31
10^4	2	Σ_2	8.26	21.02	93.89	96.97	7.96	21.19	93.90	96.90
10^4	2	Σ_3	8.37	22.06	93.87	97.07	8.40	22.05	94.29	97.16
10^4	4	Σ_1	13.66	31.25	99.52	97.90	15.23	32.46	99.64	97.98
10^4	4	Σ_2	12.89	29.02	99.53	97.57	14.45	30.67	99.63	97.82
10^4	4	Σ_3	13.82	31.77	99.60	97.95	14.77	32.18	99.70	98.11
10^4	20	Σ_1	56.55	76.37	99.41	96.59	56.60	76.38	99.42	96.26
10^4	20	Σ_2	55.97	75.06	99.35	96.44	55.49	74.89	99.37	96.17
10^4	20	Σ_3	56.78	76.51	99.41	96.64	57.36	76.67	99.44	96.31

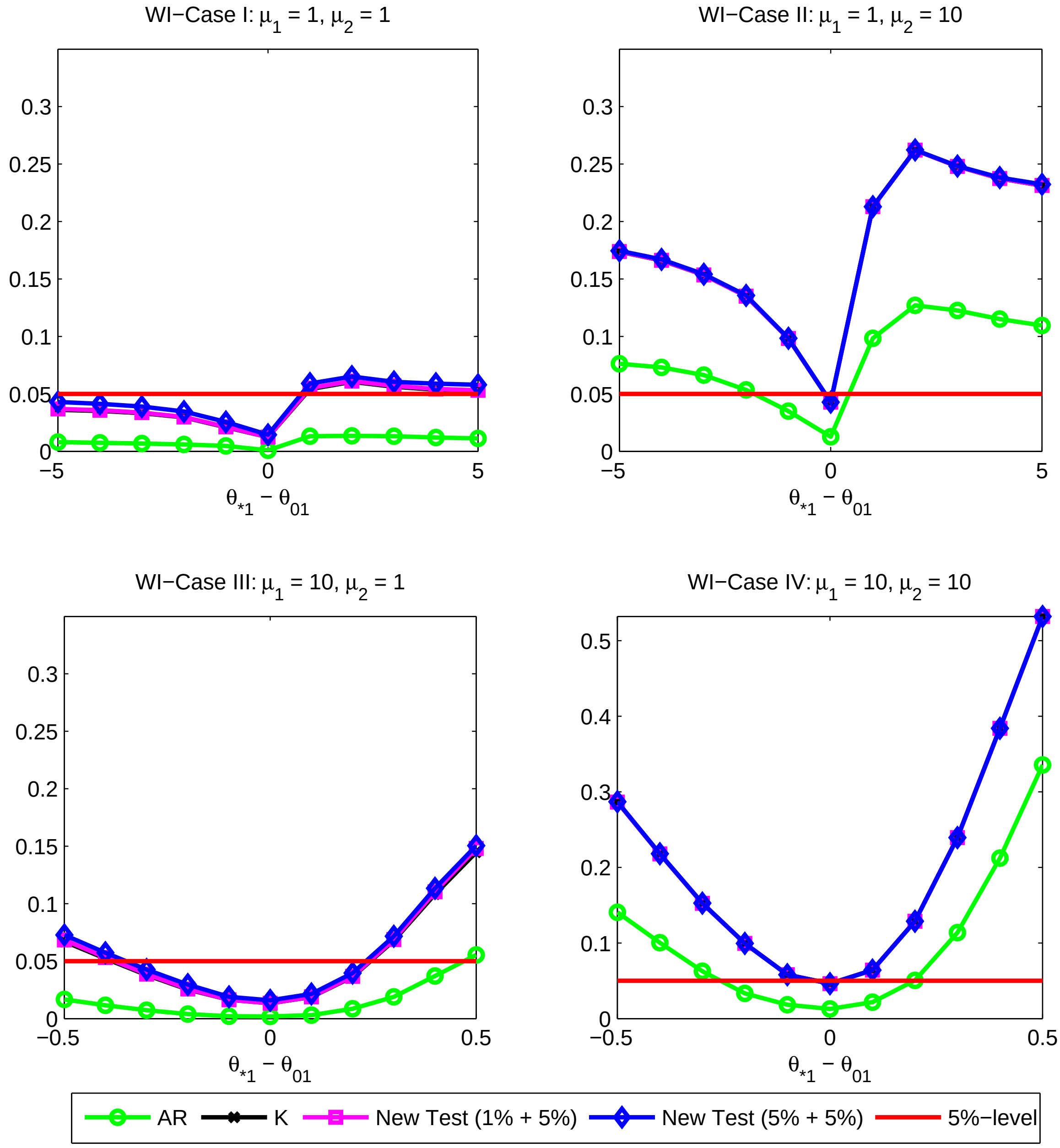

Figure 3.1: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 2, $\rho_{u1} = 0.5$, $\rho_{u2} = 0.5$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

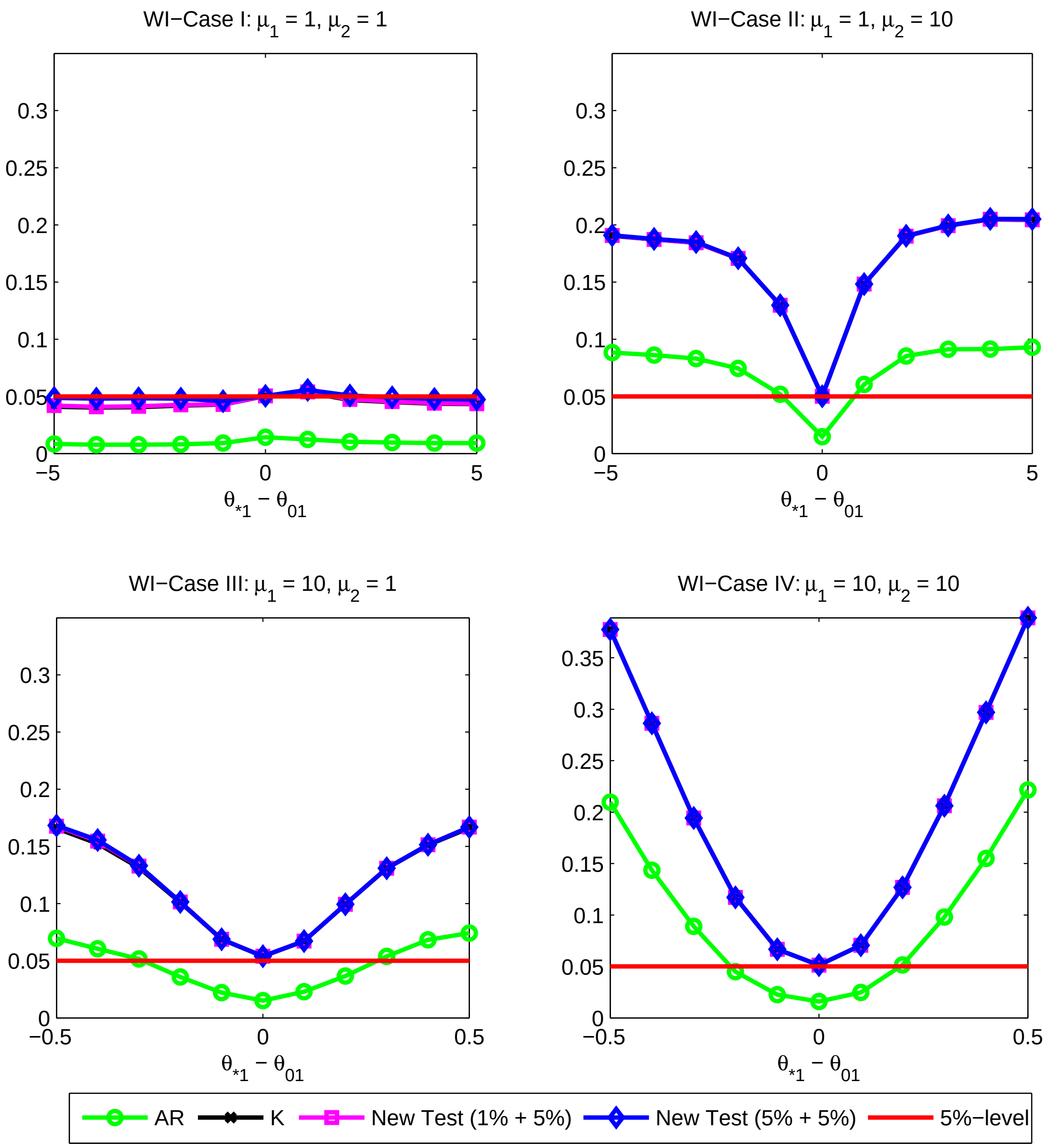

Figure 3.2: Sample Size = 100, Number of Instruments = 2, $\rho_{u1} = 0.1$, $\rho_{u2} = 0.99$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

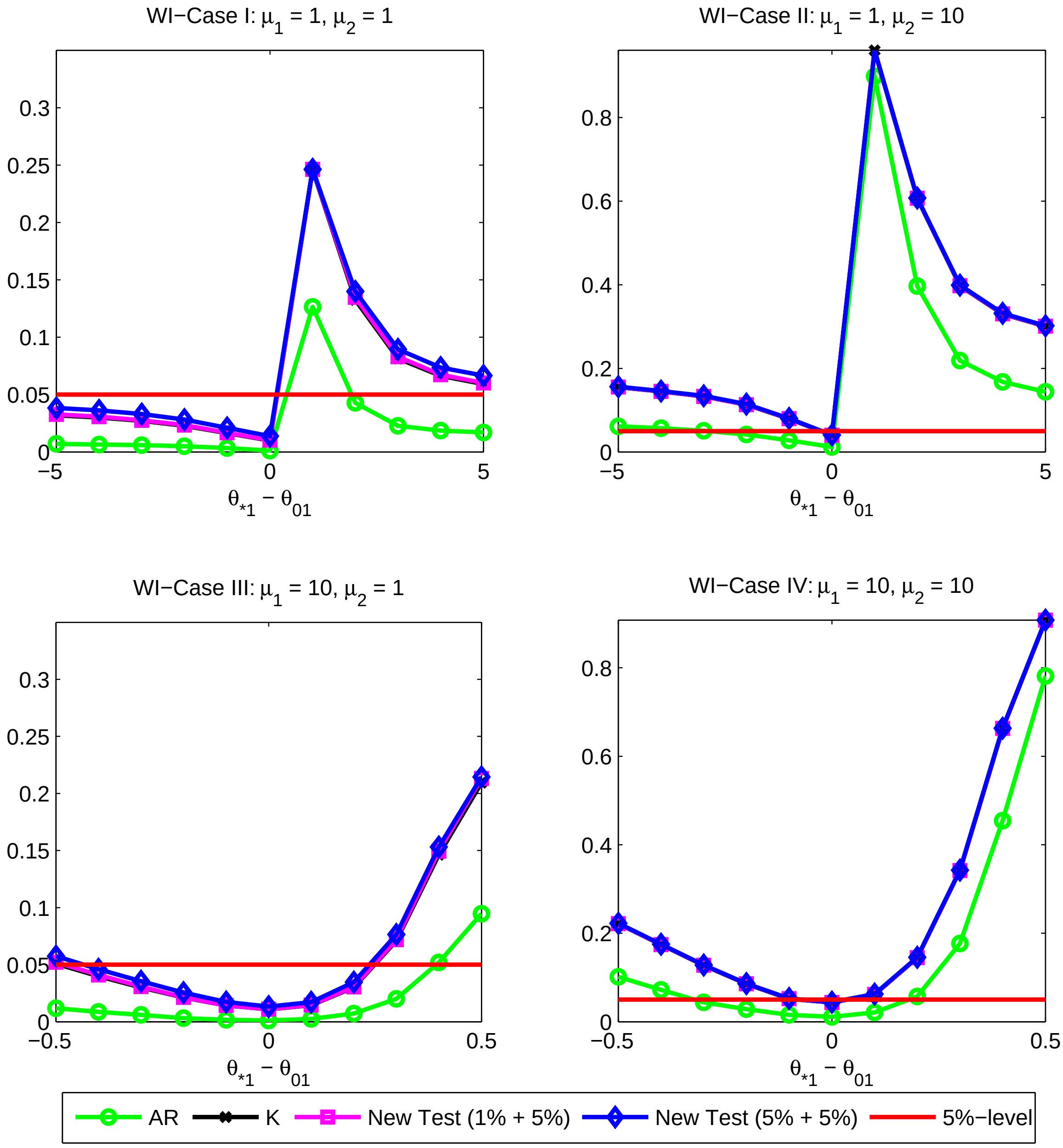

Figure 3.3: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 2, $\rho_{u1} = 0.99$, $\rho_{u2} = 0.1$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

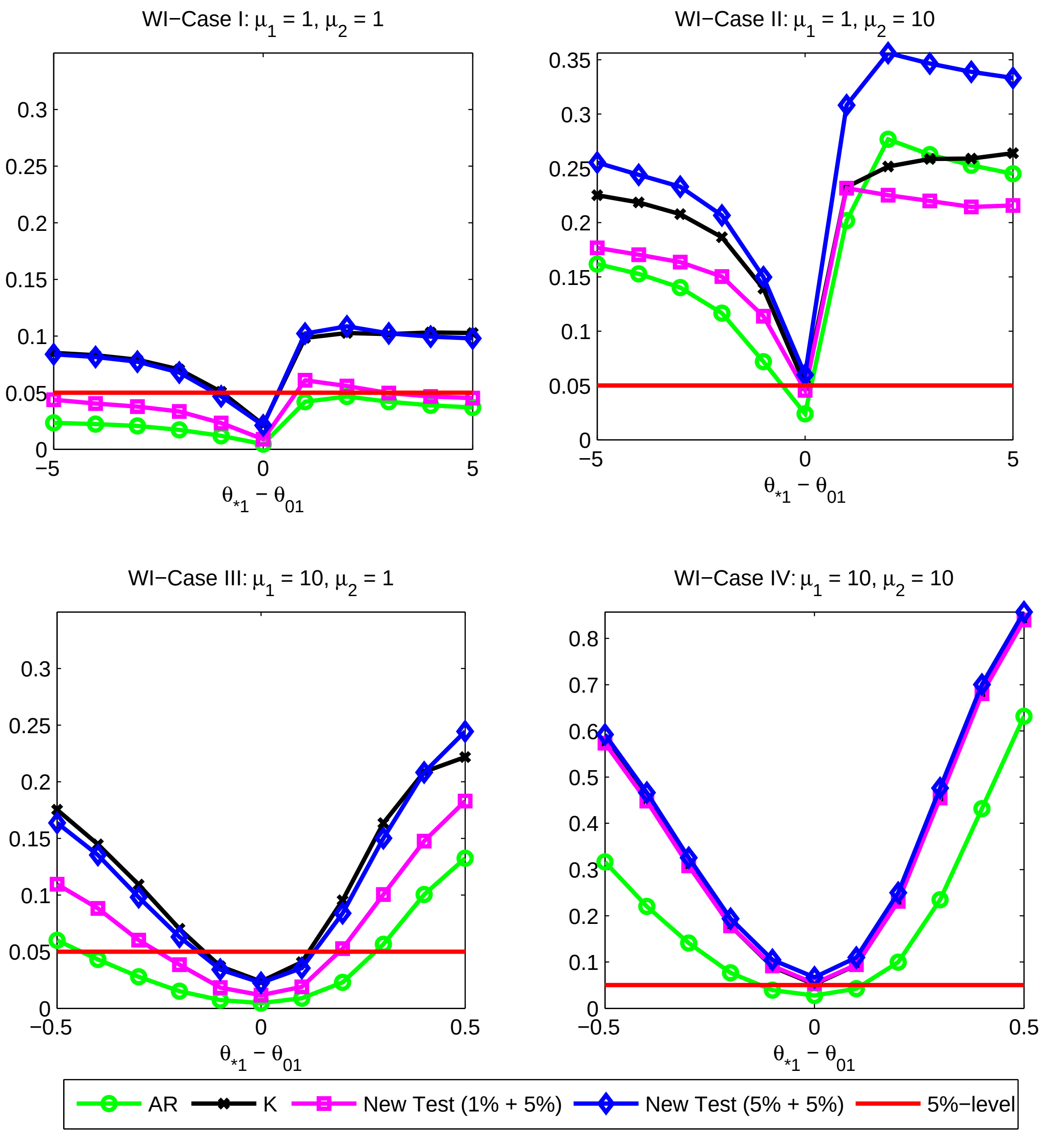

Figure 3.4: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 4, $\rho_{u1} = 0.5$, $\rho_{u2} = 0.5$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

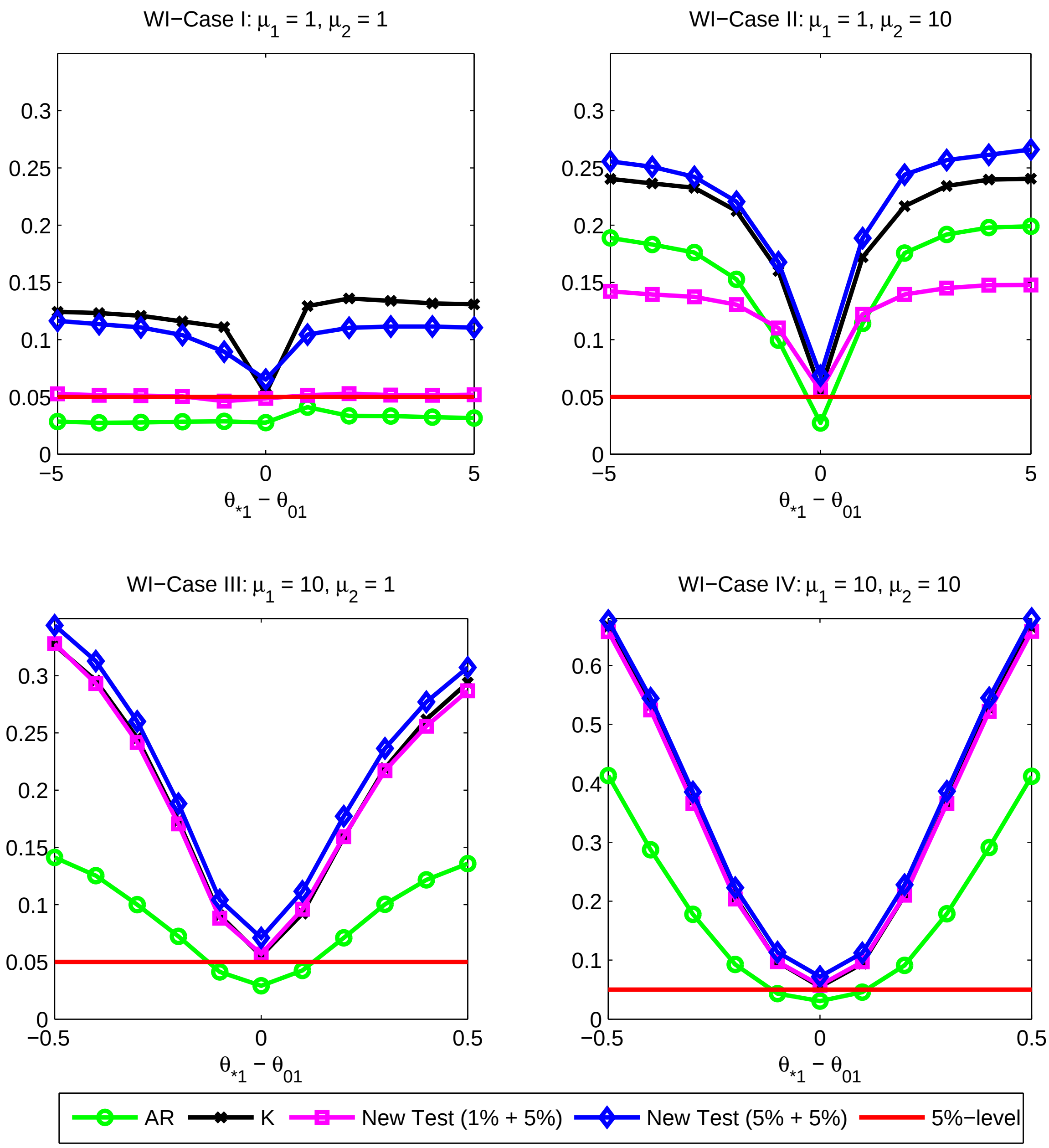

Figure 3.5: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 4, $\rho_{u1} = 0.1$, $\rho_{u2} = 0.99$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

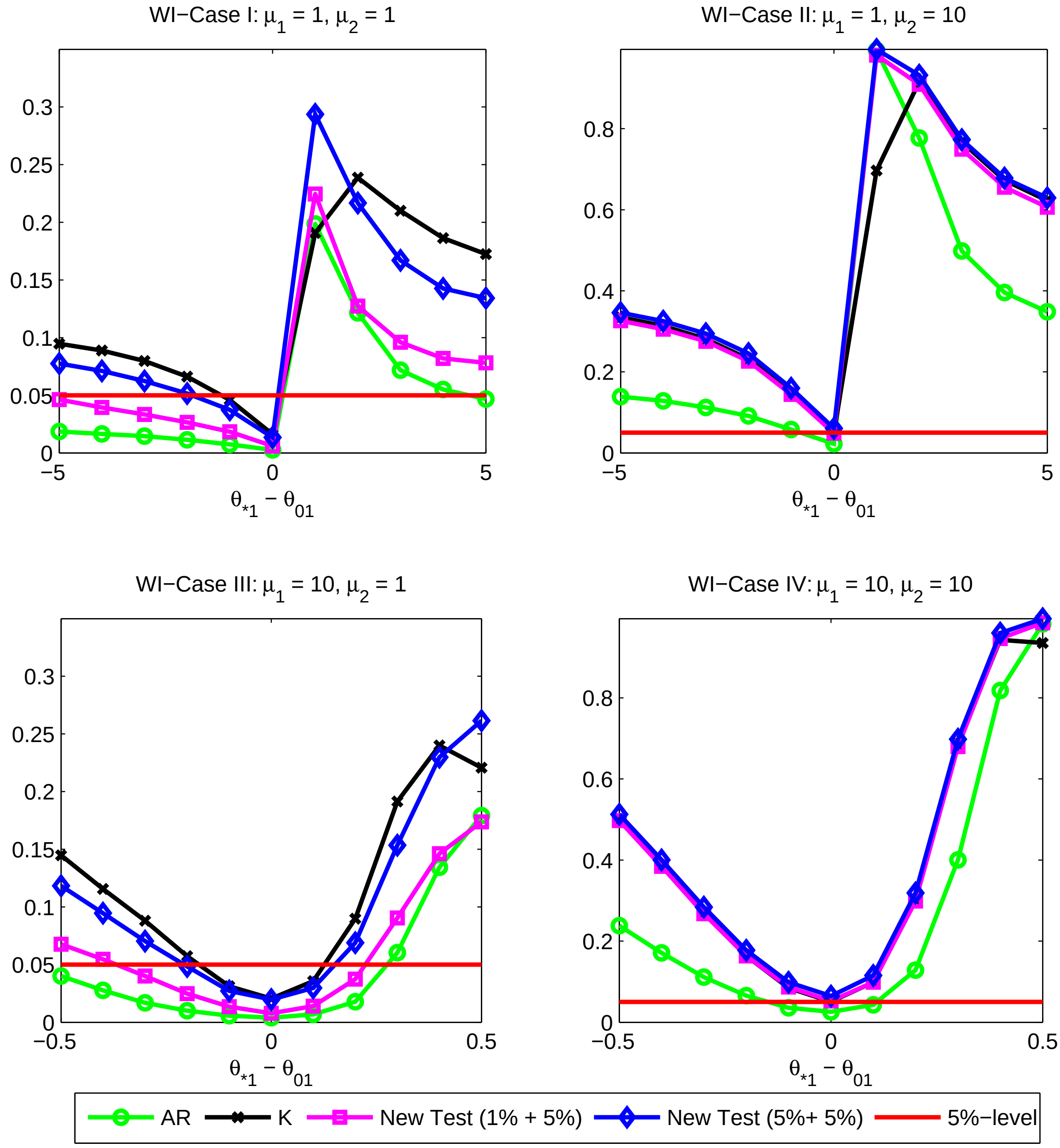

Figure 3.6: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 4, $\rho_{u1} = 0.99$, $\rho_{u2} = 0.1$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

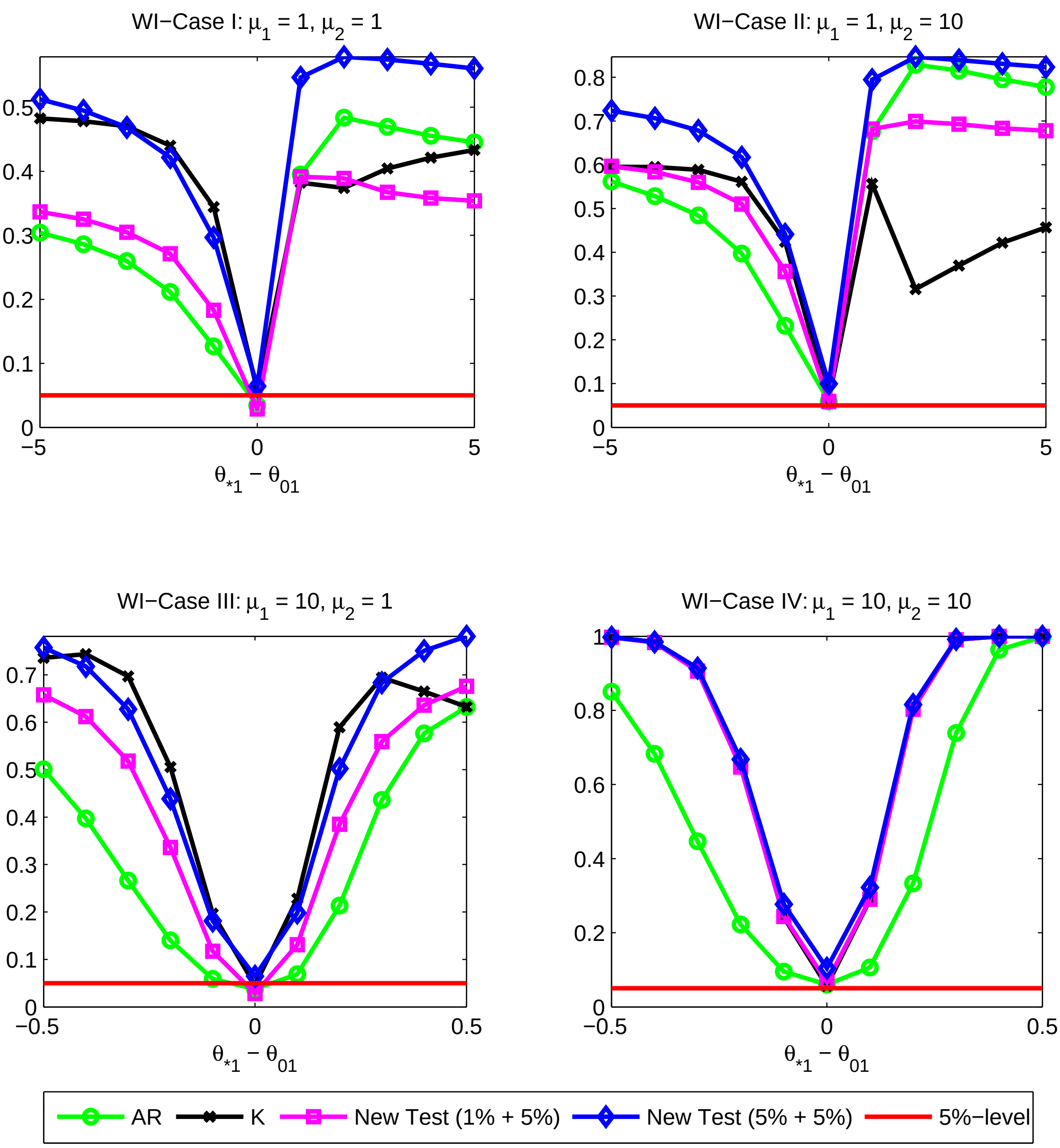

Figure 3.7: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 20, $\rho_{u1} = 0.5$, $\rho_{u2} = 0.5$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

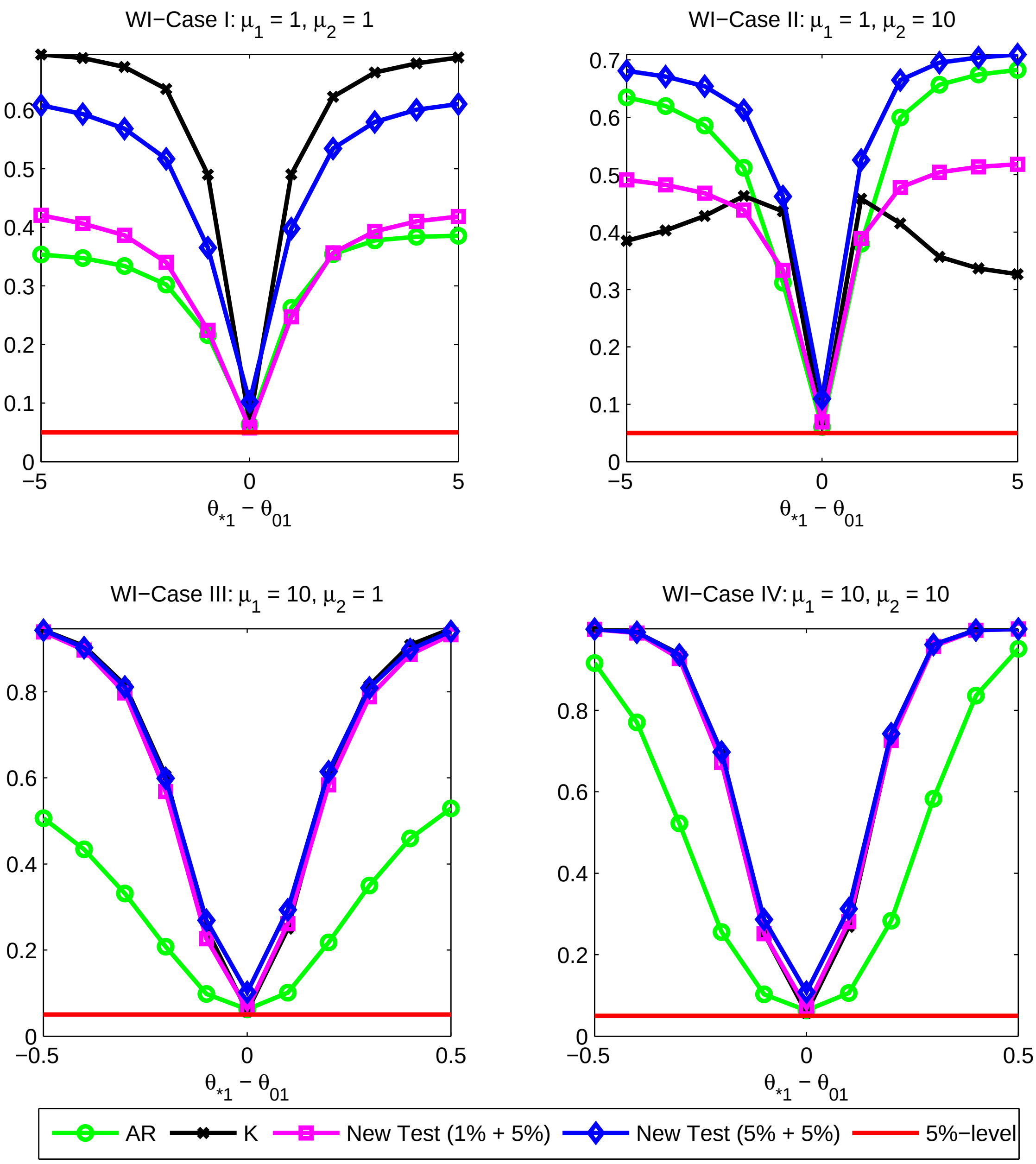

Figure 3.8: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 20, $\rho_{u1} = 0.1$, $\rho_{u2} = 0.99$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

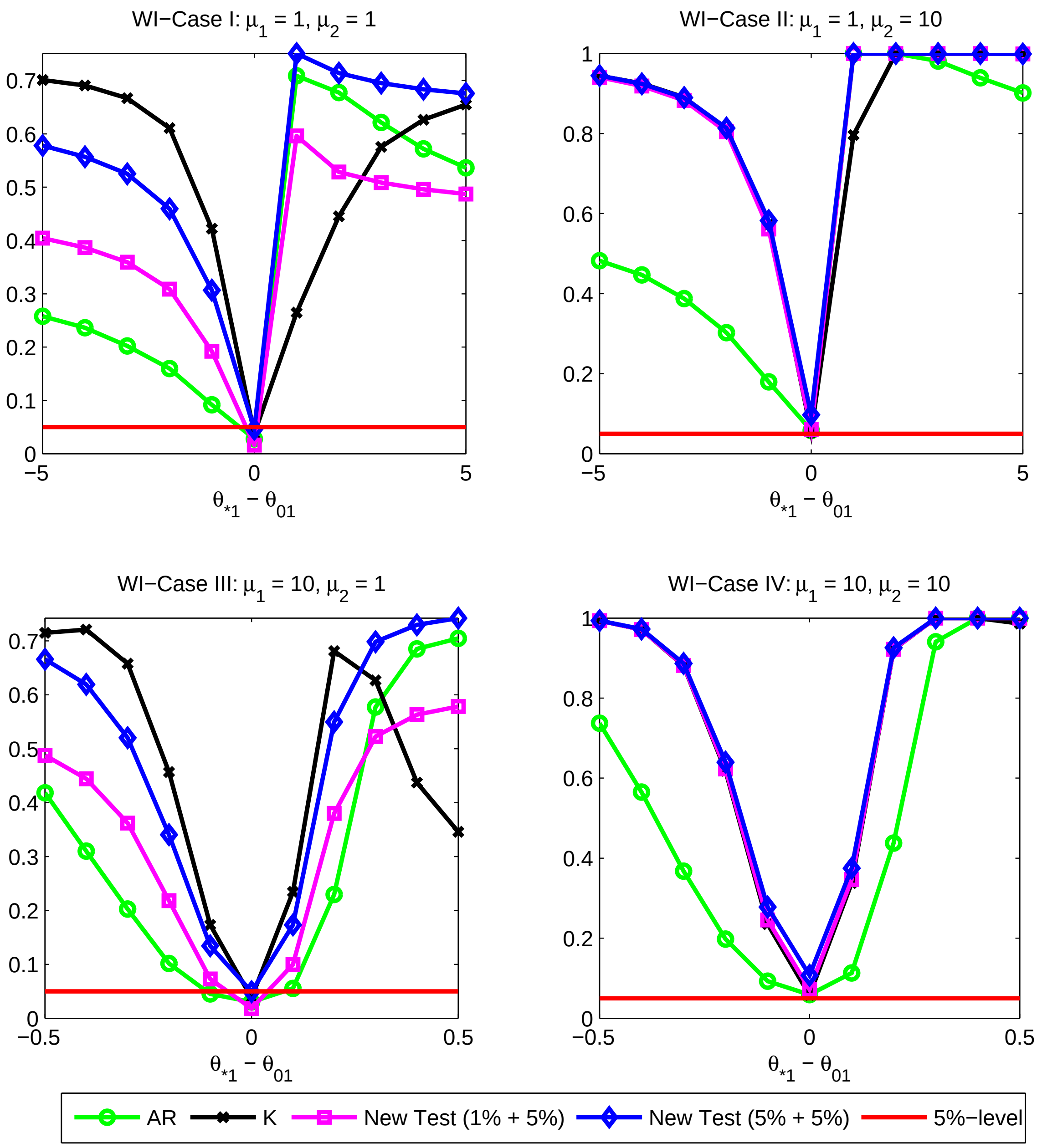

Figure 3.9: Rejection rates for $H_1 : \theta_1 = \theta_{*1}$ when n = 100, k = 20, $\rho_{u1} = 0.99$, $\rho_{u2} = 0.1$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Chapter 4
CONCLUSION

The score test relies on mild assumptions on the underlying model and often involves less computation than the Wald and LR tests. However, the asymptotic properties of the score test for subsets of parameters, in general, depend on the identifiability of both the parameters of interest and the nuisance parameters. The weak instrument setup is a common example where such identifiability restrictions are not satisfied and where the usual score test may be over-sized.

Recent research has shown that variants of the score statistic can sometimes be used to jointly test for all the parameters in the model. Dufour and his co-authors have also shown that the usual projection technique, based on the corresponding score statistic, can be used for testing subsets of parameters. However, such projection-based tests, although never over-sized, tend to be conservative.

We proposed a new method of projection-type score test for subsets of parameters. We showed that the new method is quite generally less conservative than the method of projection considered by Dufour and his co-authors. In fact, our test is locally optimal whenever local optimality can be attributed to the usual score test. At the same time, unlike the usual score test for subsets of parameters, it is also possible to impose a pre-specified upper bound to the size of our test even when the nuisance parameters are not identified.

In this thesis we have described the successful application of our test to the weak instrument/identification framework. Further application to specific types of extremum estimation is a topic of our future research.

BIBLIOGRAPHY

Aitchison, J. and Silvey, S. D. (1958). Maximum Likelihood Estimation of Parameters Subject to Constraints. *Annals of Mathematical Statistics*, 19: 813–828.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Andrews, D. W. K. (1994). Empirical Process Methods in Econometrics. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume IV, chapter 37, pages 2247–2294. Elsevier Science Publisher.

Angrist, J. and Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106: 979–1014.

Angrist, J. and Krueger, A. B. (1995). Spilt-Sample Instrumental Variables Estimates of the Return to Schooling. *Journal of Business and Economics Statistics*, 13: 225–235.

Bera, A. K. and Bilias, Y. (2001). Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: an essay on historical developments and some new results. *Journal of Statistical Planning and Inference*, 97: 9–44.

Berger, R. L. and Boos, D. D. (1994). P-Values Maximized Over a Confidence Set for the Nuisance Parameter. *Journal of the American Statistical Association*, 89: 1012–1016.

Blackburn, M. and Neumark, D. (1992). Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials. *The Quarterly Journal of Economics*, 107: 1421–1436.

Boos, D. D. (1992). On Generalized Score Tests. *The American Statistician*, 46: 327–333.

Bound, J., Jaeger, D. A., and Baker, R. (1995). Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak. *Journal of the American Statistical Association*, 90: 443–450.

Card, D. (1995). Using Geographical Variation in College Proximity to Estimate the Return to Schooling. In Christofides, L. N., Grant, E. K., and Swidinsky, R., editors, *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*. University of Toronto Press.

Chaudhuri, S., Richardson, T., Robins, J., and Zivot, E. (2007). Split-Sample Score Tests in Linear Instrumental Variables Regression. Technical Report 73, CSSS, University of Washington.

Conniffe, D. (2001). Score tests when a nuisance parameter is unidentified under the null hypothesis. *Journal of Statistical Planning and Inference*, 97: 67–83.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64: 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74: 33–43.

Dufour, J. M. (1997). Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, 65: 1365–1388.

Dufour, J. M. and Jasiak, J. (2001). Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors. *International Economic Review*, 42: 815–843.

Dufour, J. M. and Taamouti, M. (2005a). Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments. Discussion Paper.

Dufour, J. M. and Taamouti, M. (2005b). Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments. *Econometrica*, 73: 1351–1365.

Foutz, R. V. and Srivastava, R. C. (1977). The Performance of the Likelihood Ratio Test When the Model Is Incorrect. *Annals of Statistics*, 5: 1183–1194.

Gourieroux, C., Monfort, A., and Trognon, A. (1983). Testing Nested or Nonnested Hypotheses. *Journal of Econometrics*, 21: 83–115.

Gregory, A. W. and Veall, M. R. (1985). Formulating Wald Tests Of Nonlinear Restrictions. *Econometrica*, 53: 1465–1468.

Griliches, Z. (1977). Estimating the Returns to Schooling - Some Econometric Problems. *Econometrica*, 45: 1–22.

Guggenberger, P. and Smith, R. (2005). Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification. *Econometric Theory*, 21: 667–709.

Hansen, B. E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. *Econometrica*, 64: 413–430.

Holly, A. (1982). A Remark on Hausman's specification Test. *Econometrica*, 50: 749–759.

Kent, J. T. (1982). Robust Properties of Likelihood Ratio Tests. *Biometrika*, 69: 19–27.

Kleibergen, F. (2002). Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, 70: 1781–1803.

Kleibergen, F. (2004). Testing Subsets of Parameters In The Instrumental Variables Regression Model. *The Review of Economics and Statistics*, 86: 418–423.

Kleibergen, F. (2005). Testing Parameters In GMM Without Assuming That They Are Identified. *Econometrica*, 73: 1103–1123.

Kleibergen, F. (2007). Subset statistic in the linear IV regression model. Technical report, Brown University.

Moreira, M. J. (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, 71: 1027–1048.

Nelson, C. R. and Startz, R. (1990a). Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, 54: 967–976.

Nelson, C. R. and Startz, R. (1990b). The Distribution of the Instrumental Variable Estimator and its t-ratio When the Instrument is a Poor One. *Journal of Business*, 63: 125–140.

Nelson, C. R. and Startz, R. (2007). The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics*, 138: 47–62.

Newey, W. K. and McFadden, D. L. (1994). Large Sample Estimation and Hypothesis Testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume IV, chapter 36, pages 2212–2245. Elsevier Science Publisher.

Newey, W. K. and West, K. D. (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review*, 28: 777–787.

Neyman, J. (1959). Optimal Asymptotic Test of Composite Statistical Hypothesis. In Grenander, U., editor, *Probability and Statistics, the Harald Cramer Volume*, pages 313–334. Almqvist and Wiksell, Uppsala.

Okamoto, M. (1973). Distinctness of the Eigenvalues of a Quadratic form in a Multivariate Sample. *Annals of Statistics*, 1: 763–765.

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

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44: 50–57.

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In Lin, D. Y. and Heagerty, P., editors, *Proceedings of the Second Seattle Symposium on Biostatistics*. New York: Springer.

Silvapulle, M. J. (1996). A Test in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*, 91: 1690–1693.

Silvey, S. D. (1959). The Lagrange Multiplier test. *Annals of Mathematical Staistics*, 30: 389–407.

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

Stock, J. H. and Wright, J. H. (2000). GMM with Weak Identification. *Econometrica*, 68: 1055–1096.

Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wang, J. and Zivot, E. (1998). Inference on a Structural Parameter in Instrumental Variables Regression with Weak Instruments. *Econometrica*, 66: 1389–1404.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50: 1–25.

Zivot, E., Startz, R., and Nelson, C. (1998). Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review*, 39: 1119–1144.

Zivot, E., Startz, R., and Nelson, C. (2006). Inference in Weakly Identified Instrumental Variables Regression. In Corbae, D., Durlauf, S. N., and Hansen, B. E., editors, *Frontiers in Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, pages 125–166. Cambridge University Press.

Appendix A
PROOF OF RESULTS

Lemma A.1 *Let $\mathcal{N}_c \subset \mathbb{R}^\nu$ be compact and let $\mathcal{N} = \mathcal{N}_c - \mathcal{N}_b$ where $\mathcal{N}_b$ denotes the boundary of $\mathcal{N}_c$.*

(i) Let $\widehat{a}_n(\theta) - a_n(\theta) = o_p(1)$ and $a_n(\theta) - a(\theta) = o(1)$ for $\theta \in \mathcal{N}$. Then $\widehat{a}_n(\theta_n) - a(\theta_0) = o_p(1)$ if $a(\theta)$ is continuous at θ_0 and $\theta_n - \theta_0 = o_p(1)$.

(ii) In addition, let $\widehat{b}_n(\theta) - b_n(\theta) = o_p(1)$ and $b_n(\theta) - b(\theta) = o(1)$ for $\theta \in \mathcal{N}$. If $a(\theta)$ and $b(\theta)$ are bounded on $\mathcal{N}$, then $\widehat{a}_n(\theta_n)\widehat{b}_n(\theta_n) \xrightarrow{P} a(\theta_0)b(\theta_0)$ if $a(\theta)$ and $b(\theta)$ are continuous at θ_0 and $\theta_n - \theta_0 = o_p(1)$.

The resultant convergence in probability is uniform if the convergence in probability and the continuity are uniform in the statement of Lemma A.1. Below is a rough sketch of the proof of Lemma A.1.

(i) Using the Triangle Inequality, the results follow once we note that: (a) for n large enough $\theta_n \in \mathcal{N}$ w.p.a.1 and $\|\widehat{a}_n(\theta_n) - a(\theta_0)\| \leq \|\widehat{a}_n(\theta) - a_n(\theta)\| + \|a_n(\theta) - a(\theta)\| + \|a(\theta) - a(\theta_0)\|$.

(ii) Let $\widehat{a}_n(.)$ and $a(.)$ be $p_a \times p$ and $\widehat{b}_n(.)$ and $b(.)$ be $p \times p_b$ finite-dimensional matrices. Defining $I_a = \{(i,j) : 1 \leq i \leq p_a, 1 \leq j \leq p\}$ and $I_b = \{(i,j) : 1 \leq i \leq p, 1 \leq j \leq p_b\}$, let $\sup_{\theta \in \mathcal{N}} \max_{(i,j) \in I_a} a_{(i,j)}(\theta) \leq R_a = O(1)$ and $\sup_{\theta \in \mathcal{N}} \max_{(i,j) \in I_b} b_{(i,j)}(\theta) \leq R_b = O(1)$. Then the results follow using the same technique as in (i) once we note that the Triangle Inequality and the Cauchy-Schwartz Inequality give $\|\widehat{a}_n(\theta)\widehat{b}_n(\theta) - a(\theta)b(\theta)\| \leq \|\widehat{a}_n(\theta) - a(\theta)\|\|\widehat{b}_n(\theta) - b(\theta)\| + \sqrt{p_a p}|R_a|\|\widehat{b}_n(\theta) - b(\theta)\| + \|\widehat{a}_n(\theta) - a(\theta)\|\sqrt{p p_b}|R_b|$. ∎

In the following, for all the Mean-Value expansions of some functions of θ, the Mean-Value is generically denoted by $\bar{\theta}$ (unless it is extremely confusing). This Mean-Value obviously is different for each expansion and also for any particular expansion, the rows of this vector vary. We apologize for any confusion due to our notation.

It will be helpful to prove the following lemmas before proving Theorem 2.1. These intermediate results are standard and similar proofs can be found in most graduate econometrics textbooks. Nevertheless, we provide all the proofs for the sake of completeness.

Lemma A.2 *Let $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1$ where $d_1 \in \mathbb{R}^{\nu_1}$ is fixed. Under Assumptions A1, A2 and A3,*

(i) $\widehat{\theta}_n \xrightarrow{P} \theta_0$ and (ii) $\tilde{\theta}_{n2}(\theta_{*1}) \xrightarrow{P} \theta_{02}$.

Proof: (i) Choose any $\epsilon > 0$ arbitrarily. Then with probability approaching one (w.p.a.1) we have: (a) $n^{-1}Q_n(\widehat{\theta}_n) > n^{-1}Q_n(\theta_0) - \epsilon/3$ (by definition), (b) $Q(\widehat{\theta}_n) > n^{-1}Q_n(\widehat{\theta}_n) - \epsilon/3$ (by A1) and (c) $n^{-1}Q_n(\theta_0) > Q(\theta_0) - \epsilon/3$ (by A1). Therefore, w.p.a.1, $Q(\widehat{\theta}_n) > n^{-1}Q_n(\widehat{\theta}_n) - \epsilon/3 > n^{-1}Q_n(\theta_0) - 2\epsilon/3 > Q(\theta_0) - \epsilon$. Since $\epsilon > 0$ was chosen arbitrarily, we have for any $\epsilon > 0$, $Q(\widehat{\theta}_n) > Q(\theta_0) - \epsilon$ w.p.a.1. Now consider any open neighborhood $\mathcal{N} \subset \Theta$ containing θ_0. Since $\mathcal{N}^c \cap \Theta$ is compact, by A1, we have $\sup_{\theta \in \mathcal{N}^c \cap \Theta} Q(\theta) = Q(\theta^\dagger) < Q(\theta_0)$ for some $\theta^\dagger \in \mathcal{N}^c \cap \Theta$. Now choosing $\epsilon = Q(\theta_0) - \sup_{\theta \in \mathcal{N}^c \cap \Theta} Q(\theta)$, it follows that w.p.a.1, $Q(\widehat{\theta}_n) > \sup_{\theta \in \mathcal{N}^c \cap \Theta} Q(\theta)$ and hence $\widehat{\theta}_n \in \mathcal{N}$. Since $\mathcal{N}$ was chosen arbitrarily, it follows that $\widehat{\theta}_n \xrightarrow{P} \theta_0$.

(ii) Denote $\theta_* = (\theta_{*1}', \theta_2')'$. A second order Taylor series expansion gives

$$\frac{1}{n}Q_n(\theta_*) = \frac{1}{n}Q_n(\theta_0) + \frac{1}{n}\nabla_\theta Q_n(\theta_0)(\theta_* - \theta_0) + \frac{1}{2n}(\theta_* - \theta_0)'\nabla_{\theta\theta}Q_n(\bar{\theta})(\theta_* - \theta_0) \quad \textbf{(A.1)}$$

for some $\bar{\theta}$ such that $\|\bar{\theta} - \theta_0\| \leq \|\theta_* - \theta_0\|$. Taking probability limits on both sides of (A.1) and using A2 and A3, we get $Q(\theta_*) = Q(\theta_0) + \frac{1}{2}(\theta_2 - \theta_{02})' A_{22}(\bar{\theta})(\theta_2 - \theta_{02})$, meaning $Q(\theta_*) < Q(\theta_0)$ by the negative definiteness of $A(\theta)$. Also by A1 and (A.1) it follows that $n^{-1}Q_n(\theta_*)$ converges uniformly to $Q(\theta_*) = Q(\theta_0) + \frac{1}{2}(\theta_2 - \theta_{02})' A_{22}(\bar{\theta})(\theta_2 - \theta_{02})$ which is continuous and has a unique maximum at $\theta_2 = \theta_{02}$. Hence consistency of $\tilde{\theta}_{n2}(\theta_{*1})$ follows from the arguments in (i). ∎

It follows from the proof of Lemma A.2 that $\widehat{\theta}_n \in$ interior(Θ) w.p.a.1; and for $\theta_{*1} \in$ interior(Θ_1), it follows that $\tilde{\theta}_{n2}(\theta_{*1}) \in$ interior(Θ_2). This, along with Assumption A1, implies that $\partial Q_n(\widehat{\theta}_n)/\partial\theta = 0$ and $\partial Q_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1}))/\partial\theta_2 = 0$. We use this argument implicitly by stating that "Assumption A1 (combined with the definition of . . .)" in all the following derivations.

Lemma A.3 *Let $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1$ where $d_1 \in \mathbb{R}^{\nu_1}$ is fixed. Under Assumptions A1, A2 and A3,*

(i) $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} -A^{-1}[\Psi_1', \Psi_2']'$;

(ii) $\sqrt{n}(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}) \xrightarrow{d} -A_{22}^{-1}\,[A_{21}d_1 + \Psi_2]$.

Proof: (i) Lemma A.2 gives the consistency of $\widehat{\theta}_n$. Now, Assumption A1 (combined with the definition of $\widehat{\theta}_n$) and a Mean-Value expansion of $n^{-1/2}\nabla_\theta Q_n(\widehat{\theta}_n)$ give

$$0 = \frac{1}{\sqrt{n}}\nabla_\theta Q_n(\widehat{\theta}_n) = \frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_0) + \left[\frac{1}{n}\nabla_{\theta\theta} Q_n(\bar{\theta})\right]\sqrt{n}(\widehat{\theta}_n - \theta_0)$$

where $\bar{\theta}$ is such that $\|\bar{\theta} - \theta_0\| \leq \|\widehat{\theta}_n - \theta_0\| = o_p(1)$. Therefore, using Slutsky's Theorem and Assumptions A2 and A3 we get

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} -A^{-1}\Psi.$$

(ii) Again, as before Lemma A.2 gives the consistency of $\tilde{\theta}_{n2}(\theta_{*1})$. Define $\tilde{\theta}_* = (\theta'_{*1}, \tilde{\theta}'_{n2}(\theta_{*1}))'$. Assumption A1 (combined with the definition of $\tilde{\theta}_{n2}(\theta_{*1})$) and a Mean-Value expansion of $n^{-1/2}\nabla_2 Q_n(\tilde{\theta}_*)$ give

$$
\begin{aligned}
0 &= \frac{1}{\sqrt{n}}\nabla_2 Q_n(\tilde{\theta}_*) \\
&= \frac{1}{\sqrt{n}}\nabla_2 Q_n(\theta_0) + \left[\frac{1}{n}\nabla_{21} Q_n(\bar{\theta}_{(1)})\right] d_1 + \left[\frac{1}{n}\nabla_{22} Q_n(\bar{\theta}_{(2)})\right] \sqrt{n}(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02})
\end{aligned}
$$

for some $\bar{\theta}_{(1)}$ and $\bar{\theta}_{(2)}$ such that for $i = 1, 2$, $\|\bar{\theta}_{(i)} - \theta_0\| \leq \|\tilde{\theta}_* - \theta_0\| = o_p(1)$. Hence as before in part (i) and using Assumptions A2 and A3,

$$
\sqrt{n}(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}) \xrightarrow{d} -A_{22}^{-1}\left[A_{21}d_1 + \Psi_2\right]. \quad \blacksquare
$$

Lemma A.4 *Let $\theta_{*i} = \theta_{0i} + d_i/\sqrt{n} \in \Theta_i$ where $d_i \in \mathbb{R}^{\nu_i}$ is fixed for $i = 1, 2$. Define $\Psi_{1.2} = \Psi_1 - A_{12}A_{22}^{-1}\Psi_2$, $\theta_* = (\theta'_{*1}, \theta'_{*2})'$ and $\tilde{\theta}_* = (\theta'_{*1}, \tilde{\theta}'_{n2}(\theta_{*1}))'$. Let $\widehat{A}(\theta) \xrightarrow{P} A(\theta)$ and $\widehat{B}(\theta) \xrightarrow{P} B(\theta)$ for $\theta \in \Theta$. Under Assumptions A1, A2 and A3,*

(i) $W_1(\theta_{*1}) \xrightarrow{d} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right]' \Omega_{11}^{-1} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right]$;

(ii) $R_1(\tilde{\theta}_*) \equiv R^{alt}(\tilde{\theta}_*) \equiv R^{alt2}(\tilde{\theta}_*) \xrightarrow{d} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right]' \Omega_{11}^{-1} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right]$;

(iii) $R_1(\theta_*) = R_1(\tilde{\theta}_*) + o_p(1)$.

Proof: (i) Lemma A.2 gives the consistency of $\widehat{\theta}_n$. Hence by the Continuous Mapping Theorem $\widehat{A}(\widehat{\theta}_n) \xrightarrow{P} A$. Assumption A2 states that there exists an open neighborhood of θ_0 where $B(\theta)$ is continuous. By consistency of $\widehat{\theta}_n$, it follows that for n large enough, $\widehat{\theta}_n$ belongs to that neighborhood w.p.a.1 and hence using the Continuous Mapping Theorem we get $\widehat{B}(\widehat{\theta}_n) \xrightarrow{P} B$. Therefore by continuity of matrix inversion, and Assumptions A2 and A3, we have $\widehat{\Omega}(\widehat{\theta}_n) \xrightarrow{P} A^{-1}BA^{-1} = \Omega$ (say). Letting Ω_{11} denote the top left $\nu_1 \times \nu_1$ block of Ω, it can be seen that

$$\Omega_{11} = A_{11.2}^{-1} GBG' A_{11.2}^{-1} \text{ where } G = \left[I_{\nu_1}, -A_{12}A_{22}^{-1}\right].$$

It follows from Lemma A.3 (i) that

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_{n1} - \theta_{*1}) &= \sqrt{n}(\widehat{\theta}_{n1} - \theta_{01}) - \sqrt{n}(\theta_{*1} - \theta_{01}) \\ &\xrightarrow{d} -A_{11.2}^{-1}\Psi_{1.2} - d_1 \sim N(-d_1, A_{11.2}^{-1}GBG'A_{11.2}^{-1} = \Omega_{11}).\end{aligned}$$

Therefore, $W_1(\theta_{*1}) \xrightarrow{d} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right]' \Omega_{11}^{-1} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right] \sim \chi^2_{\nu_1}\left(d_1'\Omega_{11}^{-1}d_1\right)$.

(ii) $R_1(\tilde{\theta}_*) \equiv R^{\text{alt}}(\tilde{\theta}_*) \equiv R^{\text{alt2}}(\tilde{\theta}_*)$ by Assumption A1 and the definition of $\tilde{\theta}_{n2}(\theta_{*1})$. A Mean-Value expansion of $n^{-1/2}\nabla_1 Q_n(\tilde{\theta}_*)$ gives

$$\frac{1}{\sqrt{n}}\nabla_1 Q_n(\tilde{\theta}_*) = \frac{1}{\sqrt{n}}\nabla_1 Q_n(\theta_0) + \left[\frac{1}{n}\nabla_{11} Q_n(\bar{\theta}_{(1)})\right] d_1 + \left[\frac{1}{n}\nabla_{12} Q_n(\bar{\theta}_{(2)})\right] \sqrt{n}(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02})$$

for some $\bar{\theta}_{(1)}$ and $\bar{\theta}_{(2)}$ such that for $i = 1, 2$, $\|\bar{\theta}_{(i)} - \theta_0\| \le \|\tilde{\theta}_* - \theta_0\| = o_p(1)$. Hence using Assumptions A2 and A3, and Lemma A.3 (ii)

$$\begin{aligned}\frac{1}{\sqrt{n}}\nabla_1 Q_n(\tilde{\theta}_*) &\xrightarrow{d} \Psi_1 + A_{11}d_1 - A_{12}A_{22}^{-1}[A_{21}d_1 + \Psi_2] = \Psi_{1.2} + A_{11.2}d_1 \\ &\sim N(A_{11.2}d_1, GBG' = A_{11.2}\Omega_{11}A_{11.2}).\end{aligned}$$

As in part (i), it follows from the fact that $\tilde{\theta}_* \xrightarrow{P} \theta_0$ and the Continuous Mapping Theorem that $\left[\widehat{G}_1(\tilde{\theta}_*)\widehat{B}(\tilde{\theta}_*)\widehat{G}'(\tilde{\theta}_*)\right]^+ \xrightarrow{P} (GBG')^{-1}$ and hence

$$R_1(\tilde{\theta}_*) \equiv R^{\text{alt}}(\tilde{\theta}_*) \equiv R^{\text{alt2}}(\tilde{\theta}_*) \xrightarrow{d} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right]' \Omega_{11}^{-1} \left[A_{11.2}^{-1}\Psi_{1.2} + d_1\right] \sim \chi^2_{\nu_1}\left(d_1'\Omega_{11}^{-1}d_1\right).$$

(iii) A Mean-Value expansion of $n^{-1/2}\nabla_\theta Q_n(\theta_*)$, along with Assumptions A2 and

A3, gives

$$\frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_*) = \frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_0) + \left[\frac{1}{n}\nabla_{\theta\theta}Q_n(\bar{\theta})\right] d_\theta$$

for some $\bar{\theta}$ such that $\sqrt{n}\|\bar{\theta}-\theta_0\| \le \sqrt{n}\|\theta_*-\theta_0\| = \sqrt{d_\theta' d_\theta}$ where $d_\theta = (d_1', d_2')'$. Therefore, as before, noting that $\widehat{A}(\theta_*) \xrightarrow{P} A$ and $\widehat{B}(\theta_*) \xrightarrow{P} B$, we get

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_*) \xrightarrow{d} \Psi + Ad_\theta, \left[\widehat{G}(\theta_*)\widehat{B}(\theta_*)\widehat{G}'(\theta_*)\right]^+ \xrightarrow{P} (GBG')^{-1} \text{ and} \\
&\frac{1}{\sqrt{n}}\nabla_1 Q_n(\theta_*) - \widehat{A}_{12}(\theta_*)\widehat{A}_{22}^{-1}(\theta_*)\frac{1}{\sqrt{n}}\nabla_2 Q_n(\theta_*) \\
&\xrightarrow{d} [\Psi_1 + (A_{11}d_1 + A_{12}d_2)] - A_{12}A_{22}^{-1}[\Psi_2 + (A_{21}d_1 + A_{22}d_2)] = \Psi_{1.2} + A_{11.2}d_1
\end{aligned}$$

implying $R_1(\theta_*) = R_1(\tilde{\theta}_*) + o_p(1)$. ∎

Proof of Theorem 2.1: (i) In the following, whenever we refer to $\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_1)} R_1(\theta_1,\theta_{*2})$, it is implied that $\mathcal{C}_2(1-\zeta,\theta_1)$ is non-empty. The asymptotic size of the new projection-type score test is

$$\begin{aligned}
& \lim_{n\to\infty} Pr_{\theta_{01}}\left[\{\mathcal{C}_2(1-\zeta,\theta_{01})=\varnothing\} \cup \{\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})} R_1(\theta_{01},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon)\}\right] \\
\le\ & 1 - \lim_{n\to\infty} Pr_{\theta_{01}}\left[\{\theta_{02}\in\mathcal{C}_2(1-\zeta,\theta_{01})\} \cap \{\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})} R_1(\theta_{01},\theta_{*2}) \le \chi^2_{\nu_1}(1-\epsilon)\}\right] \\
=\ & 1 - \lim_{n\to\infty} Pr_{\theta_{01}}\left[\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})} R_1(\theta_{01},\theta_{*2}) \le \chi^2_{\nu_1}(1-\epsilon) \mid \theta_{02}\in\mathcal{C}_2(1-\zeta,\theta_{01})\right] \\
& \quad \times \lim_{n\to\infty} Pr_{\theta_{01}}\left[\,\theta_{02}\in\mathcal{C}_2(1-\zeta,\theta_{01})\right] \\
\le\ & 1 - \lim_{n\to\infty} Pr_{\theta_{01}}\left[R_1(\theta_{01},\theta_{02}) \le \chi^2_{\nu_1}(1-\epsilon)\right] \lim_{n\to\infty} Pr_{\theta_{01}}\left[\,\theta_{02}\in\mathcal{C}_2(1-\zeta,\theta_{01})\right] \\
\le\ & 1-(1-\epsilon)(1-\zeta) \\
\le\ & \epsilon+\zeta
\end{aligned}$$

using respectively Lemma A.4(iii) and the condition of the Theorem.

(ii) By the condition of the Theorem,

$$\begin{aligned}
& \lim_{n\to\infty} Pr_{\theta_{01}} \left[\{\mathcal{C}_2(1-\zeta,\theta_{*1}) = \varnothing\} \cup \{ \inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_{*1},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon)\} \right] \\
= & \lim_{n\to\infty} Pr_{\theta_{01}} \left[\mathcal{C}_2(1-\zeta,\theta_{*1}) = \varnothing\right] + \lim_{n\to\infty} Pr_{\theta_{01}} \left[\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_{*1},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon) \right] \\
= & \; 0 + \lim_{n\to\infty} Pr_{\theta_{01}} \left[\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_{*1},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon) \right].
\end{aligned}$$

Now note that it is also assumed that any $\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{*1})$ is in the $\sqrt{n}$-neighborhood of θ_{02}. Thus the value $\theta_2^{\inf}(\theta_{*1})$, where the infimum of $R_1(\theta_{*1},\theta_2)$ is attained, is also in the $\sqrt{n}$-neighborhood of θ_{02}. Hence using Lemma A.4(iii), we get

$$\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_{*1},\theta_{*2}) \equiv R_1(\theta_{*1},\theta_2^{\inf}(\theta_{*1})) = R_1(\theta_{*1},\theta_{02}) + o_p(1),$$

implying that

$$\begin{aligned}
& \lim_{n\to\infty} Pr_{\theta_{01}} \left[\{\mathcal{C}_2(1-\zeta,\theta_{*1}) = \varnothing\} \cup \{ \inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} R_1(\theta_{*1},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon)\} \right] \\
= & \lim_{n\to\infty} Pr_{\theta_{01}} \left[R_1(\theta_{*1},\theta_{02}) > \chi^2_{\nu_1}(1-\epsilon) \right]. \quad \blacksquare
\end{aligned}$$

Lemma A.5 *Let $\theta_{ni} = \theta_{0i} + d_i/\sqrt{n} \in \Theta$ and let $d_\theta := (d_1', d_2')'$ such that for $i = 1,2$, $d_i \in \mathbb{R}^{\nu_i}$ is fixed. For $i = 1,2$, let $L_i(\theta)$ and $\Psi_{i.g}$ be such that $L(\theta) = [L_1'(\theta), L_2'(\theta)]'$ and $\Psi_{\nabla.g} = \left[\Psi_{1.g}', \Psi_{2.g}'\right]'$. Define $\widehat{h}_{Ti}(\theta) = \sum_{t=1}^n \widehat{h}_{ti}(\theta)$ where $\widehat{h}_{ti}(\theta) = \left[vec\nabla_i g_t(\theta) - \widehat{V}_{ig}(\theta)\widehat{V}_{gg}^{-1}(\theta) g_t(\theta)\right]$ for $i = 1,2$. Under Assumptions M and W,*

$$\begin{bmatrix} n^{-1/2} g_T(\theta_n) \\ n^{-\delta_1}\widehat{h}_{T1}(\theta_n) \\ n^{-\delta_2}\widehat{h}_{T2}(\theta_n) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g + \sum_{i=1}^2 1_{[\delta_i=1]} M_i(\theta_{0i}) d_i \\ vecJ_1(\theta_0) + (1 - 1_{[\delta_1=1]})\left[\Psi_{1.g} + L_1(\theta_0)d_\theta\right] \\ vecJ_2(\theta_0) + (1 - 1_{[\delta_2=1]})\left[\Psi_{2.g} + L_2(\theta_0)d_\theta\right] \end{bmatrix}.$$

Proof: Define $V_{\nabla\nabla.g}(\theta) := V_{\nabla\nabla}(\theta) - V_{\nabla g}(\theta)V_{gg}^{-1}(\theta)V_{g\nabla}(\theta)$. Following the obvious partition with respect to θ_1 and θ_2, let $V_{\nabla g} = [V'_{1g}, V'_{2g}]'$, $V_{\nabla .g} = [V'_{1.g}, V'_{2.g}]'$ and for $i = 1, 2$, let $h_{Ti}(\theta) = \sum_{t=1}^{n} h_{ti}(\theta)$ where $h_{ti}(\theta) := \big[vec\nabla_i g_t(\theta) - V_{ig}(\theta_0)V_{gg}^{-1}(\theta_0)g_t(\theta)\big]$. Letting $h_T(\theta) = [h'_{T1}(\theta), h'_{T2}(\theta)]'$, Assumptions M and W give

$$\frac{1}{\sqrt{n}}\begin{bmatrix} g_T(\theta_0) \\ \hline h_T(\theta_0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g \\ \Psi_{\nabla .g} \end{bmatrix} \sim N\left[0, \begin{pmatrix} V_{gg}(\theta_0) & 0 \\ 0 & V_{\nabla\nabla .g}(\theta_0) \end{pmatrix}\right] \text{ and hence}$$

$$\frac{1}{\sqrt{n}} g_T(\theta_0) \xrightarrow{d} \Psi_g \quad \text{and for } i = 1, 2 \;\; \frac{1}{n^{\delta_i}} h_{Ti}(\theta_0) \xrightarrow{d} vecJ_i(\theta_0) + (1 - 1_{[\delta_i = 1]})\Psi_{i.g}. \quad \textbf{(A.2)}$$

A Mean-Value Expansion of $\frac{1}{\sqrt{n}} g_T(\theta_n)$ around θ_0 gives

$$\frac{1}{\sqrt{n}} g_T(\theta_n) = \frac{1}{\sqrt{n}} g_T(\theta_0) + \frac{1}{n}\nabla_\theta g_T(\bar{\theta})d_\theta$$

for some $\bar{\theta}$ such that $\|\bar{\theta} - \theta_0\| \leq \|\theta_n - \theta_0\| = O(1/\sqrt{n})$. Hence using (A.2) and Assumption W, we get

$$\frac{1}{\sqrt{n}} g_T(\theta_n) = \Psi_g + \sum_{i=1}^{2} 1_{[\delta_i = 1]} M_i(\theta_{0i})d_i + o_p(1).$$

Using Lemma A.1 and the fact that continuity is preserved by matrix inversion, for $i = 1, 2$,

$$\begin{aligned} \frac{1}{n^{\delta_i}} \widehat{h}_{Ti}(\theta_n) &= \frac{1}{n^{\delta_i}} \left[vec\nabla_i g_T(\theta_n) - \widehat{V}_{ig}(\theta_n)\widehat{V}_{gg}^{-1}(\theta_n)g_T(\theta_n)\right] \\ &= \frac{1}{n^{\delta_i}} \left[vec\nabla_i g_T(\theta_n) - V_{ig}(\theta_0)V_{gg}^{-1}(\theta_0)g_T(\theta_n)\right] + o_p(1) \\ &= \frac{1}{n^{\delta}} h_{Ti}(\theta_n) + o_p(1) \\ &= \frac{1}{n^{\delta_i}} h_{Ti}(\theta_0) + \frac{1}{n^{\delta_i + \frac{1}{2}}} \nabla_\theta h_{Ti}(\bar{\theta})d_\theta + o_p(1) \end{aligned}$$

for some $\bar{\theta}$ such that $\|\bar{\theta} - \theta_0\| \leq \|\theta_n - \theta_0\| = O(1/\sqrt{n})$ which follows from the Mean-Value Expansion of $h_{Ti}(\theta_n)$ around θ_0. Hence Assumption M2 and Lemma A.1

and (A.2) give for $i = 1, 2$,

$$\begin{aligned} \frac{1}{n^{\delta_i}}\widehat{h}_{Ti}(\theta_n) &= \frac{1}{n^{\delta_i}}h_{Ti}(\theta_0) + (1 - 1_{[\delta_i=1]})L_i(\theta_0)d_\theta + o_p(1) \\ &\xrightarrow{d} vecJ_i(\theta_0) + (1 - 1_{[\delta_i=1]})\left[\Psi_{i.g} + L_i(\theta_0)d_\theta\right] \quad \blacksquare \end{aligned}$$

Proof of Lemma 3.1: The following proof follows from Stock and Wright (2000) with negligible modifications. Noting that under WI-Cases II and IV,

(a) $E\frac{1}{n}g_T(\theta) = 1_{[\delta_1=1]}m_1(\theta_1) + 1_{[\delta_1=\frac{1}{2}]}\frac{1}{\sqrt{n}}\tilde{m}_{n1}(\theta) + m_2(\theta_2)$ where $m_1(\theta_1) \to m_1(\theta_{01})$ for $\theta_1 \to \theta_{01}$ and $\tilde{m}_{n1}(\theta) \to \tilde{m}_1(\theta)$ uniformly in $\theta \in \Theta$, and

(b) $\widehat{V}_{gg}^{-1}(\theta) \xrightarrow{P} V_{gg}^{-1}(\theta)$ uniformly where $V_{gg}^{-1}(\theta)$ is positive definite, continuous and bounded in $\theta \in \Theta$,

it follows from Assumption W that

$$\frac{1}{n}Q_n(\theta_{*1}, \theta_2) \xrightarrow{P} m_2'(\theta_2)V_{gg}^{-1}(\theta_{01}, \theta_2)m_2(\theta_2)$$

uniformly in $\theta_2 \in \Theta_2$. The right-hand side is zero iff $\theta_2 = \theta_{02}$ and hence continuity of the argmin operator gives

$$\tilde{\theta}_{n2}(\theta_{*1}) \xrightarrow{P} \theta_{02}. \tag{A.3}$$

Let $\tilde{\theta}_* = (\theta_{*1}', \tilde{\theta}_{n2}'(\theta_{*1}))'$ and $\theta_{*0} = (\theta_{*1}', \theta_{02}')'$. By definition of CUE $\tilde{\theta}_{n2}(\theta_{*1})$,

$$\begin{aligned} 0 &\geq Q_n(\tilde{\theta}_*) - Q_n(\theta_{*0}) \\ &= \left[\frac{\nabla_\theta g_T(\bar{\theta})}{\sqrt{n}}(\tilde{\theta}_* - \theta_0)\right]' \widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\left[\frac{\nabla_\theta g_T(\bar{\theta})}{\sqrt{n}}(\tilde{\theta}_* - \theta_0)\right] + \Delta_{1n} \\ &\quad +2\left[\frac{\nabla_\theta g_T(\bar{\theta})}{\sqrt{n}}(\tilde{\theta}_* - \theta_0)\right]' \widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\frac{g_T(\theta_0)}{\sqrt{n}} \end{aligned} \tag{A.4}$$

where the mean-value $\bar{\theta} \in \Theta$ is such that $\|\bar{\theta} - \theta_0\| \leq \|\tilde{\theta}_* - \theta_0\| = o_p(1)$ and

$$\Delta_{1n} = \frac{1}{\sqrt{n}} g_T'(\theta_0) \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \frac{1}{\sqrt{n}} g_T(\theta_0) - \frac{1}{\sqrt{n}} g_T'(\theta_{*0}) \widehat{V}_{gg}^{-1}(\theta_{*0}) \frac{1}{\sqrt{n}} g_T(\theta_{*0}).$$

For notational convenience let us define $\mathcal{M} \equiv \mathcal{M}(\tilde{\theta}_*, \bar{\theta}, \theta_0) := \left[\frac{\nabla_\theta g_T(\bar{\theta})}{\sqrt{n}} (\tilde{\theta}_* - \theta_0) \right]$ and let mineval(A) denote the minimum eigen-value of the matrix A. Note that,

(i) $\quad \mathcal{M}' \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \mathcal{M} \geq \|\mathcal{M}\|^2 \text{mineval} \left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \right)$

(ii) $\quad \mathcal{M}' \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \dfrac{g_T(\theta_0)}{\sqrt{n}} \geq -\|\mathcal{M}\| \left\| \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \dfrac{g_T(\theta_0)}{\sqrt{n}} \right\|$ (by the CS Inequality).

Define $\Delta_{2n} = \dfrac{\left\| \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \frac{g_T(\theta_0)}{\sqrt{n}} \right\|}{\text{mineval} \left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \right)}$ and $\Delta_{3n} = \dfrac{\Delta_{1n}}{\text{mineval} \left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \right)}$. Therefore, dividing (A.4) by mineval$\left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \right)$, we get,

$$\|\mathcal{M}\|^2 - 2\|\mathcal{M}\|\Delta_{2n} + \Delta_{3n} \leq 0$$

which implies that $\Delta_{2n} - \sqrt{\Delta_{2n}^2 - \Delta_{3n}} \leq \|\mathcal{M}\| \leq \Delta_{2n} + \sqrt{\Delta_{2n}^2 - \Delta_{3n}}$.

Noting that $\|\bar{\theta} - \theta_0\| = o(1)$, Assumptions M1 and W give $\dfrac{1}{n} \nabla_i g_T(\bar{\theta}) \to 1_{[\delta_i = 1]} M_i(\theta_{0i})$ for $i = 1, 2$. Since $\|d_1\| = O(1)$, it is clear that $\sqrt{n} \left(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02} \right) = O_p(1)$ if Δ_{2n} and Δ_{3n} are $O_p(1)$. Under Assumption M,

$$\Delta_{2n} \leq \frac{\sup_\theta \left\| \widehat{V}_{gg}^{-1}(\theta) g_T(\theta_0)/\sqrt{n} \right\|}{\inf_\theta \text{mineval} \left(\widehat{V}_{gg}^{-1}(\theta) \right)} \xrightarrow{d} \frac{\sup_\theta \left\| V_{gg}^{-1}(\theta) \Psi_g \right\|}{\inf_\theta \text{mineval} \left(V_{gg}^{-1}(\theta) \right)} = O_p(1). \qquad \textbf{(A.5)}$$

Again, noting that, for some $\bar{\theta}_{10} = (\bar{\theta}_1', \theta_{02}')'$ such that $\|\bar{\theta}_{10} - \theta_0\| \leq \|\theta_{*0} - \theta_0\| =$

$O(1/\sqrt{n})$, i.e. for some $\bar{\theta}_1 = \theta_{01} + \frac{\bar{d}_1}{\sqrt{n}}$ where $\|\bar{d}_1\| \leq \|d_1\| = O(1)$,

$$
\begin{aligned}
&|\Delta_{1n}| \\
= & \left| \frac{g_T'(\theta_0)}{\sqrt{n}} \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \frac{g_T(\theta_0)}{\sqrt{n}} - \left[\frac{g_T(\theta_0)}{\sqrt{n}} + \frac{\nabla_1 g_T(\bar{\theta}_{10})}{n} \bar{d}_1 \right]' \widehat{V}_{gg}^{-1}(\theta_{*0}) \left[\frac{g_T(\theta_0)}{\sqrt{n}} + \frac{\nabla_1 g_T(\bar{\theta}_{10})}{n} \bar{d}_1 \right] \right| \\
\leq & \left| \frac{g_T'(\theta_0)}{\sqrt{n}} \left[\widehat{V}_{gg}^{-1}(\tilde{\theta}_*) - \widehat{V}_{gg}^{-1}(\theta_{*0}) \right] \frac{g_T(\theta_0)}{\sqrt{n}} \right| + 2 \left| \frac{g_T'(\theta_0)}{\sqrt{n}} \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \frac{\nabla_1 g_T(\bar{\theta}_{10})}{n} \bar{d}_1 \right| \\
& + \left| \bar{d}_1' \frac{\nabla_1 g_T'(\bar{\theta}_{10})}{n} \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \frac{\nabla_1 g_T(\bar{\theta}_{10})}{n} \bar{d}_1 \right| \\
\xrightarrow{d} & \; 2 \times 1_{[\delta_1=1]} \left[\left| \Psi_g V_{gg}^{-1}(\theta_0) M_1(\theta_{01}) \bar{d}_1 \right| + \left| \bar{d}_1' M_1'(\theta_{01}) V_{gg}^{-1}(\theta_0) M_1(\theta_{01}) \bar{d}_1 \right| \right] = O_p(1)
\end{aligned}
$$

follows from Assumptions M and W. Again, since $V_{gg}^{-1}(\theta)$ is positive definite, similar arguments as in (A.5) give $|\Delta_{3n}| = O_p(1)$.

Therefore, $\sqrt{n}\|\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}\| = O_p(1)$. ∎

Proof of Lemma 3.2: Define $A_{ni} = \widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_T)\widehat{D}_{Ti}(\theta_n)$ for $i = 1, 2$. It follows from Lemma A.5 that $\frac{1}{n^{\delta_i}} A_{ni} \xrightarrow{d} V_{gg}^{-\frac{1}{2}'}(\theta_0) \left[J_i(\theta_0) + (1 - 1_{[\delta_i=1]}) devec_k \left[\Psi_{i.g} + L_i(\theta_0) d_\theta \right] \right]$ and

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \left(A_{n1}' N(A_{n2}) A_{n1} \right)^{-\frac{1}{2}'} A_{n1}' N(A_{n2}) \widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n) g_T(\theta_n) \\
& \qquad \xrightarrow{d} \left(\mathbb{A}_1' N(\mathbb{A}_2) \mathbb{A}_1 \right)^{-\frac{1}{2}'} \mathbb{A}_1' N(\mathbb{A}_2) \mathbb{B}, \qquad \text{(A.6)} \\
& \frac{1}{\sqrt{n}} \left(A_{n2}' A_{n2} \right)^{-\frac{1}{2}'} A_{n2}' \widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n) g_T(\theta_n) \xrightarrow{d} \left(\mathbb{A}_2' \mathbb{A}_2 \right)^{-\frac{1}{2}'} \mathbb{A}_2' \mathbb{B}, \text{ and} \qquad \text{(A.7)} \\
& \frac{1}{\sqrt{n}} \widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n) g_T(\theta_n) \xrightarrow{d} \mathbb{B} \qquad \text{(A.8)}
\end{aligned}
$$

Lemma 3.2 follows directly from (A.6), (A.7) and (A.8). ∎

Proof of Theorem 3.3: (i) From Lemma 3.2, it is clear that $\mathcal{C}_2(1-\zeta, \theta_{01})$ is a $1-\zeta$ joint confidence region for θ_2 when θ_1 is known *a priori* to be θ_{01}, and hence it contains θ_{02} with probability $1-\zeta$. Using the same strategy as in the proof of

Theorem 2.1, we have

$$\begin{aligned}&\left\{\mathcal{C}_{\theta_2}(1-\zeta,\theta_{01})=\varnothing,\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})}K_{n1}(\theta_{01},\theta_{*2})>\chi^2_{\nu_1}(1-\epsilon)\right\}\\ \subseteq\ &\left\{\theta_{02}\notin\mathcal{C}_2(1-\zeta,\theta_{01}),K_{n1}(\theta_{01},\theta_{02})>\chi^2_{\nu_1}(1-\epsilon)\right\}.\end{aligned}$$

Hence noting that Lemma 3.2 gives $K_{n1}(\theta_0)\xrightarrow{d}\chi^2_{\nu_1}$, standard Bonferroni arguments give the asymptotic size to be at most $\epsilon+\zeta$.

(ii) Lemma 3.2 implies that in WI-Cases II and IV, $\mathcal{C}_2(1-\epsilon,\theta_{*1})$ is contained in the $\sqrt{n}$-neighborhood of θ_{02} w.p.a.1 under the conditions of the Theorem. Hence $\theta_2^{\inf}(\theta_{*1})$, where the infimum $\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})}K_{n1}(\theta_{*1},\theta_{*2})$ is attained, is also in the $\sqrt{n}$-neighborhood of θ_{02}. Hence Lemma 3.2 directly applies and gives the asymptotic equivalence of the tests. ∎

Lemma A.6 *Let* $\tilde\theta_*=(\theta'_{*1},\tilde\theta'_{n2}(\theta_{*1}))'$ *where* $\tilde\theta_{n2}(\theta_{*1})$ *is such that* $\nabla_2 Q_n(\tilde\theta_*)=0$ *for* $\theta_{*1}\in\Theta_1$. *Then under Assumptions M and W,*

$$Pr_{\theta_{01}}\left[\inf_{\theta_{*2}\in\Theta_2}K_n(\theta_{*1},\theta_{*2})>\chi^2_\nu(1-\epsilon)\right]\le P_{\theta_{01}}\left[K_{n1}\left(\theta_{*1},\tilde\theta_{n2}(\theta_{*1})\right)>\chi^2_{\nu_1}(1-\epsilon)\right].$$

Proof: The result follows once we note that

$$\begin{aligned}Pr_{\theta_{01}}\left[\inf_{\theta_{*2}\in\Theta_2}K_n(\theta_{*1},\theta_{*2})>\chi^2_\nu(1-\epsilon)\right]&\le Pr_{\theta_{01}}\left[K_n\left(\theta_{*1},\tilde\theta_{n2}(\theta_{*1})\right)>\chi^2_\nu(1-\epsilon)\right]\\&= Pr_{\theta_{01}}\left[K_{n1}\left(\theta_{*1},\tilde\theta_{n2}(\theta_{*1})\right)>\chi^2_\nu(1-\epsilon)\right]\\&\le Pr_{\theta_{01}}\left[K_{n1}\left(\theta_{*1},\tilde\theta_{n2}(\theta_{*1})\right)>\chi^2_{\nu_1}(1-\epsilon)\right].\end{aligned}$$

∎

VITA

Saraswata Chaudhuri was born on February 12, 1978, in Calcutta, India. He received the Bachelor's Degree with Honors in Statistics from Presidency College, Calcutta University in 1999, and the Master's Degree in Statistics from the Indian Statistical Institute in 2001. From July 2001 to August 2003 he worked with various institutions in India as an Actuarial Analyst, a Business-Consultant, a Software Engineer and a Research Assistant. In September 2003, he enrolled as a graduate student in Economics at the University of Washington and in June 2008 he graduated with a Doctor of Philosophy in Economics. He has accepted a tenure track position in the Economics Department at the University of North Carolina - Chapel Hill to begin in July 2008.