ABSTRACT

SHIN, CHANGMOCK. Entropy Based Moment Selection in Generalized Method of Moments. (Under the direction of Professor Alastair R. Hall).

GMM provides a computationally convenient estimation method and the resulting estimator can be shown to be consistent and asymptotically normal under the fairly moderate regularity conditions. It is widely known that the information content in the population moment condition has impacts on the quality of the asymptotic approximation to finite sample behavior. This dissertation focuses on a moment selection procedure that leads us to choose relevant (asymptotically efficient and non-redundant) moment conditions in the presence of weak identification. The contributions of this dissertation can be characterized as follows: in the framework of linear model, (i) the concept of nearly redundant moment conditions is introduced and the connection between near redundancy and weak identification is explored; (ii) performance of $RMSC(c)$ is evaluated when weak identification is a possibility but the parameter vector to be estimated is not weakly identified by the candidate set of moment conditions; (iii) performance of $RMSC(c)$ is also evaluated when the parameter vector is weakly identified by the candidate set; (iv) a combined strategy of Stock and Yogo's (2002) test for weak identification and $RMSC(c)$ is introduced and evaluated; (v) (i) and (ii) are extended to allow for nonlinear dynamic models. The subsequent simulation results support the analytical findings: when only a part of instruments in the set of possible candidates for instruments are relevant and the others are redundant given all or some of the relevant ones, $RMSC(c)$ chooses all the relevant instruments with high probabilities and improves the quality of the post-selection inferences; when the candidates are in order of their importance, a combined strategy of Stock and

Yogo (2002) pretest and $RMSC(c)$ improves the post-selection inferences, however it tends to select parsimonious models; when all the possible candidates are equally important, it seems that $RMSC(c)$ does not provide any merits. However, in the last case, asymptotic efficiency and non-redundancy can be achieved by basing the estimation and inference on all the possible candidates.

Entropy Based Moment Selection in Generalized Method of Moments

by

Changmock Shin

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial satisfaction of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh

2005

Approved By:

Peter Bloomfield

Dr. Peter Bloomfield

Denis Pelletier

Dr. Denis Pelletier

Alastair Hall

Dr. Alastair R. Hall
Chair of Advisory Committee

Atsushi Inoue

Dr. Atsushi Inoue

This dissertation is dedicated

to my God the Father who art in Heaven.

This dissertation is also dedicated

to my parents who give me every opportunity to realize my dreams,

and to my wife Seungmee who is always with me.

Biography

Changmock Shin

Contact:

103 Arbor Way 919-467-6237

Cary, NC 27513 cshin626@gmail.com

Birth: June 26, 1967. (Seoul, Korea)

Degrees:

1998, M.A., Economics, Duke University, Durhan, NC. U.S.A.

1993, M.A., Economics, Korea University, Seoul, Korea.

1990, B.A., Economics, Korea University, Seoul, Korea.

Important Papers:

- "Entropy Based Moment Selection in the Presence of Weak Identification" (paper with Alastair Hall and Atushi Inoue), A Conference in Honor of George Judge The 2nd Conference on Information and Entropy Econometrics: Theory, Method, and Applications, September, 23–25, 2005
- "Information in Generalized Method of Moments Estimation and Entropy based Moment Selection" (paper with Alastair Hall, Atushi Inoue, and Kalidas Jana), 2005, accepted for publication in *Journal of Econometrics*.

Skills:

Language: English (fluent), Korean(native)

Computer: Matlab, GNU Octave, Gauss, Splus, R, OxMetrics, Stata, LaTeX.

Other: Freelance Photojournalist for "Evangelizing Todays Child", Child Evangelism Fellowship of Korea (Magazine in Korean).

Acknowledgements

Hidden is a story deep down inside the pages of this dissertation; it is a story of a long journey filled with adventures, challenges, efforts, sacrifices, supports, encouragements and loves. Without these ingredients, I would never be able to finish this work.

I would like to express my deepest gratitude to my advisor Dr. Alastair R. Hall for his technical and editorial advice, for his expert guidance and mentorship, and for his encouragement and support at all levels. These were essential to the completion of this dissertation.

I would like to express my sincere thankfulness to the members of my committee: Dr. Atushi Inoue, Dr. Peter Bloomfield and Dr. Denis Pelletier for providing many constructive comments that improved the contents of this dissertation as well as my knowledge on the topic. I would also like to express my sincere thanks to Dr. Matt Holt for attending my final oral exam and for giving me valuable comments.

I would like to express my special thanks to the friendship of Jeongbeom Ma, Woon Chang, Minje Sung, Sunam Park, Jinyoung Oh, Junghwan Hyun, Kyunghwan Lee, Kwangmin Choi, Yunkyung Lee, Choongbae Kim and many others whose names are not listed here. The fragrant memories with them gave me the strength to pass over the difficult times. I would also like to give my special thanks to Sanghyun Oh for his proofreading the draft and for spotting many typos and editorial errors that I made.

I would like to express my warmest gratitude and love to my parents for their everlasting and unconditional patience, encouragements, supports, love and trust in me. Without these I am nothing. I would also like to express my profound thankfulness to my sisters and their families for leading me to the right way whenever I have gone astray and for their loves and supports. I would also like to give my very special gratitude to my parents-in-law and their families for their patience and supports and for their prayers for me and my family.

Last, but not least, with every little beat of my heart I would like to give my love and thanks to my wife Seungmee for being always with me, for being so patient, for listening to my complaints and frustrations, for believing in me and for loving me. I owe my every achievement to her. I also want to give my greatest love and thanks to my daughter Yujin and my son Hyunjin for giving me their cries and laughs that teach me how much I love them and that give me the strength to go through bad and good times. They are the origin of my happiness.

Above all of these, I would like to give thanks and glory to God who was, is, and will be with me. I must confess that I am not the one who finish this dissertation. It is Him who works in me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Many statistical techniques were invented in the nineteenth century. Unfortunately, however, in many circumstances it was computationally infeasible to apply the new statistical techniques due to the limits in the computing technology. With the advent of high-speed electronic computing and programming languages, the practice of applied econometrics draws heavily on the methodology known as Maximum Likelihood Estimation (MLE) which is known as the best available estimator within the classical statistics paradigm. For *large* samples (or *asymptotically*), Maximum Likelihood Estimators (MLEs) are optimal in the sense that: (i) MLEs are *asymptotically normally distributed*; (ii) MLEs have *asymptotically minimum variance*; (iii) MLEs are *asymptotically unbiased*[1].

These desirable properties of MLEs stem from the method's basis on the joint probability distribution of data, which is usually the likelihood function. The Likelihood Principle states that all the relevant information in the sample is contained in the likelihood function. This dependency of MLEs on likelihood function can cause a

[1] MLEs are often biased, but the bias vanishes as sample size goes to infinity

weakness because it requires the knowledge about the true joint probability distribution of data. In other words, the desirable statistical properties of MLEs are obtained only when the distribution is correctly specified. Another shortcoming of MLEs is that for many instances the MLE would be computationally very burdensome even with the highly developed computing capabilities.

In contrast, Generalized Method of Moments (GMM) is a flexible econometric procedure that can be applied under quite mild assumptions for the probabilistic behavior of the data generating process. Instead of requiring the complete knowledge of the true specification of the likelihood function, GMM estimators are attained based on a set of population moment conditions. Differently put, GMM finds the estimates of parameters by matching the moments of the model to those of data as close as possible. In this aspect, GMM shares common features with the conventional Method of Moments (MM). However, unlike MM, GMM is also applicable in the case where the number of moments conditions exceeds that of parameters to be estimated. If the population moment conditions are true, the resulting estimators can be shown to be consistent and asymptotically normal under fairly weak regularity conditions. Another advantage of GMM is that it can be obtained with fair computational burden when MLE requires a huge computational burden.

In most cases of interest, the researcher is actually faced with a candidate set of moment conditions and the estimators are based on researcher's choice of the moment conditions from the candidate set. Therefore, the choice should reflect the information content of the moment conditions relative to the desired inferences. It is desirable for the selected moment condition to satisfy some statistical criteria.[2] In a substantial

[2]One of the perspectives on the statistical criteria for selecting moment conditions can be found in Hall (2005, Section 7.1). Those criteria are identification conditions, orthogonality conditions, the efficiency conditions and non-redundancy conditions. Some other criteria can be found in Donald

body of literature, it has reported that the inclusion of the moment conditions with no incremental or insufficient information can cause a deterioration in the quality of the asymptotic approximation to finite sample behavior. These information scenarios are known as *redundant moment conditions* and *weak identification*, respectively. Breusch, Qian, Schmidt, and Wyhowski (1999) introduce the concept of redundant moment condition to depict the circumstance where the inclusion of additional set of moment conditions has no impact on the asymptotic variance of the GMM estimator. In their Monte Carlo simulation studies, Hall and Peixe (2003) and Hall and Inoue (2003) show that inclusion of redundant moment conditions can deteriorate the quality of the limiting distribution as an approximation to finite sample behavior. Some earlier studies provide the similar evidence but the term *redundancy* is not used since they predated the Breusch, Qian, Schmidt, and Wyhowski's (1999) paper. Many other studies are focused on the weak identification in the settings of GMM or IV which is a special case of GMM. For example, see Nelson and Startz (1990), Hall, Rudebusch, and Wilcox (1996), Staiger and Stock (1997), Stock and Wright (2000), Nelson, Startz, and Zivot (2000), Stock and Yogo (2002), Stock, Wright, and Yogo (2002). These studies are dealing with the detection of weak identification and solutions for avoiding the potential problem of weak identification.

Noting the potential problems of the low quality of information content in the moment conditions, Hall and Inoue (2003) develop an information criterion for the selection of the set of relevant moment conditions. Hall and Inoue's (2003) method rests on the long run canonical correlations between the population moment condition used in the estimation and the unknown true score vector associated with the

and Newey (2001). They develop a criterion to choose instruments that minimize the approximate mean squared errors of the estimators within the settings of 2SLS, LIML and a bias adjusted version of 2SLS.

data. In this dissertation, we extend Hall and Inoue's (2003) Relevant Moment Selection Criterion ($RMSC(c)$) and show that even in the presence of weak identification $RMSC(c)$ is useful to select the relevant moment conditions form a finite dimensional candidate set and to improve the quality of asymptotic approximation to the finite sample behavior in GMM estimation. This moment selection procedure $RMSC(c)$ is based on the finding that the entropy of the limiting distribution of GMM estimator can provide a useful metric for the information content in the population moment conditions. The contributions of this dissertation can be characterized as follows: in the framework of linear model, (i) the concept of nearly redundant moment conditions are introduced and the connection between near redundancy and weak identification is explored; (ii) performance of $RMSC(c)$ is evaluated when weak identification is a possibility but the parameter vector to be estimated is not weakly identified by the candidate set of moment conditions; (iii) performance of $RMSC(c)$ is also evaluated when the parameter vector is weakly identified by the candidate set; (iv) a combined strategy of Stock and Yogo's (2002) test for weak identification and $RMSC(c)$ is introduced and evaluated; (v) (i) and (ii) are extended to allow for a class of nonlinear dynamic models known as Generalize Instrument Variable estimation.

The rest of this dissertation is organized as follows. In Chapter 2, we review GMM estimation procedure and its major results. In that chapter, we also introduce the entropy based moment selection criterion and establish its consistency. In Chapter 3, we reveal the behavior of the relevant moment selection criterion in the linear IV model setting and establish the consistency results of the moment selection procedure when weak identification is a possibility. In Chapter 4, we we generalize the results in Chapter 3 and shows the main results therein are still valid in the nonlinear settings as well. Chapter 5 provides the simulation design and the results. Finally, Chapter 6 provides concluding remarks and directions for future works.

Chapter 2

GMM and Entropy based Moment Selection

In this chapter, we review the basic concept and main results of Generalized Method of Moments (GMM) estimation; we also review the concepts of redundancy and weak identification. It is also shown that entropy of the limiting distribution of GMM can be used as a useful metric for the information content in GMM. We also describe the entropy based moment selection criterion proposed by Hall and Inoue (2003) and explore its statistical properties.

2.1 Generalized Method of Moments Estimation

GMM provides a framework for the estimation of an unknown parameter vector of interest based only on a set of moment conditions. Unlike Maximum Likelihood estimation, GMM does not require the complete knowledge about statistical environment of the data generating process. The main idea behind GMM is to find the estimator so that the sample orthogonality condition implied by the underlying economic/econometric theory as close as possible to its population counterpart. Due to this feature, GMM is widely used in a variety of settings, especially in complex models such as nonlinear dynamic models.[1] This method was proposed by Hansen (1982). GMM is quite a general principle and it includes as special cases Least Squares, Instrumental Variables (IV) and Maximum Likelihood estimations.

Throughout our discussion of GMM in its most general form, we consider the case in which the data satisfy the following condition.

Assumption 2.1. $\{v_t \in \mathcal{V}, t = 1, 2, \ldots\}$ *is a sequence of strictly stationary and ergodic random vectors where* $\mathcal{V} \subseteq \Re^s$.

The basis of the GMM estimation of the unknown $p \times 1$ parameter vector θ_0 is a set of population moment conditions of the form

$$E[f(v_t, \theta_0)] = 0, \quad \text{for all } t \tag{2.1}$$

where $f : \mathcal{V} \times \Theta \to \Re^q$ and $q \geq p$. This moment function f is assumed to satisfy the following regularity conditions.

Assumption 2.2. *The function* $f : \mathcal{V} \times \Theta \to \Re^q$ *satisties: (i) it is continuous on* Θ *for each* v*; (ii)* $E[f(v_t, \theta)]$ *exitsts and is finite for every* $\theta \in \Theta$*; (iii)* $E[f(v_t, \theta)]$ *is continuous on* Θ.

[1] Examples of the application of GMM are well documented in Hall (2005, Chapter 1).

In practice, the population moment conditions are constructed according to the underlying economic/econometric models and are assumed to be valid when evaluated at the true parameter vector θ_0. However, it should be noted that the population moment condition can only be useful as a basis for estimation when it can provides enough information for the identification of the parameters of interest. The basic condition for parameter identification is given by the following global identification condition.

Assumption 2.3 (Global Identification). $E[f(v_t, \bar{\theta})] \neq 0$ *for all* $\bar{\theta} \in \Theta$ *such that* $\bar{\theta} \neq \theta_0$.

This *global* identification condition states that the population moment condition only holds at one value in the entire parameter space Θ. In the context of linear model, a convenient condition for global identification can be easily derived. Unfortunately, however, it is seldom possible in the context of nonlinear model. By this reason, the researchers often limit their attention to some suitably defined neighborhood of θ_0 and use the concept of *local* identification. To derived the condition for local identification, we need to impose the following regularity conditions on $\frac{\partial f(v_t,\theta)}{\partial \theta'}$.

Assumption 2.4. *(i) The derivative matrix* $\frac{\partial f(v_t,\theta)}{\partial \theta'}$ *exists and is continuous on* Θ *for each* $v \in \mathcal{V}$*; (ii)* θ_0 *is an interior point of* Θ*; (iii)* $E\left[\frac{\partial f(v_t,\theta)}{\partial \theta'}\right]$ *exists and is finite.*

Under these regularity conditions, we can derive the condition for local identification by restricting attention to a sufficiently small ϵ neighborhood of θ_0 such that the following first order Taylor series expansion is exact,

$$f(v_t, \theta) = f(v_t, \theta_0) + \left\{ \frac{\partial f(v_t, \theta_0)}{\partial \theta'} \right\} (\theta - \theta_0) \tag{2.2}$$

Taking expectations on both sides of (2.2) and using the population moment condition yields

$$E[f(v_t, \theta)] = E\left[\frac{\partial f(v_t, \theta_0)}{\partial \theta'} \right] (\theta - \theta_0) \tag{2.3}$$

From (2.3) we can deduce the following local identification condition.

Assumption 2.5 (Local Identification)**.** rank $\left\{ E\left[\frac{\partial f(v_t, \theta_0)}{\partial \theta'} \right] \right\} = p$.

If this local identification condition holds, in some suitably defined neighborhood of θ_0, $E[f(v_t, \theta)] = 0$ only at $\theta = \theta_0$. It is worth noting that in nonlinear models identification may be sensitive to the value of θ_0. That is, it may be possible that the population moment condition provides enough information for the identification of parameter to be estimated at some θ_0 but not at any others.[2] It should be also noted that the fact that rank $\left\{ E\left[\frac{\partial f(v_t, \theta_0)}{\partial \theta'} \right] \right\} \leq \min\{p, q\}$ and the condition for local identification implies that $p \leq q$, that is we need to have at least as many as moment conditions as parameters of interest.

The GMM estimator of θ_0, $\hat{\theta}_T$ is defined to be:

$$\begin{aligned} \hat{\theta}_T(f) &= \text{Argmin}_{\theta \in \Theta} \; Q_T(\theta) \\ &= \text{Argmin}_{\theta \in \Theta} \; g_T(\theta)' \, W_T \, g_T(\theta) \end{aligned} \tag{2.4}$$

where $g_T(\theta) = T^{-1} \sum_{t=1}^{T} f(v_t, \theta)$ and W_T is a positive semi–definite weighting matrix which converges in probability to $S(f)^{-1}$ where

$$S(f) \;=\; \lim_{T \to \infty} Var[T^{1/2} g_T(\theta_0)] \tag{2.5}$$

[2]Examples can be found in Hall (2005, Chapter 3).

Under certain regularity conditions it can be shown that the estimator is consistent for θ_0 and asymptotic normal. These results are well documented in Hansen (1982) or Hall (2005, Chapter 3). Here, we just introduce the basic results in the following assumption.

Assumption 2.6 (Asymptotic Normality)**.** *$T^{1/2}[\hat{\theta}_T(f) - \theta_0] \xrightarrow{d} N(0, V_\theta(f))$ where $V_\theta(f) = [G(f)'S(f)^{-1}G(f)]^{-1}$ and $G(f) = E[\partial f(v_t, \theta_0)/\partial\theta']$.*

For the purpose of this dissertation, we will focus on the class of GMM estimators known as Generalized Instrumental Variables (GIV) estimators which is introduced by Hansen and Singleton (1982). In this framework, the population moment condition is expressed as the statistical orthogonality condition between two vectors, $u_t(\theta_0)$ and $z_{t-\delta}$. The vector $u_t(\theta_0)$ consists of functions of the data and the unknown parameter vector and the following conditional moment condition is satisfied.

$$E[u_t(\theta_0) \,|\, \mathbf{\Omega}_{t-\delta}] = 0 \tag{2.6}$$

where $\mathbf{\Omega}_{t-\delta}$ is the information set at time period $t-\delta$ for some non-negative integer δ. By applying the law of iterated expectation, we have the following population moment condition.

$$E[z_{t-\delta} \otimes u_t(\theta_0)] = 0 \tag{2.7}$$

for any $z_{t-\delta} \in \mathbf{\Omega}_{t-\delta}$. The vector $z_{t-\delta}$ is known as the instrument vector. One convenient feature of this GIV estimation is that the choice of moment condition is reduced to the choice of instrument vector.

2.2 Relevant Moment Selection Criterion

To apply the GMM estimation in practice, researchers usually face a set of candidates from which to choose the elements of the population moment condition. It is desirable to select the elements of the population moment conditions based on statistical criteria that are devised to choose the valid moment conditions. The literature on the information content of population moment conditions has focused on three particular scenarios: optimal moment conditions, redundant moment conditions and weakly identified moment conditions. From a view point of information content, the optimal moment conditions provides the maximum information; redundant moment conditions provide no additional information; weak identified moment conditions provide insufficient information about the parameters to be estimated. Among these three particular scenarios, we focus on the second and the third ones and consider the Relevant Moment Selection Criterion proposed by Hall and Inoue (2003). $RMSC(c)$ leads us to choose the moment conditions that satisfy asymptotic efficiency and non-redundancy conditions.

2.2.1 Desirable Properties of Moment Conditions

It is useful to begin with the description of desirable properties of selected moment conditions. These desirable properties depend on the statistical criteria for selection of desirable moment conditions. These criteria should reflect the purpose of the analysis. Following Hall's (2005, Chapter 7) perspective, we assume that the objective of the analysis is to make inferences about θ_0 based on the asymptotic distribution theory summarized in the previous section. From this perspective, Hall (2005, Section 7.1) introduces four desirable properties for the selected moment conditions: orthogonal condition, efficiency condition, non-redundancy condition and relevance condition.

To consider the problem of moment selection and their desirable properties, we introduced some new notations. It is natural to assume that the candidate set of scalar functions which can form the basis for the population moment condition is finite. Let $q_{max} \times 1$ vector $f_{max}(\cdot)$ be a stack of these scalar functions. Following Andrews (1999), we use a $q_{max} \times 1$ selection vector c ane we now index $f(\cdot)$ by c. $c_i = 1$ implies the i^{th} element of $f_{max}(\cdot)$ is included in $f(\cdot; c)$, and $c_i = 0$ implies this element is excluded. Note that $|c| = c'c$ equals the number of elements in $f(\cdot; c)$. The set of all possible selection vectors C is defined as follows:

$$C \; = \{ \, c \in \Re^{q_{max}}; \; c_i \; = \; 0, 1, \; \text{for } i = 1, 2 \ldots q_{max}, \;\; \text{and} \;\; c = (c_1, \ldots c_{q_{max}})', \; |c| \geq p \, \} \tag{2.8}$$

Through out this dissertation, we assume that the objective of the estimation is to make inferences about the true parameter vector θ_0 based on the two-step or iterated GMM estimation and its asymptotic distribution theory. Now, let $\hat{\theta}_T(c)$ denote the GMM estimator based on $E[f(v_t, \theta_0; c)] = 0$ and let $V_\theta(c)$ denote the asymptotic variance of $\hat{\theta}_T(c)$. This asymptotic variance can be written as

$$V_\theta(c) = \left[G_0(c)' S(c)^{-1} G_0(c) \right]^{-1} \tag{2.9}$$

where $G_0(c) = E\left[\frac{\partial f(v_t, \theta_0; c)}{\partial \theta'}\right]$ and $S(c) = \lim_{T \to \infty} Var\left[T^{-1/2} \sum_{t=1}^{T} f(v_t, \theta_0; c)\right]$. Applying the asymptotic normality results in Assumption 2.6, the limiting distribution of $\hat{\theta}_T(c)$ can be written as follows.

$$T^{1/2}[\hat{\theta}_T(c) - \theta_0] \xrightarrow{d} N(0, V_\theta(c)) \tag{2.10}$$

By considering the first and second moment properties of this asymptotic distribution and the quality of this distribution as an approximation to finite sample behavior, we introduce the following three desirable properties of the selected moment conditions.

First of all, it is desirable that the selected moment conditions satisfy the following orthogonality condition.

Definition 2.1 (Orthogonality Condition)**.** *Let c_{sel} denote the selection vector chosen by the researcher. The selected moment condition satisfies population moment condition in (2.1), that is* $E[f(v_t, \theta_0; c_{sel})] = 0$.

The consistency of $\hat{\theta}_T(c)$ is embedded in the limiting distribution of (2.10), since it has a zero mean. The population moment condition plays an important role in derivation of this result. This fact leads to the need of the orthogonality condition. If this condition is satisfied, the limiting distribution has the desirable first moment properties. Also note that it is clearly desirable to base the inference on the moment conditions that yield the smallest variance in a matrix sense. This leads to the following efficiency conditions.

Definition 2.2 (Efficiency Condition)**.** *The selected moment condition is efficient, that is $V_\theta(c) - V_\theta(c_{sel})$ is positive semi-definite for all $c \in C$ such that* $E[f(v_t, \theta_0; c_{sel})] = 0$.

If efficiency condition is satisfied, we can say that the limiting distribution has the desirable second moment properties. However, it should be noticed that by construction increasing the number of moment conditions can never increase the asymptotic variance of the GMM estimator. The proof of this argument can be found for example in

Hall (2005, Theorem 6.1). This means that if the moment selection procedure is based only on the orthogonality and efficiency condition, the optimal selection strategy is to choose as many *correct* moment conditions[3] as possible. However, there exists a situation where the augmentation of the population moment condition has no effect on the asymptotic variance of the GMM estimator. To depict this circumstance, Breusch, Qian, Schmidt, and Wyhowski (1999) introduce the term redundancy. In other words, the augmented set of moment conditions provides no further information for the estimation. The formal definitions of redundancy and non-redundancy are as follows.

Redundancy: Suppose that $f(v_t, \theta) = [f_1(v_t, \theta)', \, f_2(v_t, \theta)']'$ then $E[f_2(v_t, \theta_0)] = 0$ is said to be redundant for θ_0 given $E[f_1(v_t, \theta_0)] = 0$ if $V(f) = V(f_1)$. Therefore if $E[f_2(v_t, \theta_0)] = 0$ is redundant given $E[f_1(v_t, \theta_0)] = 0$ then it provides no information about θ_0 beyond that already in $E[f_1(v_t, \theta_0)] = 0$.

Non-redundancy: The converse of redundancy is termed *non*-redundancy. If the moment condition $E[f_2(v_t, \theta_0)] = 0$ is non-redundant given $E[f_1(v_t, \theta_0)] = 0$ then $V(f_1) - V(f)$ is positive semi-definite and so $E[f_2(v_t, \theta_0)] = 0$ provides additional information.

By definition, the inclusion of redundant moment conditions does not affect the asymptotic properties of the GMM estimator. However, their inclusion can lead to a serious deterioration in the quality of the asymptotic approximation to finite sample

[3]This term is used by Andrews (1999) to denote the moment conditions that satisfy the orthogonality condition.

behavior. This deterioration is explicitly demonstrated in Hall and Peixe (2003). This observation leads us to the following non-redundancy condition.

Definition 2.3 (Non-Redundancy Condition)**.** *No individual element of the selected moment condition* $E[f(v_t, \theta_0; c_{sel})] = 0$ *is redundant for the estimation of* θ_0 *given the remaining elements.*

The discussion above motivates Hall (2005) to argue that it is desirable that researchers wish to base the inference on the subset of the available moment conditions which are asymptotically efficient but contain no redundant moment conditions. For ease of exposition, Hall (2005) uses the term *relevant* moment conditions to denote the subset of the available moment conditions which are asymptotically efficient but contain no redundant moment conditions. A formal definition of relevance follows.

Definition 2.4 (Selection Vector Associated with Relevance Condition)**.** c_r *is the selection vector associated with the relevant moment conditions if the following three properties hold: (i)* $c_r \in C$*; (ii)* $V_\theta(\iota_{q_{max}}) = V_\theta(c_r)$ *where* $\iota_{q_{max}}$ *is a* $q_{max} \times 1$ *vector of ones; (iii)* $V_\theta(c_{r,1}) - V_\theta(c_r)$ *is positive semi–definite for* $c_r = c_{r,1} + c_{r,2}$ *and* $c_{r,1} \in C$*.*

A few observations about this definition are in order. Since the inclusion of redundant moment conditions has no asymptotic cost, researchers can always obtain the asymptotic efficiency by using the complete candidate set as the moment conditions in the estimation. Part *(ii)* implies that the asymptotic efficiency is also achieved by the estimation based on the relevant subset. Note here that asymptotic efficiency is relative to all possible choices of moment condition from the candidate set. An important implication of this property is that if $c_r \neq \iota_{q_{max}}$ then all remaining elements of the candidate set are redundant given the relevant subset that is, $V_\theta(c_r) = V_\theta(c_r + c_i)$ for $c_r' c_i = 0$ and $(c_r + c_i) \in C$.

2.2.2 Entropy and Moment Selection in GMM Estimation

In this subsection, we show that the entropy of limiting distribution of $\hat{\theta}_T(f)$ provides a continuous measure of the information content in the population moment condition. Then we describe the entropy based moment selection criterion and establish the consistency of this moment selection criterion. Here, the term *consistency* means that the selection criterion will choose the relevant vector of moment conditions with probability one as the sample size grows.

Entropy of a vector of continuous random variables X with a probability distribution P is defined by

$$ent(X) \equiv ent(p(x)) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \ln p(x)\, dP(x) \tag{2.11}$$

where $p(x) = dP(x)$ is the probability density function for the absolutely continuous distribution P. Entropy is a measure of disparity or uncertainty of the density $p(x)$. Entropy is used in econometrics for analyzing finite or limited data sets as well as ill-conditioned or ill-behaved data sets.[4] Within the context of our discussion, the data set can be viewed as ill-conditioned or ill-behaved in the presence of (near) redundancy or weak identification. Ahmed and Gokhale (1989) derive the point estimation of

[4]Current development of information and entropy econometrics are well treated in Golan (2002).

entropy for the multivariate normal distribution, denoted by $ent(NORM)$, as follows:

$$ent(NORM) = 0.5p\,[1 + \ln(2\pi)] \;+\; 0.5\,\ln\big|\Sigma\,\big| \tag{2.12}$$

where Σ is the variance-covariance matrix. Applying their result to the limiting distribution of $\hat{\theta}_T(f)$ in Assumption 2.6, the entropy of this distribution can be written as

$$ent_\theta(f) \;=\; 0.5p\,\Big[1 + \ln(2\pi)\Big] \;-\; 0.5\,\ln\Big[\Big|G(f)'S(f)^{-1}G(f)\Big|\Big] \tag{2.13}$$

One feature of the entropy in (2.13) should be noted. The number of parameters, p, is invariant across all candidates for moments considered and hence the only part of $ent_\theta(f)$ that varies with the choice of moments is $-\ln\big[|G(f)'S(f)^{-1}G(f)|\big] = \ln\big[|V_\theta(f)|\big]$.

The following theorem states that this entropy of the limiting distribution can be used as a measure of the information content in population moment conditions and motivates the entropy based relevant moment selection criterion.

Theorem 2.1. *Let v_t satisfy Assumption 2.1 and define $\mathcal{F}$ as the set of moment functions that satisfied the asymptotic normality stated in Assumption 2.6. That is,*

$$\mathcal{F} = \{f(\cdot) \textit{ such that Assumption 2.6 holds}\}.$$

Then, we have the followings.

(i) Let f^0 be the optimal choice of moment condition from $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{F}} \subseteq \mathcal{F}$. Then $ent_\theta(f^0) \leq ent_\theta(f)$ for all $f \in \tilde{\mathcal{F}}$.

(ii) Define $f(v_t, \theta) = [f_1(v_t, \theta)', f_2(v_t, \theta)']'$. Assume $f_i \in \mathcal{F}$ for $i = 1, 2$. If the moment condition $E[f_2(v_t, \theta_0)] = 0$ is redundant for θ_0 given $E[f_1(v_t, \theta_0)] = 0$

then $ent_\theta(f) = ent_\theta(f_1)$*. If* $E[f_2(v_t, \theta_0)] = 0$ *is non-redundant for* θ_0 *given* $E[f_1(v_t, \theta_0)] = 0$ *then* $ent_\theta(f) < ent_\theta(f_1)$*.*

(iii) If $\text{rank}\{G(f)\} < p$ *then* $ent_\theta(f) = \infty$*.*

Proof. For the proof of Theorem 2.1, it is useful to introduce the following lemma based on results due to Rao (1973).

Lemma 2.1. *If* A *and* B *are* $p \times p$ *real symmetric positive definite matrices and* $A - B$ *is positive semi-definite, then* $|A| \geq |B|$ *and the inequality is strict if* $A \neq B$*.*

Proof. Let $\{\lambda_i; i = 1, 2, \ldots, p\}$ be the solutions to the determinental equation $|A - \lambda B| = 0$. Rao (1973, Problem 9 (i), p.70) states that $\lambda_i \geq 0$ for $i = 1, 2, \ldots, p$, and Rao (1973, Problem 9 (ii), p.70) states that $\frac{|A|}{|B|} = \prod_{i=1}^{p} \lambda_i$. These two results immediately imply that $|A| \geq |B|$. Now we need to show that strict inequality holds when $A \neq B$. According to Rao's (1973, p.41) results of the canonical reduction of a pair of real symmetric matrices, there exists a matrix R such that $A = R^{-1'} \Lambda R^{-1}$ and $B = R^{-1'} R^{-1}$, where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$. These canonical reductions imply that

$$R'(A - B)R = \Lambda - I_p \tag{2.14}$$

Since $A - B$ is positive semi-definite by assumption and R is nonsingular, we have that if $A \neq B$, $R'(A - B)R$ is non-null and positive semi-definite. This means that at least one of its eigenvalues is positive. Hence we have

$$\mathrm{trace}(R'(A-B)R) \geq 0 \tag{2.15}$$

(2.14) and (2.15) with Rao (1973, Problem 9 (i), p.70) imply that $\lambda_i > 1$ for at least one i. $\square$

Now we turn to the proof of Theorem 2.1.

part *(i)*. By construction, we have that: (i) $V_\theta(f)$ and $V_\theta(f^0)$ are positive definite; (ii) $V_\theta(f) - V_\theta(f^0)$ is positive semi-definite. The result follows from Lemma 2.1 and (2.13).

part *(ii)*. If $E[f_2(v_t, \theta_0)] = 0$ is redundant for θ_0 given $E[f_1(v_t, \theta_0)] = 0$, then by definition $V_\theta(f) = V_\theta(f_1)$. The results immediately follows from (2.13). If $E[f_2(v_t, \theta_0)] = 0$ is non-redundant for θ_0 given $E[f_1(v_t, \theta_0)] = 0$, then by definition $V_\theta(f_1) - V_\theta(f)$ is positive semi-definite and so $|V_\theta(f_1)| \geq |V_\theta(f)|$. The result follows from Lemma 2.1 and (2.13).

part *(iii)*. If $\mathrm{rank}\{G(f)\} < p$, then $G_0' S^{-1} G_0$ is singular and therefore $|G_0' S^{-1} G_0| = 0$. The result immediately follows from (2.13). $\square$

Theorem 2.1 implies that we want to select moment conditions that minimize $ent_\theta(f)$ and hence $\ln\big[|V_\theta(f)|\big]$ across all choices of f in the candidate set. Now the form of entropy in (2.13) and the results in Theorem 2.1 motivate us to consider the following entropy based moment selection criterion. RMSC stands for *relevant* *moment selection criterion.*

$$RMSC(c) = \ln\Big[\big|\hat{V}_{\theta,T}(c)\big|\Big] + \kappa\big(|c|, T\big) \tag{2.16}$$

where $\hat{V}_{\theta,T}(c)$ denotes a consistent estimator of $V_\theta(c)$ and $\kappa(|c|, T)$ is a deterministic penalty that is an increasing function of the number of moments, $|c|$. A natural choice for the covariance matrix estimator is

$$\hat{V}_{\theta,T}(c) = [G_T(\hat{\theta}_T(c); c)'\hat{S}_T^{-1}(c)G_T(\hat{\theta}_T(c); c)]^{-1} \tag{2.17}$$

where $\hat{\theta}_T(c)$ is the GMM estimator based on the population moment condition associated with the selection vector c, $E[f(v_t, \theta_0; c)] = 0$, and

$$\begin{aligned} G_T(\theta; c) &= T^{-1}\sum_{t=1}^{T}\frac{\partial f(v_t, \theta; c)}{\partial\theta'}; \\ \hat{S}_T(c) &\xrightarrow{p} S(c); \\ S(c) &= \lim_{T\to\infty} Var\Big[T^{-1/2}\sum_{t=1}^{T} f(v_t, \theta_0; c)\Big]. \end{aligned}$$

These considerations lead us to base estimation on the selection vector that minimizes $RMSC(c)$ over C. Now the selection vector $\hat{c}_T$ is defined as

$$\hat{c}_T \;=\; \mathrm{Argmin}_{c\in C}\, RMSC(c) \tag{2.18}$$

To analyze the asymptotic properties of the selection vector $\hat{c}_T$, we require certain regularity conditions. To present these regularity conditions, it is necessary to define the set of selection vectors that are asymptotically efficient relative to the candidate set,

$$\mathcal{C} \;=\; \{c;\, V_\theta(\iota_{q_{max}}) = V_\theta(c),\, c \in C\} \tag{2.19}$$

and also the subset of $\mathcal{C}$ of minimum length,

$$\mathcal{C}_{min} \;=\; \{c;\, c \in \mathcal{C},\, |c| \le |\bar{c}| \text{ for all } \bar{c} \in \mathcal{C}\} \tag{2.20}$$

Using this notation, we impose the following assumptions.

Assumption 2.7. *(i) c_r satisfies Definition 2.4 and $\mathcal{C}_{min} = \{c_r\}$; (ii) $E[f(v_t, \theta_0; c)] = 0$ if and only if $\theta = \theta_0$ for all $c \in C$; (iii) $\hat{V}_{\theta,T}(c) = V_\theta(c) + O_p(\tau_T^{-1})$ where $\tau_T \to \infty$ as $T \to \infty$; (iv) for any $\tilde{c}, \bar{c} \in C$ such that $|\bar{c}| > |\tilde{c}|$, $\tau_T[\kappa(|\bar{c}|, T) - \kappa(|\tilde{c}|, T)] \to +\infty$ as $T \to \infty$, and $\kappa(|c|, T) = o(1)$ for every $c \in C$.*

Assumptions 2.7(i) and (ii) are about the identification conditions for $RMSC(c)$ and θ_0, respectively. In practice, the choice of τ_T is as follows: when the weighting matrix is the inverse of the sum of a fixed number of autocovariances, $\tau_T = T^{1/2}$; when the weighting matrix is the inverse of a heteroscedasticity autocorrelation covariance (HAC) matrix calculated with bandwidth ℓ_T such that $\ell_T \to \infty$ as $T \to \infty$ and $\ell_T = o(T^{1/2})$ then $\tau_T = (T/\ell_T)^{1/2}$. Andrews (1991) provides more primitive conditions for Assumption 2.7(iii) for this case (e.g., his Assumptions B and C). Assumption 2.7(iv) states the very general properties of the deterministic penalty term which is designed by Andrews (1999). It is the most desirable to develop an optimal choice for the penalty term, but this is left for the future research. In their simulation studies, both Andrews (1999) and Hall and Inoue (2003) found that the BIC-type penalty works best in this context. The BIC-type penalty is given as

$$\kappa(|c|, T) = (|c| - p)\ln(\tau_T)/\tau_T \tag{2.21}$$

Within this framework, Hall and Inoue (2003) establish the consistency of $\hat{c}_T$ for c_r. For the completion of this chapter, their theorem and proof are presented here.

Theorem 2.2. *If Assumption 2.7 holds, then we have*

$$\hat{c}_T \xrightarrow{p} c_r \tag{2.22}$$

Proof. First, we define

$$\begin{aligned}
\Delta_T(c, c_r) &= RMSC(c) - RMSC(c_r) \\
&= \ln\big[|\hat{V}_{\theta,T}(c)|\big] + \kappa(|c|, T) - \Big\{\ln\big[|\hat{V}_{\theta,T}(c_r)|\big] + \kappa(|c_r|, T)\Big\} \\
&= \Big\{\ln\big[|\hat{V}_{\theta,T}(c)|\big] - \ln\big[|V_\theta(c)|\big]\Big\} - \Big\{\ln\big[|\hat{V}_{\theta,T}(c_r)|\big] - \ln\big[|V_\theta(c_r)|\big]\Big\} \\
&\quad + \Big\{\ln\big[|V_\theta(c)|\big] - \ln\big[|V_\theta(c_r)|\big]\Big\} + \{\kappa(|c|, T) - \kappa(|c_r|, T)\}
\end{aligned} \tag{2.23}$$

From Definition 2.4(ii), we have that $V_\theta(\iota_{q_{max}}) = V_\theta(c_r)$ and so it suffices to consider $\Delta_T(c, c_r)$ for two choices of c: (i) c such that $V_\theta(c) = V_\theta(c_r)$; (ii) c such that $V_\theta(c) - V_\theta(c_r) = M(c)$ where $M(c)$ is a non-null positive semi-definite matrix.

***Case (i)**:* In this case, we show that if the relevant set of moments is augmented with any irrelevant set of moments, than $RMSC(c)$ increases with probability one. Since we have c such that $V_\theta(c) = V_\theta(c_r)$, (2.23) is reduced to

$$\begin{aligned}
\Delta_T(c, c_r) = &\Big\{\ln\big[|\hat{V}_{\theta,T}(c)|\big] - \ln\big[|V_\theta(c)|\big]\Big\} - \Big\{\ln\big[|\hat{V}_{\theta,T}(c_r)|\big] - \ln\big[|V_\theta(c_r)|\big]\Big\} \\
&+ \{\kappa(|c|, T) - \kappa(|c_r|, T)\}
\end{aligned} \tag{2.24}$$

By Assumption 2.7(iii), we have: $\tau_T \Delta_T(c, c_r) = O_p(1) + \tau_T[\kappa(|c|, T) - \kappa(|c_r|, T)]$. Assumption 2.7(i) states that $|c| > |c_r|$ and so it follows from Assumption 2.7(iv) that: $\lim_{T\to\infty} \tau_T[\kappa(|c|, T) - \kappa(|c_r|, T)] = +\infty$. Thus, $\tau_T \Delta_T(c, c_r)$ is positive with probability one in the limit as $T \to \infty$.

Case (ii): In this case, we show that $RMSC(c)$ can always be reduced by including any omitted relevant set of moments. Since $V_\theta(c) - V_\theta(c_r) = M(c)$ where $M(c)$ is a non-null positive semi-definite matrix, from Lemma 2.1 it follows that:

$$\ln\big[|V_\theta(c)|\big] - \ln\big[|V_\theta(c_r)|\big] = m(c, c_r) \tag{2.25}$$

where $m(c, c_r) : C \times C \to [0, +\infty)$. From Assumptions 2.7(iii) and (iv), it follows that:

$$\ln\big[|\hat{V}_{\theta,T}(c)|\big] - \ln\big[|\hat{V}_{\theta,T}(c_r)|\big] = m(c, c_r) + o_p(1) \tag{2.26}$$

Therefore $\Delta_T(c, c_r)$ is positive with probability one in the limit as $T \to \infty$.

Taken together the results in Cases (i) and (ii) yield the desired result. □

It should be noted that the consistency result in Theorem 2.2 is premised on the assumption that θ_0 is identified by all the subsets of candidate set considered. This observation motivates us to consider the weak identification. Since the seminal papers Nelson and Startz (1990), Staiger and Stock (1997) and Stock and Wright (2000), the problem of weak identification has received considerable attention in the literature. To the best of author's knowledge, there is no consensus about how to define the concept of weak identification. One definition of weak identification can be found in Stock and Yogo (2002). They argue persuasively that weak identification means that standard asymptotics do not provide a good approximation and that the definition of weak identification depends, therefore, on what aspect of the asymptotic approximation one is interested in. Despite of the many ways of defining weak identification, it is widely accepted that in the presence of weak identification the limiting distribution in Assumption 2.6 may provide a poor approximation to the finite sample behavior. Therefore, it is worthwhile to consider the limiting behavior of $RMSC(c)$ when weak identification is a possibility. This is the topic of the next two chapters. Chapters 3 and 4 will consider the limiting behavior of $RMSC(c)$ in the linear and nonlinear settings, respectively.

Chapter 3

Near Redundancy and Weak Identification in the Linear Model

In the previous chapter, the concepts of redundancy and weak identification are briefly introduced. These concepts capture the situation where some set of moments can only provide a marginal information at best about the estimation and have very similar potential problem of the deterioration of the small sample properties. Even though these concepts share some common aspects, in their original forms they can not be easily compared each other, since the weak identification contains the Pitman drift but the redundancy does not. In this chapter, we will introduce the concept of *near* redundancy and compare it to weak identification. The limiting behavior of $RMSC(c)$ under the presence of weak identification will also be exploited. Through out this chapter, we confine our discussion to the context of linear IV model estimated via 2 Stage Least Squares (2SLS).

3.1 Redundancy and Weak Identification in the Linear Model

To introduce the concept of near redundancy, it is convenient to start by considering the following linear IV model.

$$y_t = x_t'\theta_0 + u_t, \tag{3.1}$$

$$x_t = \Pi_1 z_{1,t} + \Pi_2 z_{2,t} + e_t, \tag{3.2}$$

where y_t is a scalar, x_t is $p \times 1$ vector, $z_{i,t}$ is $q_i \times 1$ for $i = 1, 2$. In addition, we set $z_t = (z_{1,t}', z_{2,t}')'$ and set $q = q_1 + q_2$. Let Z_i be the $T \times q_i$ matrix whose t^{th} row is $z_{i,t}'$, $Z = (Z_1, Z_2)$, and X be the $T \times p$ matrix whose t^{th} row is x_t', and u be the $T \times 1$ vector with t^{th} element u_t. We define $v_t = (x_t', z_t', u_t, e_t')'$ and assume $\{v_t;\, t = 1, 2 \ldots T\}$ is an independently and identically distributed sequence of random vectors. Furthermore, it is assumed that $E[u_t|z_t] = 0$ and $E[u_t^2|z_t] = \sigma_0^2$.

Now we consider the condition for redundancy in this setting. According to Breusch, Qian, Schmidt, and Wyhowski (1999), the condition for redundancy of $E[z_{2,t}u_t] = 0$ for the estimation of θ_0 given $E[z_{1,t}u_t] = 0$ is

$$E[z_{2,t}x_t'] - E[z_{2,t}z_{1,t}'](E[z_{1,t}z_{1,t}'])^{-1}E[z_{1,t}x_t'] = 0 \tag{3.3}$$

Since redundancy is about the limiting behavior, the following sample analog to this condition in the limit is the important one.

$$T^{-1}Z_2'X - T^{-1}Z_2'Z_1(T^{-1}Z_1'Z_1)^{-1}T^{-1}Z_1'X \xrightarrow{p} 0 \tag{3.4}$$

This relation holds since the redundancy condition in (3.3) holds for every t.

We now impose some parametric restrictions on the data generating process in (3.1) and (3.2) in order to introduce the concept of near redundancy. It is assumed that the structural equation (3.1) still holds, but the first stage equation (3.2) is modified as

$$x_t \;=\; \Pi_1 z_{1,t} \;+\; \Pi_{2,T} z_{2,t} \;+\; e_t \tag{3.5}$$

It should be noticed that the coefficient on $z_{2,t}$ depends on T. The implication of this modification is that x_t and, in turn, v_t depend on the sample size T. For notational simplicity, this dependence will be omitted unless it is needed for emphasis and clarification. Besides these variables, it is assumed that all the other variables have distributions that are independent of T. Define $E[T^{-1}Z_i'Z_j] = \Omega_{i,j}$, for $i,j = 1,2$, $E[T^{-1}Z_i'X] = \Omega_{i,x,T}$, $\Omega_{i,x} = \lim_{T\to\infty}\Omega_{i,x,T}$, and finally let

$$\Omega_{z,z} \;=\; \begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix}, \qquad \Omega_{z,x} \;=\; \begin{bmatrix} \Omega_{1,x} \\ \Omega_{2,x} \end{bmatrix}$$

We impose the following high level assumptions.

Assumption 3.1. *(i)* $\operatorname{rank}\{\Omega_{i,i}\} = q_i$ *for* $i = 1,2$*; (ii)* $T^{-1}Z'Z \xrightarrow{p} \Omega_{z,z}$*; (iii)* $T^{-1/2}Z'u \xrightarrow{d} N(0, \sigma_0^2\Omega_{z,z})$*; (iv)* $T^{-1/2}\sum_{t=1}^{T} z_t \otimes e_t \xrightarrow{d} N(0,\Sigma_1)$*; (v)* $T^{-1}u'u \xrightarrow{p} \sigma_0^2$*; (vi)* $T^{-1/2}\sum_{t=1}^{T}(e_t u_t - \sigma_{e,u}) \xrightarrow{d} N(0,\Sigma_2)$ *where* $\sigma_{e,u} = E[e_t u_t]$.

Finally, the formal definition of near redundancy in this context is given as follows.

Definition 3.1 (Near Redundancy 1)**.** *Let the data be generated via (3.1) and (3.5) and Assumption 3.1 hold. The moment condition* $E[z_{2,t}u_t] = 0$ *is said to be nearly redundant for the estimation of* θ_0 *given* $E[z_{1,t}u_t] = 0$ *if*

$$\Omega_{2,x,T} \;-\; \Omega_{2,1}\Omega_{1,1}^{-1}\Omega_{1,x,T} \;=\; T^{-1/2}\eta \tag{3.6}$$

where η is a non-zero vector of finite constants.

(3.6) in the definition of near redundancy implies

$$\Omega_{2,x} - \Omega_{2,1}\Omega_{1,1}^{-1}\Omega_{1,x} = 0 \tag{3.7}$$

This, in turn, implies that (3.4) holds. These observations suggests that nearly redundant moment conditions make no contribution to the limiting variance of the GMM estimator. This intuition is confirmed in the following theorem.

Theorem 3.1. *Let $\hat{\theta}_T$ be the 2SLS estimator of θ_0 based on $E[z_t u_t] = 0$. Assume that $E[z_{2,t} u_t] = 0$ is nearly redundant for θ_0 given $E[z_{1,t} u_t] = 0$. Let Assumption 3.1 hold, the data be generated via (3.1) and (3.5) and* $\text{rank}(\Omega_{1,x}) = p$. *The limiting distribution of this GMM estimator is:*

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0, \sigma_0^2[\Omega_{1,x}'\Omega_{1,1}^{-1}\Omega_{1,x}]^{-1}\right) \tag{3.8}$$

Proof. Consider

$$T^{1/2}(\hat{\theta}_T - \theta_0) = \mathcal{Q}_T T^{-1/2} Z' u$$

where $\mathcal{Q}_T = [T^{-1}X'Z(T^{-1}Z'Z)^{-1}T^{-1}Z'X]^{-1}T^{-1}X'Z(T^{-1}Z'Z)^{-1}$. From Assumption 3.1, it follows that

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0, [\Omega_{z,x}'\Omega_{z,z}^{-1}\Omega_{z,x}]^{-1}\right) \tag{3.9}$$

Now consider

$$\mathtt{V} = \Omega'_{z,x}\Omega^{-1}_{z,z}\Omega_{z,x} = \begin{bmatrix} \Omega'_{1,x} & \Omega'_{2,x} \end{bmatrix} \begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix}^{-1} \begin{bmatrix} \Omega_{1,x} \\ \Omega_{2,x} \end{bmatrix} \tag{3.10}$$

Using the result of inverting partitioned matrices from Hayashi (2000, p. 673) it follows that

$$\begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix}^{-1} = \begin{bmatrix} \Omega^{-1}_{1,1} + \Omega^{-1}_{1,1}\Omega_{1,2}\mathtt{F}\Omega_{2,1}\Omega^{-1}_{1,1} & -\Omega^{-1}_{1,1}\Omega_{1,2}\mathtt{F} \\ -\mathtt{F}\Omega_{2,1}\Omega^{-1}_{1,1} & \mathtt{F} \end{bmatrix} \tag{3.11}$$

where $\mathtt{F} = (\Omega_{2,2} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,2})^{-1}$. Combining these two and collecting terms yield

$$\begin{aligned} \mathtt{V} &= \Omega'_{1,x}(\Omega^{-1}_{1,1} + \Omega^{-1}_{1,1}\Omega_{1,2}\mathtt{F}\Omega_{2,1}\Omega^{-1}_{1,1})\Omega_{1,x} \\ &\quad - \Omega'_{2,x}\mathtt{F}\Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,x} - \Omega'_{1,x}\mathtt{F}\Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{2,x} + \Omega'_{2,x}\mathtt{F}\Omega_{2,x} \\ &= \Omega'_{1,x}\Omega^{-1}_{1,1}\Omega_{1,x} + [\Omega'_{2,x} - \Omega'_{1,x}\Omega^{-1}_{1,1}\Omega_{1,2}]\mathtt{F}[\Omega_{2,x} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,x}] \\ &= \Omega'_{1,x}\Omega^{-1}_{1,1}\Omega_{1,x} + \mathtt{G}'\mathtt{F}\mathtt{G} \end{aligned} \tag{3.12}$$

where $\mathtt{G} = \Omega_{2,x} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,x}$. It follows that from (3.7) $\mathtt{G} = 0$ and so $\mathtt{V} = \Omega'_{1,x}\Omega^{-1}_{1,1}\Omega_{1,x}$ which gives the desired result. □

In order for the comparison with weak identification, we impose the parametric restriction on the data generating process in (3.1) and (3.5). By definition, we have

$$\Omega_{2,x,T} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,x,T} = E[T^{-1}Z'_2X] - E[T^{-1}Z'_2Z_1]\left\{E[T^{-1}Z'_1Z_1]\right\}^{-1}E[T^{-1}Z'_1X] \tag{3.13}$$

Using (3.5) and Assumption 3.1, (3.13) can be written as

$$\Omega_{2,x,T} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,x,T} = (\Omega_{2,2} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,2})\Pi'_{2,T} \tag{3.14}$$

According to Assumption 3.1(i) $\Omega_{2,2} - \Omega_{2,1}\Omega_{1,1}^{-1}\Omega_{1,2}$ is a nonsingular matrix of constants. Thus, we can conclude that $E[z_{2,t}u_t] = 0$ is nearly redundant for the estimation of θ_0 given $E[z_{1,t}u_t] = 0$ if and only if the parametric restriction $\Pi_{2,T} = \mathtt{C}_1 T^{-1/2}$ holds for some non-null matrix of constants $\mathtt{C}_1$.

We now consider weak identification. As mentioned above, there are several ways to define the concept of weak identification. Among those, Staiger and Stock's (1997) *classic* version of weak identification in the linear model is sufficient for the current discussion. See Nelson, Startz, and Zivot (2000) or Stock, Wright, and Yogo (2002) for different treatments of weak identification. Again, we assume the data is generated by (3.1) and (3.5). However, for the ease of comparison with near redundancy, we impose some additional restrictions on (3.5). We assume $\Pi_1 = 0$ and $\Pi_{2,T} = T^{-1/2}\mathtt{C}_1$ and so (3.5) is rewritten as

$$x_t \;=\; T^{-1/2}\mathtt{C}_1\, z_{2,t} \;+\; e_t \tag{3.15}$$

We also assume that the estimation is based on $E[z_{2,t}u_t] = 0$. In this case, the key derivative matrix is

$$\Omega_{2,x,T} \;=\; E[T^{-1}Z_2'X] \;=\; T^{-1/2}\Omega_{2,2}\mathtt{C}_1' \tag{3.16}$$

and so $\Omega_{2,x} = 0_{q_2\times p}$ causing identification to fail as $T \to \infty$.

Note that $\Pi_{2,T}$ behaves the same way under both near redundancy and weak identification. From the observations above, we can say that if θ_0 is weakly identified based on $E[z_{2,t}u_t] = 0$ alone then these moments are nearly redundant once the moment condition is augmented by $E[z_{1,t}u_t] = 0$. It is now natural to ask whether a set of moment conditions can be nearly redundant when other moments are included but be associated with weak identification if these other moments are excluded. To answer this question, we again slightly modify the first stage equation of the data generating process by imposing some other parametric restrictions. These restrictions are rank$\{\Pi_1\} = p$ and $\Pi_{2,T} = T^{-1/2}\mathtt{C}_1$. Now the first stage equation of the data generating process is written as

$$x_t = \Pi_1 z_{1,t} + T^{-1/2}\mathtt{C}_1 \, z_{2,t} + e_t \tag{3.17}$$

When the data is generated by (3.1) and (3.17), the next statement is straightforward by the definition of the near redundancy. If estimation is based on $E[z_t u_t] = 0$ then $E[z_{2,t}u_t] = 0$ is nearly redundant because of the inclusion of $E[z_{1,t}u_t] = 0$. However, if estimation is based on $E[z_{2,t}u_t] = 0$ alone then this set of moments is inevitably the only source of information about θ_0. Does this mean that θ_0 is weakly identified? The answer is may be or may be not. To see this, note that within this specification the key derivative matrix is

$$\Omega_{2,x,T} = \Omega_{2,1}\Pi_1' + T^{-1/2}\Omega_{2,2}\mathtt{C}_1' \tag{3.18}$$

Therefore, θ_0 is identified provided rank$\{\Omega_{2,1}\Pi_1'\} = p$, but weakly identified if this condition fails. Or, put another way, θ_0 is identified provided $z_{2,t}$ inherits enough of the explanatory power of $z_{1,t}$ for x_t when the latter is omitted.

3.2 Behavior of RMSC with Weak Identification in the Linear Model

As noted above, the consistency result of $RMSC(c)$ in Theorem 2.2 is premised on certain regularity conditions. One of these conditions is the requirement that θ_0 is identified by all the subsets of the candidate set over which the minimization is performed. Even though it is a feasible situation, the comparison between near redundancy and weak identification in the previous section opens the possibility that some redundant moment conditions given the relevant subset may fail to identify θ_0 when some or all of that relevant subset are omitted from the estimation. Therefore, in this section, we consider the limiting behavior of $RMSC(C)$ when weak identification is a possibility, and then use these results to derive the limiting behavior of $\hat{c}_T$ when the parameter vector may be weakly identified for some of the combinations considered. All of these analyses will be undertaken in the context of the linear model. In later chapter, these will be extended to allow for the nonlinear model.

Since there is no reason to suppose that weak identification affects all elements of θ_0 equally, this time, we partition to x_t into $(x'_{1,t}, x'_{2,t})'$ where $x_{i,t}$ is $p_i \times 1$ for $i = 1, 2$, and partition θ_0 conformably into $(\theta'_{0,1}, \theta'_{0,2})'$. Taking account of this partition scheme in the first stage equation of the data generating process, we rewrite (3.2) as

$$\begin{aligned} x_{1,t} &= \Pi_{1,1,T} z_{1,t} + \Pi_{1,2,T} z_{2,t} + e_{1,t} \\ x_{2,t} &= \Pi_{2,1,T} z_{1,t} + \Pi_{2,2,T} z_{2,t} + e_{2,t} \end{aligned} \tag{3.19}$$

where $\Pi_{i,j,T}$ is $p_i \times q_j$. All other definitions and assumptions of previous discussion will be remain the same. Once again, we let $\hat{\theta}_T$ denote the 2SLS estimator of θ_0 based on $E[z_t u_t] = 0$. In this setting, the limiting variance of 2SLS estimator is written as

$$V_\theta = \sigma_0^2 (\Omega_{x,z} \Omega_{z,z}^{-1} \Omega_{z,x})^{-1} \tag{3.20}$$

Given these structure, the obvious candidate for the covariance matrix estimator is

$$\hat{V}_{\theta,T} = \hat{\sigma}_T^2 [T^{-1} X'Z (T^{-1} Z'Z)^{-1} T^{-1} Z'X]^{-1} \tag{3.21}$$

where $\hat{\sigma}_T^2 = T^{-1}(y - X\hat{\theta}_T)'(y - X\hat{\theta}_T)$. Then we have

$$\ln\big[|\hat{V}_{\theta,T}|\big] = p \ln(\hat{\sigma}_T^2) - \ln\big[|T^{-1} X'Z (T^{-1} Z'Z)^{-1} T^{-1} Z'X|\big] \tag{3.22}$$

For the analysis of the behavior of $RMSC(c)$ when θ_0 is only weakly identified by some subsets of the candidate set of moment conditions, we will consider first the behavior of $\ln[|\hat{V}_{\theta,T}|]$ for the following three distinct scenarios.

Scenario 3.1. (θ_0 ***is weakly identified.***)

$\Pi_{i,j,T} = T^{-1/2}\mathtt{C}_{i,j}$ for some matrices of constants $\mathtt{C}_{i,j}$, $i,j = 1,2$ and $\text{rank}\{\mathtt{C}\} = p$ where

$$\mathtt{C} = \begin{bmatrix} \mathtt{C}_{1,1} & \mathtt{C}_{1,2} \\ \mathtt{C}_{2,1} & \mathtt{C}_{2,2} \end{bmatrix}$$

Scenario 3.2. ($\theta_{0,1}$ ***is identified but*** $\theta_{0,2}$ ***is weakly identified.***)

$\Pi_{1,1,T} = \Pi_{1,1}$ with $\text{rank}(\Pi_{1,1}) = p_1$; $\Pi_{1,2,T} = T^{-1/2}\mathtt{C}_{1,2}$, for some matrix of constants $\mathtt{C}_{1,2}$; $\Pi_{2,j,T} = T^{-1/2}\mathtt{C}_{2,j}$ for some matrices of constants $\mathtt{C}_{2,j}$, $j = 1,2$ and $\text{rank}\{[\mathtt{C}_{2,1}, \mathtt{C}_{2,2}]\} = p_2$.

Scenario 3.3. (θ_0 ***is identified.***)

$\Pi_{i,1,T} = \Pi_{i,1} + T^{-1/2}\mathtt{C}_{i,1}$ with $\text{rank}(\Pi_{i,1}) = p_i$ for $i = 1,2$; $\Pi_{i,2,T} = T^{-1/2}\mathtt{C}_{i,2}$ for some matrices of constants $\mathtt{C}_{i,j}$, $i,j = 1,2$.

Some comments on Scenario 3.3 will follow. First, note the specification for $\Pi_{i,1,T}$ implies that θ_0 is identified by $E[z_{1,t}u_t] = 0$ but allows identification to rest on different elements of the moment condition for $\theta_{0,1}$ and $\theta_{0,2}$. Second, some elements of $E[z_t u_t] = 0$ are nearly redundant given other elements. That is, $E[z_{2,t}u_t] = 0$ is nearly redundant given $E[z_{1,t}u_t] = 0$. Some elements of $E[z_{1,t}u_t] = 0$ may also be nearly redundant given the remaining elements of this vector depending on the elements of $\Pi_{i,1}$ and $\mathtt{C}_{i,1}$. The following theorem presents the large sample behavior of $ln[|\hat{V}_{\theta,T}|]$ under these three scenarios.

Theorem 3.2. *Let the data be generated via (3.1) and (3.19) and Assumption 3.1 holds.*

(i) Under Scenario 3.1: $\ln\big[|\hat{V}_{\theta,T}|\big] = p\ln(T) + O_p(1)$;

(ii) Under Scenario 3.2: $\ln\big[|\hat{V}_{\theta,T}|\big] = p_2\ln(T) + O_p(1)$;

(iii) Under Scenario 3.3: $\ln\big[|\hat{V}_{\theta,T}|\big] = p\ \ln[\sigma_0^2] - \ln\big[|\Omega'_{1,u}\Omega_{1,1}^{-1}\Omega_{1,u}|\big] + o_p(1) = O_p(1)$.

Proof. Through out this proof, we use matrices B_T, D_T and A_T that are defined as follows.

$$\begin{aligned} B_T &= T^{-1}X'Z \\ D_T &= (T^{-1}Z'Z)^{-1} \\ A_T &= B_T D_T B_T' \end{aligned} \tag{3.23}$$

Now consider the following partitions.

$$B_T = \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix}, \quad D_T = \begin{bmatrix} D_{1,1} & D_{1,2} \\ D_{2,1} & D_{2,2} \end{bmatrix}, \quad A_T = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \tag{3.24}$$

where $B_{i,j}$ is $p_i \times q_j$, $D_{i,j}$ is $q_i \times q_j$ for $i, j = 1, 2$ (and the T subscript on $B_{i,j}$ and $D_{i,j}$

is suppressed for notational simplicity) and $A_{i,j}$'s are given as

$$
\begin{aligned}
A_{1,1} &= B_{1,1}D_{1,1}B'_{1,1} + B_{1,2}D_{2,1}B'_{1,1} + B_{1,1}D_{1,2}B'_{1,2} + B_{1,2}D_{2,2}B'_{1,2} \\
A_{1,2} &= B_{1,1}D_{1,1}B'_{2,1} + B_{1,2}D_{2,1}B'_{2,1} + B_{1,1}D_{1,2}B'_{2,2} + B_{1,2}D_{2,2}B'_{2,2} \\
A_{2,1} &= B_{2,1}D_{1,1}B'_{1,1} + B_{2,2}D_{2,1}B'_{1,1} + B_{2,1}D_{1,2}B'_{1,2} + B_{2,2}D_{2,2}B'_{1,2} \\
A_{2,2} &= B_{2,1}D_{1,1}B'_{2,1} + B_{2,2}D_{2,1}B'_{2,1} + B_{2,1}D_{1,2}B'_{2,2} + B_{2,2}D_{2,2}B'_{2,2}
\end{aligned}
\tag{3.25}
$$

Notice that T subscription is suppressed for the notational brevity. From Dhrymes (2000, Proposition 2.30), the determinant of partitioned matrix A_T can be calculated as

$$
\left|A_T\right| = \left|A_{2,2}\right| \left|A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}\right| \tag{3.26}
$$

Part(i): For the proof of part (i), we used the result that $\hat{\sigma}_T^2 = O_p(1)$ due to Staiger and Stock (1997, Theorem 1(b)). Consider $\ln[|A_T|]$. The order of $|A_T|$ can therefore be deduced from (3.26) once the orders of $A_{i,j}$'s are known. Assumption 3.1 implies that $D_{i,j} = O_p(1)$, and Scenario 3.1 implies that $B_{i,j} = O_p(T^{-1/2})$. Therefore, $A_{i,j} = O_p(T^{-1})$. Define $\tilde{A}_T = TA_T$, and $\tilde{A}_{i,j} = TA_{i,j}$. Notice that $\tilde{A}_T = O_p(1)$ by construction. We now show that $\tilde{A}_T$ is positive definite with probability one in the limit as $T \to \infty$. To this end, we write $\tilde{A}_T = \tilde{B}_T D_T \tilde{B}'_T$ where $\tilde{B}_T = T^{1/2}B_T$, and consider the following quadratic form: $v'\tilde{A}_T v = \left\{v'(T^{1/2}B_T)\right\} D_T \left\{(T^{1/2}B'_T)v\right\} = v'\tilde{B}_T D_T \tilde{B}'_T v$; for some non-zero vector v. Since D_T is positive definite by construction, we need to consider $(T^{1/2}B'_T)v = \tilde{B}'_T v$.

In order to do this, we express $x_{1,t}$ and $x_{2,t}$ in the matrix form as follows,

$$
X_i \;=\; Z_1\Pi'_{i,1,T} + Z_2\Pi'_{i,2,T} + E_i, \qquad \text{for } i = 1,2 \tag{3.27}
$$

where X_i and E_i are $T \times p_i$, Z_i is $T \times q_i$, for $i = 1, 2$. Now let $X = [X_1\ X_2]$ and $Z = [Z_1\ Z_2]$, then B_T can be written as:

$$B_T = T^{-1}X'Z = \begin{bmatrix} T^{-1}X_1'Z_1 & T^{-1}X_1'Z_2 \\ T^{-1}X_2'Z_1 & T^{-1}X_2'Z_2 \end{bmatrix} = \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix} \tag{3.28}$$

Each block of B_T can be written as

$$B_{i,j} \;=\; \Pi_{i,1,T}\left(T^{-1}Z_1'Z_j\right) + \Pi_{i,2,T}\left(T^{-1}Z_2'Z_j\right) + T^{-1}E_i'Z_j \tag{3.29}$$

for $i, j = 1, 2$. Then the each block of $\tilde{B}_T$, $\tilde{B}_{i,j}$ (supressing the T again), is given by: $\tilde{B}_{i,j} = T^{1/2}B_{i,j}$ for $i, j = 1, 2$.

Using the assumptions of Scenario 3.1, $\Pi_{i,j,T} = T^{1/2}\mathtt{C}_{i,j}$, we can conclude that

$$\tilde{B}_T'v \;\xrightarrow{d}\; \begin{bmatrix} \mathtt{C}_{1,1}\Omega_{1,1} + \mathtt{C}_{1,2}\Omega_{2,1} & \mathtt{C}_{1,1}\Omega_{1,2} + \mathtt{C}_{1,2}\Omega_{2,2} \\ \mathtt{C}_{2,1}\Omega_{1,1} + \mathtt{C}_{2,2}\Omega_{2,1} & \mathtt{C}_{2,1}\Omega_{1,2} + \mathtt{C}_{2,2}\Omega_{2,2} \end{bmatrix}' v + \tilde{B}_{normal}'v \tag{3.30}$$

where $\tilde{B}_{normal}$ is a matrix whose elements are normally distributed. Now let

$$\mathtt{C} = \begin{bmatrix} \mathtt{C}_{1,1} & \mathtt{C}_{1,2} \\ \mathtt{C}_{2,1} & \mathtt{C}_{2,2} \end{bmatrix}$$

then we can write, $\tilde{B}_T'v \xrightarrow{d} (\mathtt{C}\Omega_{z,z})'v + \tilde{B}_{normal}'v$. By construction, $\Omega_{z,z}$ is positive definite, hence if the matrix $\mathtt{C}$ is of full rank (i.e. rank$(\mathtt{C}) = p$), $\tilde{A}_T$ is positive definite with probability one.

Returning to A_T, if we substitute in (3.26) for $A_{i,j}$ in terms of $\tilde{A}_{i,j}$ then we obtain: $|A_T| = T^{-p}|\tilde{A}_T| = O_p(T^{-p})$; where the last equality follows from the properties of $\tilde{A}_T$ derived above. The desired result then follows from (3.22).

Part (ii): We first consider $\ln[|A_T|]$. The analysis evolves along similar lines to the proof of part (i) and uses the partitions defined therein. Once again, Assumption 3.1 implies that $D_{i,j} = O_p(1)$. Scenario 3.2 implies that $B_{1,j} = O_p(1)$ and $B_{2,j} = O_p(T^{-1/2})$. Therefore, it follows from (3.25) that $A_{1,1} = O_p(1)$, $A_{1,2} = O_p(T^{-1/2})$, $A_{2,1} = O_p(T^{-1/2})$ and $A_{2,2} = O_p(T^{-1})$. Define

$$\bar{A}_T = \begin{bmatrix} \bar{A}_{1,1} & \bar{A}_{1,2} \\ \bar{A}_{2,1} & \bar{A}_{2,2} \end{bmatrix} \tag{3.31}$$

where $\bar{A}_{1,1} = A_{1,1}$, $\bar{A}_{1,2} = T^{1/2}A_{1,2}$, $\bar{A}_{2,1} = T^{1/2}A_{2,1}$ and $\bar{A}_{2,2} = TA_{2,2}$.

Now we consider the properties of $\bar{A}_T$. Note $\bar{A}_T = O_p(1)$ by construction. We now show that $\bar{A}_T$ is positive definite with probability one in the limit as $T \to \infty$. To do this, we define,

$$\bar{B}_T = \begin{bmatrix} B_{1,1} & B_{1,2} \\ T^{1/2}B_{2,1} & T^{1/2}B_{2,2} \end{bmatrix} \tag{3.32}$$

Then, $\bar{A}_T$ can be written as

$$\bar{A}_T = \bar{B}_T D_T \bar{B}'_T \tag{3.33}$$

By the same logic as in part (i), we need to consider $\bar{B}'_T v$. Each block of $\bar{B}_T$ is:

$$\begin{aligned} \bar{B}_{1,j} &= \Pi_{1,1,T}\left(T^{-1}Z'_1Z_j\right) + \Pi_{1,2,T}\left(T^{-1}Z'_2Z_j\right) + T^{-1}E'_1Z_j \\ \bar{B}_{2,j} &= T^{1/2}\Pi_{2,1,T}\left(T^{-1}Z'_1Z_j\right) + T^{1/2}\Pi_{2,2,T}\left(T^{-1}Z'_2Z_j\right) + T^{-1/2}E'_2Z_j \end{aligned} \tag{3.34}$$

for $j = 1, 2$. Using the assumptions in Scenario 3.2, these can be rewritten as

$$\begin{aligned} \bar{B}_{1,j} &= \Pi_{1,1,T}\left(T^{-1}Z'_1Z_j\right) + T^{-1/2}\mathtt{C}_{1,2}\left(T^{-1}Z'_2Z_j\right) + T^{-1}E'_1Z_j \\ \bar{B}_{2,j} &= \mathtt{C}_{2,1}\left(T^{-1}Z'_1Z_j\right) + \mathtt{C}_{2,2}\left(T^{-1}Z'_2Z_j\right) + T^{-1/2}E'_2Z_j \end{aligned} \tag{3.35}$$

for $j = 1, 2$. From the expressions above, it can be easily concluded that

$$\bar{B}_T' v \xrightarrow{d} (\bar{\mathtt{C}}\Omega_{z,z})'v + \bar{B}_{normal}' v \tag{3.36}$$

where

$$\bar{\mathtt{C}} = \begin{bmatrix} \Pi_{1,1} & 0 \\ \mathtt{C}_{2,1} & \mathtt{C}_{2,2} \end{bmatrix}$$

By the same logic as part (i), we can conclude that if the matrix, $\bar{\mathtt{C}}$, is of full rank (i.e. $\text{rank}([\mathtt{C}_{2,1}\ \mathtt{C}_{2,2}]) = p_2$), $\bar{A}_T$ is positive definite with probability one and $O_p(1)$.

Substituting for $A_{i,j}$ in (3.26), we obtain: $|A_T| = T^{-p_2}|\bar{A}_T| = O_p(T^{-p_2})$; where the last equality follows from the properties of $\bar{A}_T$ derived above.

We now show that $\hat{\sigma}_T^2 = O_p(1)$. By definition, we have

$$T\hat{\sigma}_T^2 = u'u - 2u'X(\hat{\theta}_T - \theta_0) + (\hat{\theta}_T - \theta_0)'X'X(\hat{\theta}_T - \theta_0) \tag{3.37}$$

From Assumption 3.1(v), it follows that $u'u = O_p(T)$, and from Assumption 3.1(iii) and (v), it follows that $u'X = O_p(T)$. Therefore, we focus on $\hat{\theta}_T - \theta_0$. Let $\hat{\theta}_{T,i}$ be the 2SLS estimator of $\hat{\theta}_{0,i}$. Using the notation from the proof of part (i), $\hat{\theta}_T - \theta_0 = (B_T D_T B_T^{'})^{-1} B_T D_T Z'u$, and so it follows that

$$\begin{bmatrix} \hat{\theta}_{T,1} - \theta_{0,1} \\ \hat{\theta}_{T,2} - \theta_{0,2} \end{bmatrix} = \begin{bmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix} \begin{bmatrix} D_{1,1} & D_{1,2} \\ D_{2,1} & D_{2,2} \end{bmatrix} \begin{bmatrix} Z_1'u \\ Z_2'u \end{bmatrix} \tag{3.38}$$

where, from Dhrymes (2000, Proposition 2.31),

$$\begin{aligned} H_{1,1} &= (A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1})^{-1} \\ H_{1,2} &= -A_{1,1}^{-1}A_{1,2}(A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2})^{-1} \\ H_{2,1} &= -A_{2,2}^{-1}A_{2,1}(A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1})^{-1} \\ H_{2,2} &= (A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2})^{-1} \end{aligned} \tag{3.39}$$

Using the order statements given above, it can be shown that $H_{1,1} = O_p(1)$, $H_{1,2} = O_p(T^{1/2})$, $H_{2,1} = O_p(T^{1/2})$ and $H_{2,2} = O_p(T)$. Multiplying out (3.38), we obtain

$$\begin{aligned} \hat{\theta}_{T,1} - \theta_{0,1} &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} H_{1,i}B_{i,j}D_{j,k}Z_k'u \\ \hat{\theta}_{T,2} - \theta_{0,2} &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} H_{2,i}B_{i,j}D_{j,k}Z_k'u \end{aligned} \tag{3.40}$$

Using the order statements given above, it follows from (3.40) that $\hat{\theta}_{T,1} - \theta_{0,1} = O_p(T^{-1/2})$ and $\hat{\theta}_{T,2} - \theta_{0,2} = O_p(1)$. Using these order statements along with the others above, it follows from (3.37) that $\hat{\sigma}_T^2 = O_p(1)$. The desired result then follows from (3.22).

Part (iii): It follows from Theorem 3.1 that $\hat{\theta}_T - \theta_0 = O_p(T^{-1/2})$. Furthermore, from Assumption 3.1, we have that $X'u = O_p(T)$ $X'X = O_p(T)$ and $T^{-1}u'u = \sigma_0^2 + o_p(1)$. Therefore, it follows from (3.37) that $\hat{\sigma}_T^2 \xrightarrow{p} \sigma_0^2$. Now we consider $\ln[|T^{-1}X'Z(T^{-1}Z'Z)^{-1}T^{-1}Z'X|]$. Since both natural logarithm and determinant of matrices are continuous functions, it follows from Assumption 3.1, Slutsky's Theorem and (3.12) that

$$\ln[|T^{-1}X'Z(T^{-1}Z'Z)^{-1}T^{-1}Z'X|] \xrightarrow{p} \ln[|\Omega_{1,x}'\Omega_{1,1}^{-1}\Omega_{1,x}|] \tag{3.41}$$

which completes the proof. $\square$

Using the results in Theorem 3.2, we now consider the behavior of the moment selection procedure based on $RMSC(c)$. In this particular linear model, the chosen selection vector can be expressed as

$$\hat{c}_T = \text{Argmin}_{c\in C}\Big\{\ln[\hat{\sigma}_T^2(c)] - \ln\big[\big|T^{-1}X'Z(c)\{T^{-1}Z(c)'Z(c)\}^{-1}T^{-1}Z(c)'X\big|\big] + \kappa(|c|,T)\Big\} \quad (3.42)$$

where $Z(c)$ denotes the matrix of instruments chosen by the selection vector c. $Z(c)$ is the $T \times |c|$ matrix whose t^{th} row is $z_t(c)'$, $\hat{\sigma}_T^2(c) = T^{-1}[y - X\hat{\theta}_T(c)]'[y - X\hat{\theta}_T(c)]$ and $\hat{\theta}_T(c)$ is the 2SLS estimator of θ_0 based on $E[z_t(c)u_t] = 0$.

We introduce two additional assumptions in order to establish the consistency results of $RMSC(c)$ when weak identification is a possibility.

Assumption 3.2. *$C = C_I \cup C_{II} \cup C_{III}$ where C_I yields models that fit within Scenario 3.1, C_{II} yields models that fit within Scenario 3.2 and C_{III} yields models that fit within Scenario 3.3.*

Assumption 3.3. *There is a $c_r \in C_{III}$ that satisfies the properties in Definition 2.4 with $q_{max} = q$ and $\mathcal{C}_{min} = \{c_r\}$.*

These two additional assumptions specify a partition of C and an identification condition for the moment selection procedure and lead us to establish the following theorem.

Theorem 3.3. *Let the data be generated by (3.1) and (3.19) and Assumptions 2.7–3.3 hold, then* $\hat{c}_T \xrightarrow{p} c_r$.

Proof. From Theorem 3.2(i)-(ii), it follows that $RMSC(c) \to \infty$ as $T \to \infty$ with probability one for all $c \in C_I \cup C_{II}$. From Theorem 3.2(iii), $RMSC(c) = O_p(1)$ for $c \in C_{III}$. Therefore, $\lim_{T\to\infty} \text{Prob}(\hat{c}_T \in C_{III}) = 1$. The rest of the proof follows by the same argument as the proof of Theorem 2.2. $\square$

In Theorem 2.2 of the previous chapter, we establish the consistency of $RMSC(c)$ without considering redundancy or weak identification. Under certain regularity conditions, $RMSC(c)$ picks the selection vector c_r associated with the relevant moments with probability one in the limit as $T \to \infty$. In this chapter, focusing on the linear IV model estimated via 2SLS, we generalized this results to allow for the possibility of the presence of weak identification. As it can be seen in Theorem 3.3, $RMSC(c)$ is consistent for c_r even when subsets of the candidate set provide only weak identification. In the next chapter, we will generalize this further and consider the nonlinear model.

Chapter 4

Near Redundancy and Weak Identification in the Nonlinear Model

The discussion in the previous chapter frames the concepts of (near) redundancy and weak identification in the context of linear IV model estimated via 2SLS. In this chapter, we extend the previous discussions to allow for the nonlinear dynamic models as well as for the multiple equation GMM estimations. As it would be anticipate, we obtain the same qualitative results in this more general setting.

4.1 GIV and Parametric Restrictions on the Data Generating Process

For our purpose here, it is useful and sufficient to frame the discussion within a class of GMM estimators, widely known as Generalized Instrumental Variable (GIV). The reason of exploiting GIV is twofold; (i) GIV provides a convenient tool for the estimation and inference within the context of the nonlinear dynamic models, such as the Consumption Based Asset Pricing Models, (ii) GIV reduces a problem of moment selection to that of instrument selection. In the framework of GIV, the population moment condition denotes the statistical orthogonality of two random vectors, $u_t(\theta_0)$ and z_t.[1] The $M \times 1$ econometric disturbance vector $u_t(\theta_0) = \big(u_{1,t}(\theta_0), u_{2,t}(\theta_0), \ldots, u_{M,t}(\theta_0)\big)'$ consists of functions of the data and the unknown true parameter vector, and it is assumed that

$$E[u_{m,t}(\theta_0) \,|\, \mathbf{\Omega}_t] = 0, \quad \text{for } m = 1, 2, \ldots, M \tag{4.1}$$

where $\mathbf{\Omega}_t$ is the information set as of time period t and θ_0 is a $p \times 1$ true parameter vector that is unknown to the econometrician. In practice, (4.1) is usually implied by the first order conditions of the agent's optimization problem of underlying economic theory. Put differently, (4.1) represents the information derived from the underlying economic/econometric model and it states that the conditional expectations of each of the disturbance function evaluated at the true parameter value is zero. The $q \times 1$ vector of instruments z_t consists of a vector of functions of the elements in $\mathbf{\Omega}_t$, hence,

[1] In more general setting, $z_{t-\delta}$ for some non-negative integer δ should be used. For example, see Hall (2005) and Hansen and Singleton (1982). However, for the expositional brevity, it is assumed that $\delta = 0$ in this chapter.

it is satisfied that

$$z_t \in \mathbf{\Omega}_t \tag{4.2}$$

(4.1) means that the error is uncorrelated with any variables in the information set $\mathbf{\Omega}_t$. Since the instrument vector z_t is an element of $\mathbf{\Omega}_t$, applying the law of iterated expectation we can deduce the following population moment conditions from (4.1) and (4.2)

$$E[u_{m,t}(\theta_0)\, z_t] = 0, \quad \text{for } m = 1, 2, \ldots, M \tag{4.3}$$

It is worthy of noting that we have qM population moment conditions, since z_t is a $q \times 1$ vector and (4.3) holds for each error term $u_{m,t}(\theta_0)$ for $m = 1, 2, \ldots, M$. We assume that $qM > p$ for the identification of the parameters. Using the Kronecker product, we can rewrite the population moment conditions as

$$E[z_t \otimes u_t(\theta_0)] = 0_{(qM \times 1)} \tag{4.4}$$

Setting $f(v_t, \theta_0) = z_t \otimes u_t(\theta_0)$, (4.4) can be rewritten as

$$E[f(v_t, \theta_0)] = 0_{(qM \times 1)} \tag{4.4$'$}$$

Now we partition z_t into $\left(z'_{1,t}, z'_{2,t}\right)'$, where $z_{i,t}$ is $q_i \times 1$ for $i = 1, 2$ and $q_1 + q_2 = q$. Then the moment function $f(v_t, \theta_0)$ can be written as

$$f(v_t, \theta_0) = \begin{bmatrix} f_1(v_t, \theta_0) \\ f_2(v_t, \theta_0) \end{bmatrix} = \begin{bmatrix} z_{1,t} \otimes u_t(\theta_0) \\ z_{2,t} \otimes u_t(\theta_0) \end{bmatrix} \tag{4.5}$$

where $f_i(v_t, \theta_0)$ is $q_i M \times 1$ for $i = 1, 2$.

We define the following matrices which are helpful for our discussion of the connection between near redundancy and weak identification.

$$\begin{aligned}
\Omega_{f,f} &= E[f(v_t,\theta_0)f(v_t,\theta_0)'] \\
&= E\left[\begin{pmatrix} f_1(v_t,\theta_0) \\ f_2(v_t,\theta_0) \end{pmatrix} \begin{pmatrix} f_1(v_t,\theta_0)' & f_2(v_t,\theta_0)' \end{pmatrix}\right] \\
&= E\left[\begin{pmatrix} z_{1,t} \otimes u_t(\theta_0) \\ z_{2,t} \otimes u_t(\theta_0) \end{pmatrix} \begin{pmatrix} z'_{1,t} \otimes u_t(\theta_0)' & z'_{2,t} \otimes u_t(\theta_0)' \end{pmatrix}\right] \\
&= E\begin{bmatrix} z_{1,t}z'_{1,t} \otimes u_t(\theta_0)u_t(\theta_0)' & z_{1,t}z'_{2,t} \otimes u_t(\theta_0)u_t(\theta_0)' \\ z_{2,t}z'_{1,t} \otimes u_t(\theta_0)u_t(\theta_0)' & z_{2,t}z'_{2,t} \otimes u_t(\theta_0)u_t(\theta_0)' \end{bmatrix} \\
&= \begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix}
\end{aligned} \tag{4.6}$$

Notice that $\Omega_{f,f}$ is a $qM \times qM$ matrix and its $(i,j)^{th}$ block $\Omega_{i,j} = E[f_i(v_t,\theta_0)f_j(v_t,\theta_0)'] = E[z_{i,t}z'_{j,t} \otimes u_t(\theta_0)u_t(\theta_0)']$ is $q_iM \times q_jM$, for $i,j = 1,2$. We also define

$$\begin{aligned}
\Omega_{f,u,T} &= E\left[\frac{\partial f(v_t,\theta_0)}{\partial\theta'}\right] = E\left[z_t \otimes \frac{\partial u_t(\theta_0)}{\partial\theta'}\right] \\
&= E\begin{bmatrix} \frac{\partial f_1(v_t,\theta_0)}{\partial\theta'} \\ \frac{\partial f_2(v_t,\theta_0)}{\partial\theta'} \end{bmatrix} = E\begin{bmatrix} z_{1,t} \otimes \frac{\partial u_t(\theta_0)}{\partial\theta'} \\ z_{2,t} \otimes \frac{\partial u_t(\theta_0)}{\partial\theta'} \end{bmatrix} \\
&= \begin{bmatrix} \Omega_{1,u,T} \\ \Omega_{2,u,T} \end{bmatrix}
\end{aligned} \tag{4.7}$$

Note that $\Omega_{f,u,T}$ is $qM \times p$ and its i^{th} block $\Omega_{i,u,T} = E\left[\frac{\partial f_i(v_t,\theta_0)}{\partial\theta'}\right]$ is $q_iM \times p$, for $i = 1,2$. Further, we let

$$\lim_{T\to\infty} \Omega_{f,u,T} = \Omega_{f,u} = \begin{bmatrix} \Omega_{1,u} \\ \Omega_{2,u} \end{bmatrix} \tag{4.8}$$

As in the linear model discussed in the previous chapter, we impose some parametric restrictions on $\frac{\partial u_t(\theta)}{\partial \theta'}$. In the linear model $\frac{\partial u_t(\theta)}{\partial \theta'}$ has a simple form and it was easily implemented to the comparison of near redundancy and weak identification. However, in nonlinear model this has more complicated form and so it is investigated here. We consider linear projection[2] of $\frac{\partial u_t(\theta)}{\partial \theta'}$ onto z_t. Parametric restrictions on this linear projection coefficients will be considered in the next sections to get the desirable results. The m^{th} row of $M \times p$ matrix $\frac{\partial u_t(\theta)}{\partial \theta'}$ can be expressed as

$$\frac{\partial u_{m,t}(\theta)}{\partial \theta'} = z'_{1,t}\Phi_{m,1,T}(\theta) + z'_{2,t}\Phi_{m,2,T}(\theta) + \nu_{m,t}(\theta) \tag{4.9}$$

where $\Phi_{m,j,T}$ is a $q_j \times p$ matrix of unknown parameters for $m = 1, 2, \ldots, M$ and $j = 1, 2$; $\nu_{m,t}(\theta)$ is the forecast error of the linear projection and so it is statistically orthogonal to $z_{1,t}$ and $z_{2,t}$. Then $\frac{\partial u_t(\theta)}{\partial \theta'}$ can be written as

$$\begin{aligned}\begin{bmatrix} \frac{\partial u_{1,t}(\theta)}{\partial \theta'} \\ \frac{\partial u_{2,t}(\theta)}{\partial \theta'} \\ \vdots \\ \frac{\partial u_{M,t}(\theta)}{\partial \theta'} \end{bmatrix} &= \begin{bmatrix} z'_{1,t} & 0 & \cdots & 0 \\ 0 & z'_{1,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z'_{1,t} \end{bmatrix} \begin{bmatrix} \Phi_{1,1,T}(\theta) \\ \Phi_{2,1,T}(\theta) \\ \vdots \\ \Phi_{M,1,T}(\theta) \end{bmatrix} \\ &+ \begin{bmatrix} z'_{2,t} & 0 & \cdots & 0 \\ 0 & z'_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z'_{2,t} \end{bmatrix} \begin{bmatrix} \Phi_{1,2,T}(\theta) \\ \Phi_{2,2,T}(\theta) \\ \vdots \\ \Phi_{M,2,T}(\theta) \end{bmatrix} + \begin{bmatrix} \nu_{1,t}(\theta) \\ \nu_{2,t}(\theta) \\ \vdots \\ \nu_{M,t}(\theta) \end{bmatrix}\end{aligned} \tag{4.10}$$

and in more compact form

$$\frac{\partial u_t(\theta)}{\partial \theta'} = \mathcal{Z}'_{1,t}\Phi_{1,T}(\theta) + \mathcal{Z}'_{2,t}\Phi_{2,T}(\theta) + \nu_t(\theta) \tag{4.11}$$

where $\mathcal{Z}_{i,t} = I_M \otimes z_{i,t}$ is $q_i M \times M$ and $\Phi_{i,T}$ is $q_i M \times p$ for $i = 1, 2$.

[2]A formal treatment of the linear projection can be found in Hamilton (1994, Chapter 4).

As Stock and Wright (2000) pointed out, it will be at times useful to allow θ_0 to consist of some elements that are weakly identified and some that are identified. To allow for this case, we partition θ into $\begin{bmatrix}\theta_1' & \theta_2'\end{bmatrix}'$ and $\frac{\partial u_t(\theta)}{\partial \theta'}$ into $\begin{bmatrix}\frac{\partial u_t(\theta)}{\partial \theta_1'} & \frac{\partial u_t(\theta)}{\partial \theta_2'}\end{bmatrix}$. Note that θ_i is $p_i \times 1$ and $p = p_1 + p_2$; $\frac{\partial u_t(\theta)}{\partial \theta_i'}$ is $M \times p_i$, for $i = 1, 2$. For this purpose, consider the following partition, for $m = 1, 2, \ldots, M$.

$$\begin{aligned}
\Phi_{m,j,T}(\theta) &= \begin{bmatrix}\Phi^{(1)}_{m,j,T}(\theta) & \Phi^{(2)}_{m,j,T}(\theta)\end{bmatrix}, \quad \text{for } j = 1, 2 \\
\nu_{m,t}(\theta) &= \begin{bmatrix}\nu^{(1)}_{m,t}(\theta) & \nu^{(2)}_{m,t}(\theta)\end{bmatrix}
\end{aligned}$$

where $\Phi^{(h)}_{m,j,T}(\theta)$ is $q_j \times p_h$; $\nu^{(h)}_{m,t}(\theta)$ is $1 \times p_h$, for $j = 1, 2$ and $h = 1, 2$. Then we have

$$\frac{\partial u_{m,t}(\theta)}{\partial \theta'} = \left[\frac{\partial u_{m,t}(\theta)}{\partial \theta_1'}, \quad \frac{\partial u_{m,t}(\theta)}{\partial \theta_2'}\right] \tag{4.12}$$

where

$$\begin{aligned}
\frac{\partial u_{m,t}(\theta)}{\partial \theta_1'} &= z_{1,t}'\Phi^{(1)}_{m,1,T}(\theta) + z_{2,t}'\Phi^{(1)}_{m,2,T}(\theta) + \nu^{(1)}_{m,t}(\theta) \\
\frac{\partial u_{m,t}(\theta)}{\partial \theta_2'} &= z_{1,t}'\Phi^{(2)}_{m,1,T}(\theta) + z_{2,t}'\Phi^{(2)}_{m,2,T}(\theta) + \nu^{(2)}_{m,t}(\theta)
\end{aligned} \tag{4.13}$$

where $\frac{\partial u_{m,t}(\theta)}{\partial \theta_h'}$ is $1 \times p_h$. Now $\frac{\partial u_t(\theta)}{\partial \theta_h'}$, for $h = 1, 2$, can be written as

$$\begin{aligned}
\begin{bmatrix}\frac{\partial u_{1,t}(\theta)}{\partial \theta_h'} \\ \frac{\partial u_{2,t}(\theta)}{\partial \theta_h'} \\ \vdots \\ \frac{\partial u_{M,t}(\theta)}{\partial \theta_h'}\end{bmatrix}
&= \begin{bmatrix} z_{1,t}' & 0 & \cdots & 0 \\ 0 & z_{1,t}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_{1,t}'\end{bmatrix}
\begin{bmatrix}\Phi^{(h)}_{1,1,T}(\theta) \\ \Phi^{(h)}_{2,1,T}(\theta) \\ \vdots \\ \Phi^{(h)}_{M,1,T}(\theta)\end{bmatrix} \\
&+ \begin{bmatrix} z_{2,t}' & 0 & \cdots & 0 \\ 0 & z_{2,t}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_{2,t}'\end{bmatrix}
\begin{bmatrix}\Phi^{(h)}_{1,2,T}(\theta) \\ \Phi^{(h)}_{2,2,T}(\theta) \\ \vdots \\ \Phi^{(h)}_{M,2,T}(\theta)\end{bmatrix}
+ \begin{bmatrix}\nu^{(h)}_{1,T}(\theta) \\ \nu^{(h)}_{2,T}(\theta) \\ \vdots \\ \nu^{(h)}_{M,T}(\theta)\end{bmatrix}
\end{aligned} \tag{4.14}$$

and in more compact form

$$\begin{aligned}\frac{\partial u_t(\theta)}{\partial \theta_1'} &= \mathcal{Z}_{1,t}'\Phi_{1,T}^{(1)}(\theta) + \mathcal{Z}_{2,t}'\Phi_{2,T}^{(1)}(\theta) + \nu_t^{(1)}(\theta) \\ \frac{\partial u_t(\theta)}{\partial \theta_2'} &= \mathcal{Z}_{1,t}'\Phi_{1,T}^{(2)}(\theta) + \mathcal{Z}_{2,t}'\Phi_{2,T}^{(2)}(\theta) + \nu_t^{(2)}(\theta)\end{aligned} \tag{4.15}$$

In the next sections, it is found that these different schemes of partitioning $\frac{\partial u_t(\theta)}{\partial \theta'}$ are useful for our purpose of exploring the connection between near redundancy and weak identification and analyzing the behavior of $RMSC(c)$.

4.2 Redundancy and Weak Identification in the Nonlinear Model

In the previous chapter, we introduced the concept of near redundancy within the context of linear model. Since Breusch, Qian, Schmidt, and Wyhowski's (1999) concept of redundancy is not confined just to the linear cases, we can extend the concept of near redundancy so that it holds in more general setting and compare it to weak identification. For this purpose, we introduce Stock and Wright's (2000) assumption for weak identification. Following Stock and Wright (2000, Assumption C), we decompose the parameter vector into two parts of which one is identified and the other is weakly identified. Accordingly, we define $\theta' = (\theta^{(wi)'}, \theta^{(id)'})$ where $\theta^{(wi)}$ is $p_{wi} \times 1$, $\theta^{(id)}$ is $p_{id} \times 1$, $p = p_{wi} + p_{id}$ and the superscript indicates if the subvector is weakly identified (wi) or identified (id). Using these notations, we present Stock and Wright's (2000) framework in the next assumption about the nearly uninformative moment conditions.

Assumption 4.1 (Nearly Uninformative Moment Condition)**.**

(i) $E[f(v_t,\theta)] = T^{-1/2}\mu_{1T}(\theta) + \mu_2(\theta^{(id)})$.

(ii) $\mu_{1T}(\theta) \to \mu_1(\theta)$ *uniformly in* $\theta \in \Theta$*,* $\mu_1(\theta_0) = 0$*,* $\mu_1(\theta)$ *is continuous in* θ *and is bounded on* Θ*.*

(iii) $\mu_2(\theta_0^{(id)}) = 0$*,* $\mu_2(\theta^{(id)}) \neq 0$ *for* $\theta^{(id)} \neq \theta_0^{(id)}$*,* $M_2(\theta^{(id)}) = \frac{\partial \mu_2(\theta^{(id)})}{\partial \theta^{(id)\prime}}$ *has full rank at* $\theta^{(id)} = \theta_0^{(id)}$ *and is continuous.*

From Assumption 4.1, it can be seen that the population moment condition is satisfied at θ_0. At other parameter values, $E[f(v_t,\theta)]$ consists of two parts. The first part, $T^{-1/2}\mu_{1T}(\theta)$, decays to zero at rate $T^{-1/2}$ no matter what the value of θ is. Therefore, this first part is said to be *nearly uninformative* about θ_0. In contrast, the second part, $\mu_2(\theta^{(id)})$, is non-zero for any $\theta^{(id)} \neq \theta_0^{(id)}$ and so is *informative* about $\theta_0^{(id)}$. Therefore, within this framework, $\theta_0^{(wi)}$ is weakly identified but $\theta_0^{(id)}$ is identified. It is useful to relate this framework back to the identification condition in Assumption 2.5. Amending the definition of $G(f)$ to incorporate the presence of the drift in the data generation process for v_t implied by Assumption 4.1, it follows from Assumption 4.1 that

$$G(f) \;=\; \lim_{T\to\infty} E\left[T^{-1}\sum_{t=1}^{T}\frac{\partial f(v_t,\theta_0)}{\partial \theta'}\right] \;=\; \left[0_{q\times p_{wi}}, \quad M_2(\theta^{(id)})\right] \tag{4.16}$$

where $0_{a\times b}$ is a null matrix of dimension $a \times b$. Inspection reveals that the limiting derivative matrix is rank deficient and so Assumption 2.5 is violated. This framework of Stock and Wright (2000) captures the concept of weak identification in a convenient way that leads to more reliable distributional approximations. Since the property of nearly uninformative moment conditions holds for overall $\theta \in \Theta$, Stock and Wright (2000) As it will be seen, this also provides a convenient way for discussing the properties of $RMSC(c)$. This will be seen later in this chapter.

Now we impose the following high-level assumptions.

Assumption 4.2. *(i) $\Omega_{f,f}$ is finite and nonsingular; (ii) $\Omega_{i,i}$ is finite and nonsingular for $i = 1, 2$; (iii) $T^{-1}\sum_{t=1}^{T} f(v_t, \theta_0)f(v_t, \theta_0)' = T^{-1}\sum_{t=1}^{T}(z_t \otimes u_t(\theta_0))(z_t \otimes u_t(\theta_0))' \xrightarrow{p} \Omega_{f,f}$; (iv) $T^{-1}\sum_{t=1}^{T} z_t z_t' \xrightarrow{p} M_{z,z}$; (v) $T^{-1}\sum_{t=1}^{T} u_t(\theta_0)u_t(\theta_0)' \xrightarrow{p} \Sigma_{u,u}$; (vi) $T^{-1/2}\sum_{t=1}^{T} z_t \otimes u_t(\theta_0) \xrightarrow{d} N(0, \Omega_{f,f})$; (vii) $T^{-1/2}\sum_{t=1}^{T} z_t \otimes \nu_t(\theta_0) \xrightarrow{d} N(0, \Sigma_{z,\nu})$; (viii) $T^{-1/2}\sum_{t=1}^{T}\big\{u_t(\theta_0) \otimes \nu_t - E\big[u_t(\theta_0) \otimes \nu_t(\theta_0)\big]\big\} \xrightarrow{d} N(0, \Sigma_{u,\nu})$ (ix) parameter space Θ is a compact set; (x) $E\big[\sup_{\theta\in\Theta} \|f(v_t, \theta)\|\big] = E\big[\sup_{\theta\in\Theta} \|z_t \otimes u_t(\theta)\|\big] < \infty$; (xi) the random process $\{v_t : -\infty < t < \infty\}$ is ergodic; (xii) the derivative matrix $\frac{\partial f(v_t,\theta)}{\partial\theta'} = z_t \otimes \frac{\partial u_t(\theta)}{\partial\theta'}$ exists and is continuous on Θ for each $v \in \mathcal{V}$; (xiii) θ_0 is an interior point of Θ; (xiv) $E\big[\frac{\partial f(v_t,\theta)}{\partial\theta'}\big] = E\big[z_t \otimes \frac{\partial u_t(\theta)}{\partial\theta'}\big]$ exists and is finite for all $\theta \in \Theta$; (xv) $T^{-1}\sum_{t=1}^{T} f(v_t, \theta)f(v_t, \theta)'$ is a continuous function of θ which converges to $E[f(v_t, \theta)f(v_t, \theta)']$ uniformly in θ; (xvi) $T^{-1}\sum_{t=1}^{T} \frac{\partial f(v_t,\theta)}{\partial\theta'}$ is a continuous function of θ which converges to $E\big[\frac{\partial f(v_t,\theta)}{\partial\theta'}\big]$ uniformly in θ.*

Assumptions 4.2(i)–$(viii)$ are the extension of Assumption 3.1; we need additional assumptions (ix)–(xvi) for consistency and asymptotic normality of the GMM estimator in the context of nonlinear model. Within this framework, we define the near redundancy as follows.

Definition 4.1 (Near Redundancy 2). *The moment condition* $E[f_2(v_t, \theta_0)] = 0$ *is said to be nearly redundant for the estimation of* θ_0 *given* $E[f_1(v_t, \theta_0)] = 0$ *if*

$$\Omega_{2,u,T} - \Omega_{2,1}\Omega_{1,1}^{-1}\Omega_{1,u,T} = T^{-1/2}\mathsf{C}_2 \tag{4.17}$$

for some non-zero $q_2M \times p$ *matrix of constants* C_2.

It is important to notice that (4.17) implies

$$\Omega_{2,u} - \Omega_{2,1}\Omega_{1,1}^{-1}\Omega_{1,u} = 0 \tag{4.18}$$

Recall that the condition for redundancy of $E[f_2(v_t, \theta_0)] = 0$ for the estimation of θ_0 given $E[f_1(v_t, \theta_0)] = 0$ can be written as

$$\Omega_{2,u,T} - \Omega_{2,1}\Omega_{1,1}^{-1}\Omega_{1,u,T} = 0 \tag{4.19}$$

From (4.18) and (4.19), it can be noticed that nearly redundant moment conditions do not provide no incremental information in the estimation and so we can conjecture there is no efficiency gain in the limiting variance of the estimator. This intuition, in fact, turns out to be correct and is affirmed in the following result.

Theorem 4.1. *Let* $\hat{\theta}_T$ *be the Generalized Method of Moments estimator of* θ_0 *based on the moment conditions* $E[f(v_t, \theta_0)] = 0$*, using the (optimal) weighting matrix* $\Omega_{f,f}^{-1}$*. Assume that the moment condition* $E[f_2(v_t, \theta_0)] = 0$ *is nearly redundant for the estimation of* θ_0 *given* $E[f_1(v_t, \theta_0)] = 0$ *and Assumption 4.2 holds. Then the limiting distribution of this GMM estimator is:*

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0, [\Omega_{1,u}'\Omega_{1,1}^{-1}\Omega_{1,u}]^{-1}\right) \tag{4.20}$$

Proof. From Assumption 4.2, we can establish the following asymptotic normality of the GMM estimator; for example, see Hall (2005, Chapter 3).

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0, \left[\Omega'_{f,u}\Omega^{-1}_{f,f}\Omega_{f,u}\right]^{-1}\right) \tag{4.21}$$

Now consider

$$\mathsf{V} = \Omega'_{f,u}\Omega^{-1}_{f,f}\Omega_{f,u} = \begin{bmatrix} \Omega'_{1,u} & \Omega'_{2,u} \end{bmatrix} \begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix}^{-1} \begin{bmatrix} \Omega_{1,u} \\ \Omega_{2,u} \end{bmatrix} \tag{4.22}$$

Using the result of inverting partitioned matrices from Hayashi (2000, p. 673) it follows that

$$\begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix}^{-1} = \begin{bmatrix} \Omega^{-1}_{1,1} + \Omega^{-1}_{1,1}\Omega_{1,2}\mathsf{F}\Omega_{2,1}\Omega^{-1}_{1,1} & -\Omega^{-1}_{1,1}\Omega_{1,2}\mathsf{F} \\ -\mathsf{F}\Omega_{2,1}\Omega^{-1}_{1,1} & \mathsf{F} \end{bmatrix} \tag{4.23}$$

where $\mathsf{F} = (\Omega_{2,2} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,2})^{-1}$. Combining these two and collecting terms yield

$$\begin{aligned} \mathsf{V} &= \Omega'_{1,u}(\Omega^{-1}_{1,1} + \Omega^{-1}_{1,1}\Omega_{1,2}\mathsf{F}\Omega_{2,1}\Omega^{-1}_{1,1})\Omega_{1,u} \\ &\quad - \Omega'_{2,u}\mathsf{F}\Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,u} - \Omega'_{1,u}\mathsf{F}\Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{2,u} + \Omega'_{2,u}\mathsf{F}\Omega_{2,u} \\ &= \Omega'_{1,u}\Omega^{-1}_{1,1}\Omega_{1,u} + [\Omega'_{2,u} - \Omega'_{1,u}\Omega^{-1}_{1,1}\Omega_{1,2}]\mathsf{F}[\Omega_{2,u} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,u}] \\ &= \Omega'_{1,u}\Omega^{-1}_{1,1}\Omega_{1,u} + \mathsf{G}'\mathsf{F}\mathsf{G} \end{aligned} \tag{4.24}$$

where $\mathsf{G} = \Omega_{2,u} - \Omega_{2,1}\Omega^{-1}_{1,1}\Omega_{1,u}$. It follows that from (4.18) $\mathsf{G} = 0$ and so $\mathsf{V} = \Omega'_{1,u}\Omega^{-1}_{1,1}\Omega_{1,u}$ which gives the desired result. □

The main purpose of this section is to explore the connection between the near redundancy and weak identification. For this purpose, an in-depth investigation of this near redundancy condition will be useful. Using the definitions and notations introduced in the previous section, the left hand side of (4.17) can be written as follows.

$$\begin{aligned}
&\Omega_{2,u,T} - \Omega_{2,1}\Omega_{1,1}^{-1}\Omega_{1,u,T} \\
&\quad = E\left[z_{2,t} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] - E\left[z_{2,t}z_{1,t}' \otimes u_t(\theta_0)u_t(\theta_0)'\right] \\
&\qquad\quad \times \left\{E\left[z_{1,t}z_{1,t}' \otimes u_t(\theta_0)u_t(\theta_0)'\right]\right\}^{-1} E\left[z_{1,t} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] \\
&\quad = E\left[z_{2,t} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] - (M_{2,1} \otimes \Sigma_{u,u})(M_{1,1}^{-1} \otimes \Sigma_{u,u}^{-1})E\left[z_{1,t} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] \\
&\quad = E\left[\left(z_{2,t} - M_{2,1}M_{1,1}^{-1}z_{1,t}\right) \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] \\
&\quad = E\left[\tilde{m}_{2,1} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right]
\end{aligned} \tag{4.25}$$

where

$$\begin{aligned}
M_{i,j} &= E[z_{i,t}z_{j,t}'], \quad \text{for } i,j = 1,2 \\
\Sigma_{u,u} &= E[u_t(\theta_0)u_t(\theta_0)'] \\
\tilde{m}_{2,1} &= z_{2,t} - M_{2,1}M_{1,1}^{-1}z_{1,t}
\end{aligned} \tag{4.26}$$

Pre-multiplying $\tilde{m}_{2,1} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}$ by a properly chosen permutation matrix[3] $\mathcal{P}$ results in a rearrangement of the rows of $\tilde{m}_{2,1} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}$ as follows.

$$\mathcal{P}\left[\tilde{m}_{2,1} \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] = \begin{bmatrix} \tilde{m}_{2,1} \otimes \frac{\partial u_{1,t}(\theta_0)}{\partial \theta'} \\ \tilde{m}_{2,1} \otimes \frac{\partial u_{2,t}(\theta_0)}{\partial \theta'} \\ \vdots \\ \tilde{m}_{2,1} \otimes \frac{\partial u_{M,t}(\theta_0)}{\partial \theta'} \end{bmatrix} = \begin{bmatrix} \tilde{m}_{2,1}\frac{\partial u_{1,t}(\theta_0)}{\partial \theta'} \\ \tilde{m}_{2,1}\frac{\partial u_{2,t}(\theta_0)}{\partial \theta'} \\ \vdots \\ \tilde{m}_{2,1}\frac{\partial u_{M,t}(\theta_0)}{\partial \theta'} \end{bmatrix} \tag{4.27}$$

[3]A permutation matrix $\mathcal{P}$ is a square matrix with precisely one entry whose value is "1" in each column and row, and all of whose other entries are "0." Usually a permutation matrix is obtained by rearranging the rows and/or columns of an identity matrix. Pre-multiplying a matrix A by $\mathcal{P}$ results in the rearranging the rows of A and post-multiplying results in rearranging the columns of A. Note that $\mathcal{P}$ is nonsingular by construction, thus $\text{rank}\{A\} = \text{rank}\{\mathcal{P}A\} = \text{rank}\{A\mathcal{P}\}$. For more details, see Searle (1982).

Notice that $\tilde{m}_{2,1}$ is a $q_2 \times 1$ vector and $\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}$ is the m^{th} row of $\frac{\partial u_t(\theta_0)}{\partial \theta'}$. The last equality holds because $\tilde{m}_{2,1}$ and $\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}$ are column and row vectors, respectively.

Now the near redundancy condition (4.17) can be rewritten as,

$$E\left[\left(z_{2,t} - M_{2,1}M_{1,1}^{-1}z_{1,t}\right)\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\right] = E\left[\tilde{m}_{2,1}\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\right] = T^{-1/2}\mathsf{C}_{m,2} \tag{4.28}$$

for all $m = 1, 2, \ldots, M$ and $q_2 \times p$ matrix of constants $\mathsf{C}_{m,2}$. Notice that $\tilde{m}_{2,1}$ is the error vector in the linear projection of $z_{2,t}$ on $z_{1,t}$; hence, $\tilde{m}_{2,1}$ and $z_{1,t}$ are statistically independent. For the ease of comparison with weak identification, we impose the parametric restrictions within this setup. From (4.9), we consider the following linear projection.

$$\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'} = z'_{1,t}\Phi_{m,1,T}(\theta_0) + z'_{2,t}\Phi_{m,2,T}(\theta_0) + \nu_{m,t}(\theta_0) \tag{4.29}$$

where $\Phi_{m,j,T}(\theta_0)$ is a $q_j \times p$ matrix of unknown parameters, for $m = 1, 2, \ldots, M$ and $j = 1, 2$; $\nu_{m,t}(\theta_0)$ is the projection error and so $\nu_{m,t}(\theta_0)$ is statistically orthogonal to $z_{1,t}$ and $z_{2,t}$. That is, $E[\nu_{m,t}(\theta_0)\, z_t] = 0$. Hence the linear projection of $\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}$ on z_t is

$$L\left[\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\middle| z_t\right] = z'_{1,t}\Phi_{m,1,T} + z'_{2,t}\Phi_{m,2,T} \tag{4.30}$$

for $m = 1, 2, \ldots, M$. Substituting (4.29) into (4.28) yields

$$\begin{aligned} E\left[\tilde{m}_{2,1}\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\right] &= E\left[\left(z_{2,t} - M_{2,1}M_{1,1}^{-1}z_{1,t}\right)\left(z'_{1,t}\Phi_{m,1,T}(\theta_0) + z'_{2,t}\Phi_{m,2,T}(\theta_0) + \nu_{m,t}(\theta_0)\right)\right] \\ &= \left(M_{2,2} - M_{2,1}M_{1,1}^{-1}M_{1,2}\right)\Phi_{m,2,T}(\theta_0) \end{aligned} \tag{4.31}$$

Assumption 4.2 implies $M_{2,2} - M_{2,1}M_{1,1}^{-1}M_{1,2}$ is a finite and non-singular matrix of constants and so, taken together (4.28) and (4.31), it can be inferred that $E[f_2(v_t, \theta_0)] = 0$ is nearly redundant for the estimation of θ_0 given $E[f_1(v_t, \theta_0)] = 0$ *if and only if*

$$\Phi_{m,2,T}(\theta_0) = T^{-1/2}\mathsf{C}_{m,2}, \quad \text{for } m = 1, 2, \ldots, M \tag{4.32}$$

We now consider weak identification. The standard asymptotic theory for GMM estimation is based on the assumption that the population moment conditions are satisfied uniquely; that is, the unique solution for $E[f(v_t,\theta)] = 0$ for $\theta \in \Theta$ is the true parameter vector θ_0. In nonlinear models, however, it is often difficult to verify the global identification condition *a priori*. As a result, researchers confine interest to a suitably defined neighborhood of θ_0 in Θ and use the concept of local identification. The condition for local identification is rank $\left\{E\left[\frac{\partial f(v_t,\theta_0)}{\partial \theta'}\right]\right\} = p$; see Section 2.1. If this key derivative matrix $E\left[\frac{\partial f(v_t,\theta_0)}{\partial \theta'}\right]$ fails to have full column rank, the parameter is under-identified. The main interest of this chapter is not the case where the parameter is strictly under-identified, rather the case where it is weakly identified. As we described in Assumption 4.1, the key derivative matrix can be used to capture the concept of weak identification; while having full column rank in a small sample, the key derivative matrix may vanish as the sample size grows and in turn identification fails asymptotically. This circumstance is described in the following assumption on the key derivative matrix.

$$E\left[\frac{\partial f(v_t,\theta_0)}{\partial \theta'}\right] = E\left[z_t \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] = T^{-1/2}\tilde{\mathsf{C}} \tag{4.33}$$

where $\tilde{\mathsf{C}}$ is $qM \times p$ matrix of constants that has full column rank. Pre-multiplying the key derivative matrix by a carefully chosen permutation matrix $\tilde{\mathcal{P}}$ yields

$$\tilde{\mathcal{P}}E\left[\frac{\partial f(v_t,\theta_0)}{\partial \theta'}\right] = \tilde{\mathcal{P}}E\left[z_t \otimes \frac{\partial u_t(\theta_0)}{\partial \theta'}\right] = E\begin{bmatrix} z_t\frac{\partial u_{1,t}(\theta_0)}{\partial \theta'} \\ z_t\frac{\partial u_{2,t}(\theta_0)}{\partial \theta'} \\ \vdots \\ z_t\frac{\partial u_{M,t}(\theta_0)}{\partial \theta'} \end{bmatrix} \tag{4.34}$$

Now the rank of the key derivative matrix can be deduced by considering the m^{th} block of the permuted key derivative matrix, $E\left[z_t\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\right]$.

For the purpose of comparison between near redundancy and weak identification, we consider the linear projection of $\frac{\partial u_t(\theta_0)}{\partial \theta'}$ onto $z_{1,t}$ and $z_{2,t}$. To this end, rewrite (4.29) as

$$\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'} = z_{1,t}'\Phi_{m,1}(\theta_0) + z_{2,t}'\Phi_{m,2,T}(\theta_0) + \nu_{m,t}(\theta_0) \tag{4.35}$$

For $m = 1, 2, \ldots, M$ and $q_2 M \geq p$; the latter inequality is needed for the identification and the former equality is for the ease of comparison to near redundancy. Further, we assume $\Phi_{m,2,T}(\theta_0) = T^{-1/2}\tilde{\mathsf{C}}_{m,2}$; $\tilde{\mathsf{C}}_{m,2}$ is assumed to be a $q_2 \times p$ matrix of constants that has full column rank. The linear projection of $\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}$ on z_t is

$$L\left[\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\middle| z_t\right] = z_{1,t}'\Phi_{m,1}(\theta_0) + z_{2,t}'\Phi_{m,2,T}(\theta_0) \tag{4.36}$$

Now we suppose that the estimation is based on $E[f_2(v_t, \theta_0)] = 0$ alone and consider the key derivative matrix. The rank of the key derivative matrix can be deduced by considering the m^{th} block of the permuted key derivative matrix; see (4.34).

$$\begin{aligned} E\left[z_{2,t}\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\right] &= E\left[z_{2,t}z_{1,t}'\Phi_{m,1}(\theta_0) + z_{2,t}z_{2,t}'\Phi_{m,2,T}(\theta_0)\right] \\ &= M_{2,1}\Phi_{m,1}(\theta_0) + T^{-1/2}M_{2,2}\mathsf{C}_{m,2} \quad \text{for } m = 1, 2, \ldots, M \end{aligned} \tag{4.37}$$

From (4.37), it can be seen that if $\text{rank}\{M_{2,1}\Phi_{m,1}(\theta_0)\} < p$, the key derivative matrix has full column rank in the finite sample, however as sample size goes to infinity the key derivative matrix vanishes and in turn identification fails. Taken together (4.32) and (4.37), we can say that if θ_0 is weakly identified based on $E[f_2(v_t, \theta_0)] = 0$ alone, then these moment conditions become nearly redundant for θ_0 given $E[f_1(v_t, \theta_0)] = 0$ when $E[f_1(v_t, \theta_0)] = 0$ is added in, if the latter identifies θ_0. Note that for θ_0 being identified by $E[f_1(v_t, \theta_0)] = 0$, we need to have $\text{rank}\{M_{1,1}\Phi_{m,1}(\theta_0)\} = p$ and $q_1 M \geq p$..

Now it is quite natural to ask the following question; when $E[f_2(v_t, \theta_0)] = 0$ is nearly redundant for the estimation of θ_0 given $E[f_1(v_t, \theta_0)] = 0$, does the omission of $E[f_1(v_t, \theta_0)] = 0$ makes θ_0 weakly identified? The answer is yes or no. To verify this answer, recall (4.29)

$$\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'} = z'_{1,t}\Phi_{m,1,T}(\theta_0) + z'_{2,t}\Phi_{m,2,T}(\theta_0) + \nu_{m,t}(\theta_0) \tag{4.29}$$

Now we assume $\Phi_{m,1,T}(\theta_0) = \Phi_{m,1}(\theta_0)$ with $\text{rank}\{\Phi_{m,1}(\theta_0)\} = p$ and $\Phi_{m,2,T} = T^{-1/2}\mathsf{C}_{m,2}$; and so $E[f_2(v_t, \theta_0)] = 0$ is nearly redundant for θ_0 given $E[f_1(v_t, \theta_0)] = 0$. Suppose we omit $E[f_1(v_t, \theta_0)] = 0$ and base the GMM estimation $E[f_2(v_t, \theta_0)] = 0$ alone. Again we implicitly assume that $q_2 \geq p$. In this case, the m^{th} block of the permuted key derivative matrix is same as (4.37):

$$E\left[z_{2,t}\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\right] = M_{2,1}\Phi_{m,1}(\theta_0) + T^{-1/2}M_{2,2}\mathsf{C}_{m,2} \quad \text{for } m = 1, 2, \ldots, M \tag{4.37}$$

Therefore, θ_0 is identified if $M_{2,1}\Phi_{m,1}(\theta_0)$ has full column rank, but weakly identified if $M_{2,1}\Phi_{m,1}(\theta_0)$ is rank deficient. In other words, if $z_{2,t}$ inherits enough of the explanatory power of $z_{1,t}$ for $E\left[\frac{\partial u_{m,t}(\theta_0)}{\partial \theta'}\right]$, $z_{2,t}$ is correlated with the other endogenous explanatory variables just strong enough for the identification of θ_0. To sum up, the omission of $E[f_1(v_t, \theta_0)] = 0$, under the circumstance in which $E[f_2(v_t, \theta_0)] = 0$ is nearly redundant for θ_0 given $E[f_1(v_t, \theta_0)] = 0$, may or may not make θ_0 weakly identified.

4.3 Behavior of RMSC with Weak Identification in the Nonlinear Model

In this section, we explore the behavior of $RMSC(c)$ in the context of nonlinear model. In line with the previous chapter, we start with the limiting behavior of $RMSC(c)$ when weak identification is a possibility and then derive the the limiting behavour of $\hat{c}_T$. To this end, we consider the following moment function $f(v_t, \theta)$, for $\theta \in \Theta$.

$$f(v_t,\theta) = z_t \otimes u_t(\theta) = \begin{bmatrix} f_1(v_t,\theta) \\ f_2(v_t,\theta) \end{bmatrix} = \begin{bmatrix} z_{1,t} \otimes u_t(\theta) \\ z_{1,t} \otimes u_t(\theta) \end{bmatrix} \tag{4.38}$$

Now we extend Assumption 4.2 and impose the following high level assumptions.

Assumption 4.3. *For any $\theta \in \Theta$: (i) $\Omega_{f,f}(\theta) = E[f(v_t,\theta)f(v_t,\theta)']$ is finite and nonsingular; (ii) $\Omega_{i,i}(\theta) = E[f_i(v_t,\theta)f_i(v_t,\theta)']$ is also finite and nonsingular for $i = 1,2$; (iii) $T^{-1}\sum_{t=1}^{T} f(v_t,\theta)f(v_t,\theta)' \xrightarrow{p} \Omega_{f,f}(\theta)$; (iv) $T^{-1}\sum_{t=1}^{T} z_t z_t' \xrightarrow{p} M_{z,z}$; $M_{z,z}$ is finite and nonsingular; (v) $T^{-1}\sum_{t=1}^{T} u_t(\theta)u_t(\theta)' \xrightarrow{p} \Sigma_{u,u}(\theta)$; (vi) $T^{-1/2}\sum_{t=1}^{T} z_t \otimes u_t(\theta) \xrightarrow{d} N(0,\Omega_{f,f}(\theta))$; (vii) $T^{-1/2}\sum_{t=1}^{T} z_t \otimes \nu_t(\theta) \xrightarrow{d} N(0,\Sigma_{z,\nu}(\theta))$; (viii) $T^{-1/2}\sum_{t=1}^{T}\big\{u_t(\theta) \otimes \nu_t(\theta) - E\big[u_t(\theta)\otimes\nu_t(\theta)\big]\big\} \xrightarrow{d} N(0,\Sigma_{u,\nu}(\theta))$*

In GIV setting, the asymptotic variance-covariance matrix of GMM estimator is given by $V_\theta = \big[\Omega_{f,u}'\Omega_{f,f}^{-1}\Omega_{f,u}\big]^{-1}$. Given this structure, the obvious candidate for the covariance matrix estimator is

$$\begin{aligned} \hat{V}_{\theta,T} = \Bigg\{ & \Big\{T^{-1}\sum_{t=1}^{T} z_t \otimes \frac{\partial u_t(\hat{\theta}_T)}{\partial \theta'}\Big\}' \\ & \times \Big\{T^{-1}\sum_{t=1}^{T}\big(z_t \otimes u_t(\hat{\theta}_T)\big)\big(z_t \otimes u_t(\hat{\theta}_T)\big)'\Big\}^{-1}\Big\{T^{-1}\sum_{t=1}^{T} z_t \otimes \frac{\partial u_t(\hat{\theta}_T)}{\partial \theta'}\Big\}\Bigg\}^{-1} \end{aligned} \tag{4.39}$$

As in the previous chapter, now we consider the behavior of $\ln\big[|\hat{V}_{\theta,T}|\big]$ in three different scenarios. Let the data be generated via (4.9) and (4.15), then the three scenarios are:

Scenario 4.1. (θ_0 ***is weakly identified.***)

$\Phi_{i,T}^{(h)}(\theta) = T^{-1/2}\mathsf{C}_i^{(h)}(\theta)$, for $i, h = 1, 2$ and $\text{rank}\{z \otimes \mathcal{Z}'\mathsf{C}(\theta)\} = p$, where $T^{-1}\sum_{t=1}^{T} z_t \otimes \mathcal{Z}_t'\mathsf{C}(\theta) \xrightarrow{p} z \otimes \mathcal{Z}\mathsf{C}(\theta)$; $\mathcal{Z}_t' = (\mathcal{Z}_{1,t}',\ \mathcal{Z}_{2,t}')$; $\mathsf{C}(\theta) = \begin{bmatrix} \mathsf{C}_1^{(1)}(\theta) & \mathsf{C}_1^{(2)}(\theta) \\ \mathsf{C}_2^{(1)}(\theta) & \mathsf{C}_2^{(2)}(\theta) \end{bmatrix}$.

Scenario 4.2. ($\theta_{0,1}$ ***is identified but*** $\theta_{0,2}$ ***is weakly identified.***)

$\Phi_{1,T}^{(1)}(\theta) = \Phi_1^{(1)}(\theta)$ with $\text{rank}\{\Phi_1^{(1)}(\theta)\} = p_1$; $\Phi_{2,T}^{(1)}(\theta) = T^{-1/2}\mathsf{C}_2^{(1)}(\theta)$, for some matrix of constants $\mathsf{C}_2^{(1)}(\theta)$; $\Phi_{j,T}^{(2)}(\theta) = T^{-1/2}\mathsf{C}_j^{(2)}(\theta)$, for $j = 1, 2$ and $\text{rank}\big\{z \otimes \mathcal{Z}'\bar{\mathsf{C}}(\theta)\big\} = p_2$, where $T^{-1}\sum_{t=1}^{T} z_t \otimes \mathcal{Z}_t'\bar{\mathsf{C}}(\theta) \xrightarrow{p} z \otimes \mathcal{Z}\bar{\mathsf{C}}(\theta)$; $\bar{\mathsf{C}}(\theta) = \begin{bmatrix} \Phi_1^{(1)}(\theta) & \mathsf{C}_1^{(2)}(\theta) \\ 0 & \mathsf{C}_2^{(2)}(\theta) \end{bmatrix}$.

Scenario 4.3. (θ_0 ***is identified.***)

$\Phi_{1,T}^{(h)}(\theta) = \Phi_1^{(h)}(\theta) + T^{-1/2}\mathsf{C}_1^{(h)}(\theta)$ with $\text{rank}\{\Phi_1^{(h)}(\theta)\} = p_h$, for $h = 1, 2$; $\Phi_{2,T}^{(h)}(\theta) = T^{-1/2}\mathsf{C}_2^{(h)}(\theta)$ for some matrices of constants $\mathsf{C}_j^{(h)}(\theta)$, $h, j = 1, 2$.

These Scenarios 4.1–4.3 have the same structure as Scenarios 3.1–3.3. The important aspects of Scenario 3.3 described in Chapter 3 are also applicable to Scenario 4.3. It should also be noticed that the assumptions in Scenarios 4.1–4.3 hold over the entire parameter space Θ. Now we explore the limiting behavior of $\ln\big[|\hat{V}_{\theta,T}|\big]$ under these three scenarios. The results are presented in the following theorem.

Theorem 4.2. *Let the data be generated via (4.9) and (4.15) and Assumption 4.2 holds.*

(i) Under Scenario 4.1: $\ln\big[|\hat{V}_{\theta,T}|\big] = p\ln(T) + O_p(1)$;

(ii) Under Scenario 4.2: $\ln\big[|\hat{V}_{\theta,T}|\big] = p_2\ln(T) + O_p(1)$;

(iii) Under Scenario 4.3: $\ln\big[|\hat{V}_{\theta,T}|\big] = -\ln[|\Omega_{1,u}'\Omega_{1,1}^{-1}\Omega_{1,u}|] + o_p(1) = O_p(1)$.

Proof. The outline of the proof of part *(i)* and *(ii)* is as follows. Under Assumption 4.3 we show that:

$$\text{Under Scenario 4.1:} \qquad \ln\big[|V_{\theta,T}|\big] = p\ln(T) + O_p(1) \tag{4.40}$$

$$\text{Under Scenario 4.2:} \qquad \ln\big[|V_{\theta,T}|\big] = p_2\ln(T) + O_p(1) \tag{4.41}$$

where

$$V_{\theta,T} = \Bigg\{ \Big\{ T^{-1}\sum_{t=1}^{T} z_t \otimes \frac{\partial u_t(\theta)}{\partial \theta'} \Big\}' \times \Big\{ T^{-1}\sum_{t=1}^{T} \big(z_t \otimes u_t(\theta)\big)\big(z_t \otimes u_t(\theta)\big)' \Big\}^{-1} \Big\{ T^{-1}\sum_{t=1}^{T} z_t \otimes \frac{\partial u_t(\theta)}{\partial \theta'} \Big\} \Bigg\}^{-1} \tag{4.42}$$

for any $\theta \in \Theta$. Combining (4.40) and (4.41) with the following fact,

$$\inf_{\theta\in\Theta} \ln\big[|V_{\theta,T}|\big] \;\le\; \ln\big[|\hat{V}_{\theta,T}|\big] \;\le\; \sup_{\theta\in\Theta} \ln\big[|V_{\theta,T}|\big] \tag{4.43}$$

we can establish the desirable results.

In this proof, we define B_T' and D_T as

$$\begin{aligned} B_T &= \Bigg\{ T^{-1}\sum_{t=1}^{T} z_t \otimes \frac{\partial u_t(\theta)}{\partial \theta'} \Bigg\}' \\ D_T &= \Bigg\{ T^{-1}\sum_{t=1}^{T} \big(z_t \otimes u_t(\theta)\big)\big(z_t \otimes u_t(\theta)\big)' \Bigg\}^{-1} \end{aligned} \tag{4.44}$$

and let $A_T = B_T D_T B_T'$. Note that dependence of A_T, B_T, D_T and their submatrices on θ is suppressed for notational simplicity. Now consider the following partitions.

$$A_T = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}, \quad D_T = \begin{bmatrix} D_{1,1} & D_{1,2} \\ D_{2,1} & D_{2,2} \end{bmatrix}, \quad B_T = \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix} \tag{4.45}$$

where

$$\begin{aligned}
A_{1,1} &= B_{1,1}D_{1,1}B'_{1,1} + B_{1,2}D_{2,1}B'_{1,1} + B_{1,1}D_{1,2}B'_{1,2} + B_{1,2}D_{2,2}B'_{1,2} \\
A_{1,2} &= B_{1,1}D_{1,1}B'_{2,1} + B_{1,2}D_{2,1}B'_{2,1} + B_{1,1}D_{1,2}B'_{2,2} + B_{1,2}D_{2,2}B'_{2,2} \\
A_{2,1} &= B_{2,1}D_{1,1}B'_{1,1} + B_{2,2}D_{2,1}B'_{1,1} + B_{2,1}D_{1,2}B'_{1,2} + B_{2,2}D_{2,2}B'_{1,2} \\
A_{2,2} &= B_{2,1}D_{1,1}B'_{2,1} + B_{2,2}D_{2,1}B'_{2,1} + B_{2,1}D_{1,2}B'_{2,2} + B_{2,2}D_{2,2}B'_{2,2}
\end{aligned} \tag{4.46}$$

Notice that T subscription is suppressed for the notational brevity. From Dhrymes (2000, Proposition 2.30), the determinant of partitioned matrix A_T can be calculated as

$$\left|A_T\right| = \left|A_{2,2}\right| \left|A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}\right| \tag{4.47}$$

It is useful to examine the partition of B_T in more details. Using the partition scheme in (4.15), write B'_T as follows

$$\begin{aligned}
B'_T &= \begin{bmatrix} B'_{1,1} & B'_{2,1} \\ B'_{1,2} & B'_{2,2} \end{bmatrix} \\
&= T^{-1} \sum_{t=1}^{T} z_t \otimes \frac{\partial u_t(\theta)}{\partial \theta'} = T^{-1} \sum_{t=1}^{T} \left[\left(\begin{smallmatrix} z_{1,t} \\ z_{2,t} \end{smallmatrix} \right) \otimes \left(\begin{smallmatrix} \frac{\partial u_t(\theta)}{\partial \theta'_1} & \frac{\partial u_t(\theta)}{\partial \theta'_2} \end{smallmatrix} \right) \right] \\
&= T^{-1} \sum_{t=1}^{T} \begin{bmatrix} z_{1,t} \otimes \frac{\partial u_t(\theta)}{\partial \theta'_1} & z_{1,t} \otimes \frac{\partial u_t(\theta)}{\partial \theta'_2} \\ z_{2,t} \otimes \frac{\partial u_t(\theta)}{\partial \theta'_1} & z_{2,t} \otimes \frac{\partial u_t(\theta)}{\partial \theta'_2} \end{bmatrix}
\end{aligned} \tag{4.48}$$

The last equality holds since z_t is a *column* vector. Now we consider the limiting behavior of $\ln\big[|A_T|\big]$.

part (i): Assumption 4.3 implies that $D_T = O_p(1)$ and $D_{i,j} = O_p(1)$ for $i, j = 1, 2$. Since each $B_{i,j}$ behaves the same way under Scenario 4.1, we consider an arbitrary block of B_T' in (4.48). Using (4.15) and parametric assumptions in Scenario 4.1, we write $(i,h)^{th}$ block of B_T' as follows.

$$T^{-1}\sum_{t=1}^{T} z_{i,t}\otimes\left(\mathcal{Z}_{1,t}'\Phi_{1,T}^{(h)}(\theta)+\mathcal{Z}_{2,t}'\Phi_{2,T}^{(h)}(\theta)+\nu_t^{(h)}(\theta)\right)$$
$$=T^{-1/2}\Bigg\{T^{-1}\sum_{t=1}^{T} z_{i,t}\otimes\mathcal{Z}_{1,t}'\mathsf{C}_1^{(h)}(\theta)+T^{-1}\sum_{t=1}^{T} z_{i,t}\otimes\mathcal{Z}_{2,t}'\mathsf{C}_2^{(h)}(\theta)$$
$$+T^{-1/2}\sum_{t=1}^{T} z_{i,t}\otimes\nu_t^{(h)}(\theta)\Bigg\} \quad (4.49)$$

Assumption 4.3 implies $T^{-1/2}\sum_{t=1}^{T} z_{i,t}\otimes\nu_t^{(h)}(\theta)=O_p(1)$. Now we consider $z_{i,t}\otimes\mathcal{Z}_{j,t}\mathsf{C}_j^{(h)}(\theta)$ for $i,j=1,2$. After rearranging the rows by pre-multiplying a proper permutation matrix, a block of the permuted matrix can be written as

$$z_{i,t}\otimes z_{j,t}'\mathsf{C}_j^{(h)}(\theta)=z_{i,t}\,z_{j,t}'\mathsf{C}_j^{(h)}(\theta) \quad (4.50)$$

This equality holds because $z_{i,t}$ and $z_{j,t}'\mathsf{C}_j^{(h)}(\theta)$ are vectors. By expanding the right hand side, it can be seen that each entry of this matrix is a sum of products of any two elements of z_t and one element of $\mathsf{C}_j(\theta)$, which means that $T^{-1}\sum_{t=1}^{T} z_{i,t}\otimes z_{j,t}'\mathsf{C}_j^{(h)}(\theta)$ is bounded. Hence it follows that $B_{i,j}=O_p(T^{-1/2})$ and $B_T=O_p(T^{-1/2})$. Since we defined $A_T=B_TD_TB_T'$, we have $A_T=O_p(T^{-1})$. Now we define $\tilde{A}_T=TA_T$ and $\tilde{A}_{i,j}=TA_{i,j}$. Notice that $\tilde{A}_T=O_p(1)$ by construction. We now show that $\tilde{A}_T$ is positive definite with probability one in the limit as $T\to\infty$. For this purpose, we write $\tilde{A}_T=\tilde{B}_TD_T\tilde{B}_T'$ where $\tilde{B}_T=T^{1/2}B_T$, and consider the following quadratic form: $\alpha'\tilde{A}_T\alpha = \left\{\alpha'(T^{1/2}B_T)\right\}D_T\left\{(T^{1/2}B_T')\alpha\right\}=\alpha'\tilde{B}_TD_T\tilde{B}_T'\alpha$; for some non-zero vector α. Since D_T is positive definite by construction, we need to consider

$(T^{1/2}B'_T)\alpha = \tilde{B}'_T\alpha$. From (4.49), the $(i,h)^{th}$ block of $\tilde{B}'_T\alpha$ can be written as

$$\tilde{B}'_{i,h}\alpha = \left\{T^{-1}\sum_{t=1}^{T} z_{i,t}\otimes \mathcal{Z}'_{1,t}\mathsf{C}_1^{(h)} + T^{-1}\sum_{t=1}^{T} z_{i,t}\otimes \mathcal{Z}'_{2,t}\mathsf{C}_2^{(h)} + T^{-1/2}\sum_{t=1}^{T} z_{i,t}\otimes \nu_t^{(h)}(\theta)\right\}\alpha \tag{4.51}$$

From Assumption 4.3, we know that $T^{-1/2}\sum_{t=1}^{T} z_{i,t}\otimes\nu_t^{(h)}(\theta)$ converges in distribution to a normal random matrix. From the aforementioned results, we also know that $T^{-1}\sum_{t=1}^{T} z_{i,t}\otimes \mathcal{Z}'_{j,t}\mathsf{C}_j^{(h)}(\theta) = O_p(1)$. Let its probability limit be denoted by $z_i\otimes \mathcal{Z}'_j\mathsf{C}_j^{(h)}(\theta)$. Then we have

$$\tilde{B}'_{i,h}\alpha \xrightarrow{d} \left\{z_i\otimes \mathcal{Z}'_1\mathsf{C}_1^{(h)}(\theta) + z_i\otimes \mathcal{Z}'_2\mathsf{C}_2^{(h)}(\theta)\right\}\alpha + \tilde{B}_{i,h,\text{normal}}\alpha \tag{4.52}$$

Finally, $\tilde{B}'_T\alpha$ can be written as

$$\begin{aligned}
\tilde{B}'_T\alpha \xrightarrow{d} & \begin{bmatrix} z_1\otimes \mathcal{Z}'_1\mathsf{C}_1^{(1)}(\theta) + z_1\otimes \mathcal{Z}'_2\mathsf{C}_2^{(1)}(\theta) & z_1\otimes \mathcal{Z}'_1\mathsf{C}_1^{(2)}(\theta) + z_1\otimes \mathcal{Z}'_2\mathsf{C}_2^{(2)}(\theta) \\ z_2\otimes \mathcal{Z}'_1\mathsf{C}_1^{(1)}(\theta) + z_2\otimes \mathcal{Z}'_2\mathsf{C}_2^{(1)}(\theta) & z_2\otimes \mathcal{Z}'_1\mathsf{C}_1^{(2)}(\theta) + z_2\otimes \mathcal{Z}'_2\mathsf{C}_2^{(2)}(\theta) \end{bmatrix}\alpha \\
& + \tilde{B}'_{\text{normal}}\alpha \\
= & \left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \otimes \begin{pmatrix} \mathcal{Z}'_1 & \mathcal{Z}'_2 \end{pmatrix} \begin{pmatrix} \mathsf{C}_1^{(1)}(\theta) & \mathsf{C}_1^{(2)}(\theta) \\ \mathsf{C}_2^{(1)}(\theta) & \mathsf{C}_2^{(2)}(\theta) \end{pmatrix}\right]\alpha + \tilde{B}'_{\text{normal}}\alpha \\
= & \left[z\otimes \mathcal{Z}'\mathsf{C}(\theta)\right]\alpha + \tilde{B}'_{\text{normal}}\alpha
\end{aligned} \tag{4.53}$$

where $z = (z'_1,\ z'_2)'$ and $\mathcal{Z}' = (\mathcal{Z}'_1,\ \mathcal{Z}'_2)$. Thus, if $\text{rank}\{z\otimes\mathcal{Z}'\mathsf{C}(\theta)\} = p$, $\tilde{A}_T$ is positive definite with probability one in the limit as $T\to\infty$. Returning to A_T, if we substitute in (4.47) for $A_{i,j}$ in terms of $\tilde{A}_{i,j}$ then we obtain

$$\begin{aligned} \left|A_T\right| &= \left|T^{-1}\tilde{A}_{2,2}\right| \left|T^{-1}\tilde{A}_{1,1} - T^{-1}\tilde{A}_{1,2}\{T^{-1}\tilde{A}_{2,2}\}^{-1}T^{-1}\tilde{A}_{2,1}\right| \\ &= T^{-p}\left|\tilde{A}_T\right| \\ &= O_p(T^{-p}) \end{aligned} \tag{4.54}$$

where the last equality follows from the properties of $\tilde{A}_T$ derived above. This proves (4.40). Finally, by the fact in (4.43), the desired result then follows.

part (ii): Again, Assumption 4.3 implies that $D_T = O_p(1)$ and $D_{i,j} = O_p(1)$ for $i, j = 1, 2$. Now consider the orders in probability of B_T. From (4.15), (4.48), and the parametric restrictions of Scenario 4.2, we can find the orders of $B_{i,j}$'s as follows:

$$B'_{1,i} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}'_{1,t}\Phi_1^{(1)}(\theta) + T^{-\frac{1}{2}}\Big\{\frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}'_{2,t}\mathsf{C}_2^{(1)}(\theta) + T^{-\frac{1}{2}}\sum_{t=1}^{T} z_{i,t} \otimes \nu_t^{(1)}(\theta)\Big\}$$

$$B'_{2,i} = T^{-\frac{1}{2}}\Big\{\frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}'_{1,t}\mathsf{C}_1^{(2)}(\theta) + \frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}'_{2,t}\mathsf{C}_2^{(2)}(\theta) + T^{-\frac{1}{2}}\sum_{t=1}^{T} z_{i,t} \otimes \nu_t^{(2)}(\theta)\Big\}$$

for $i = 1, 2$. Using the same logic as in *part (i)* and the results of White (1984, Proposition 2.30 and Exercise 2.35), we can conclude as follows.

$$\begin{aligned} B'_{1,i} &= O_p(1) + T^{-1/2}O_p(1) = O_p(1) \\ B'_{2,i} &= T^{-1/2}O_p(1) = O_p(T^{-1/2}) \end{aligned} \tag{4.55}$$

Using these results in (4.46) yields the orders in probability for each $A_{i,j}$ as follows.

$$A_{1,1} = O_p(1) \qquad A_{1,2} = O_p(T^{-1/2})$$

$$A_{2,1} = O_p(T^{-1/2}) \qquad A_{2,2} = O_p(T^{-1})$$

Define

$$\bar{A}_T = \begin{bmatrix} \bar{A}_{1,1} & \bar{A}_{1,2} \\ \bar{A}_{2,1} & \bar{A}_{2,2} \end{bmatrix} = \begin{bmatrix} A_{1,1} & T^{1/2}A_{1,2} \\ T^{1/2}A_{2,1} & TA_{2,2} \end{bmatrix} \tag{4.56}$$

where $\bar{A}_{1,1} = A_{1,1}$, $\bar{A}_{1,2} = T^{1/2}A_{1,2}$, $\bar{A}_{2,1} = T^{1/2}A_{2,1}$ and $\bar{A}_{2,2} = TA_{2,2}$. Now we consider the properties of $\bar{A}_T$. Note $\bar{A}_T = O_p(1)$ by construction. We now show that $\bar{A}_T$ is positive definite with probability one in the limit as $T \to \infty$. To do this, we define,

$$\bar{B}_T = \begin{bmatrix} B_{1,1} & B_{1,2} \\ T^{1/2}B_{2,1} & T^{1/2}B_{2,2} \end{bmatrix} \tag{4.57}$$

Then, $\bar{A}_T$ can be written as $\bar{A}_T = \bar{B}_T D_T \bar{B}'_T$. By the same logic as in *part (i)*, we need to consider $\bar{B}'_T\alpha$ for some non-zero vector α. Each block of $\bar{B}_T$ is:

$$\bar{B}'_{1,i} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}'_{1,t}\Phi_1^{(1)}(\theta) + T^{-\frac{1}{2}}\left\{\frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}'_{2,t}\mathsf{C}_2^{(1)}(\theta) + T^{-\frac{1}{2}}\sum_{t=1}^{T} z_{i,t} \otimes \nu_t^{(1)}(\theta)\right\}$$

$$\bar{B}'_{2,i} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}_{1,t}\mathsf{C}_1^{(2)}(\theta) + \frac{1}{T}\sum_{t=1}^{T} z_{i,t} \otimes \mathcal{Z}_{2,t}\mathsf{C}_2^{(2)}(\theta) + T^{-\frac{1}{2}}\sum_{t=1}^{T} z_{i,t} \otimes \nu_t^{(2)}(\theta)$$

for $i = 1, 2$. Similarly, $\bar{B}'_T\alpha$ can be written as

$$\begin{aligned} \bar{B}'_T\alpha &\xrightarrow{d} \begin{bmatrix} z_1 \otimes \mathcal{Z}'_1\Phi_1^{(1)}(\theta) & z_1 \otimes \mathcal{Z}'_1\mathsf{C}_1^{(2)}(\theta) + z_1 \otimes \mathcal{Z}'_2\mathsf{C}_2^{(2)}(\theta) \\ z_2 \otimes \mathcal{Z}'_1\Phi_1^{(1)}(\theta) & z_2 \otimes \mathcal{Z}'_1\mathsf{C}_1^{(2)}(\theta) + z_2 \otimes \mathcal{Z}'_2\mathsf{C}_2^{(2)}(\theta) \end{bmatrix}\alpha + \begin{bmatrix} 0 & \bar{B}'_{2,1,\text{normal}} \\ 0 & \bar{B}'_{2,2,\text{normal}} \end{bmatrix}\alpha \\ &= \left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \otimes \begin{pmatrix} \mathcal{Z}'_1 & \mathcal{Z}'_2 \end{pmatrix}\begin{pmatrix} \Phi_1^{(1)}(\theta) & \mathsf{C}_1^{(2)}(\theta) \\ 0 & \mathsf{C}_2^{(2)}(\theta) \end{pmatrix}\right]\alpha + \bar{B}'_{\text{normal}}\alpha \\ &= \left[z \otimes \mathcal{Z}'\bar{\mathsf{C}}(\theta)\right]\alpha + \bar{B}'_{\text{normal}}\alpha \end{aligned} \tag{4.58}$$

Thus, we can conclude that $\bar{A}_T$ is positive definite with probability one in the limit as $T \to \infty$, if rank $\{z \otimes \mathcal{Z}'\bar{\mathsf{C}}(\theta)\} = p_2$. Substituting in (4.47) for $A_{i,j}$ in terms of $\bar{A}_{i,j}$ gives

$$\begin{aligned} |A_T| &= |T^{-1}\bar{A}_{2,2}|\,|\bar{A}_{1,1} - T^{-1/2}\bar{A}_{1,2}\{T^{-1}\bar{A}_{2,2}\}^{-1}T^{-1/2}\bar{A}_{2,1}| \\ &= T^{-p_2}|\bar{A}_T| \\ &= O_p(T^{-p_2}) \end{aligned} \tag{4.59}$$

Then the desired result follows.

part (iii): Since $\ln(\cdot)$ is a continuous function and determinant of a matrix is also a continuous function of the entries of the matrix in question, it follows from Assumption 4.2, Slutsky's Theorem, and Theorem 4.1 that

$$\ln\big[|\hat{V}_{\theta,T}|\big] \xrightarrow{p} \ln\big(|\Omega'_{1,u}\Omega_{1,1}^{-1}\Omega_{1,u}|\big) \tag{4.60}$$

This proves (4.41). Combining this result and the fact in (4.43), we get the desired result. □

The implication of Theorem 4.2 for the moment selection based on $RMSC(c)$ in the nonlinear model with weak identification is the same as that to Theorem 3.2 in Chapter 3. Since in this nonlinear setting, $\ln\big[|\hat{V}_{\theta,T}|\big]$ behaves the same way as in the linear setting in the previous chapter, we can easily establish the consistency of $RMSC(c)$ for c_r in line with the linear case. Due to this similarity, it may be redundant to include the following results here. However, these results are presented for merely the completion of this chapter.

In this particular nonlinear model, the chosen selection vector can be expressed as

$$\tilde{c}_T = \text{Argmin}_{c \in C} \left\{ \ln\left[|\hat{V}_{\theta,T}|\right] + \kappa(|c|, T) \right\} \tag{4.61}$$

where $\ln[|\hat{V}_{\theta,T}|]$ is given as in (4.39).

We now require two additional assumptions that specify respectively a partition of C and an identification condition.

Assumption 4.4. *$C = C_I \cup C_{II} \cup C_{III}$ where C_I yields models that fit within Scenario 4.1, C_{II} yields models that fit within Scenario 4.2 and C_{III} yields models that fit within Scenario 4.3.*

Assumption 4.5. *There is a $c_r \in C_{III}$ that satisfies the properties in Definition 2.4 with $q_{max} = q$ and $\mathcal{C}_{min} = \{c_r\}$.*

Finally, we establish the following consistency result.

Theorem 4.3. *Let the data be generated as described in Section 4.1 and Assumptions 2.7 and 4.2–4.5 hold, then $\tilde{c}_T \xrightarrow{p} c_r$.*

Proof. From Theorem 4.2(i)-(ii), it follows that $RMSC(c) \to \infty$ as $T \to \infty$ with probability one for all $c \in C_I \cup C_{II}$. From Theorem 4.2(iii), $RMSC(c) = O_p(1)$ for

$c \in C_{III}$. Therefore, $\lim_{T\to\infty} \text{Prob}(\hat{c}_T \in C_{III}) = 1$. The rest of the proof follows by the same argument as the proof of Theorem 2.2. $\square$

Theorem 3.3 establishes the consistency of $RMSC(c)$ for c_r even when subsets of the candidate set provide only weak identification. However, it should be noted that the result in Theorem 3.3 is only valid in the context of linear IV model estimated via 2SLS. Theorem 4.3 extends Theorem 3.3 and establishes the consistency of $RMSC(c)$ with possible weak identification in more general context of nonlinear GMM model represented by GIV. This analytic result is the major contribution of this dissertation. In the next chapter, the simulation design and result will be presented. These results reveal the evidence for the usefulness of our entropy based moment selection procedure.

Chapter 5

Simulation Design and Results

In the previous chapters, we introduce the entropy based moment selection criterion and establish the consistency of $RMSC(c)$ under the presence of weak identification for both linear and nonlinear model. In this chapter, we conduct a Monte Carlo simulation and verify the validity of the analytical results in the previous chapters.

5.1 Simulation Design

To conclude this dissertation, we explore the finite sample behavior of $RMSC(c)$ in a setting where weak identification is a possibility. To this end, a Monte Carlo simulation study was undertaken using the following data generating process:

$$y_t = \theta_0 x_t + u_t \tag{5.1}$$

$$x_t = \pi' z_t + e_t \tag{5.2}$$

For simplicity of calculation, we assumed that both y_t and x_t are scalars. No exogenous regressors are included in the structural equation and there is only one endogenous regressor, x_t. The true value of the structural parameter set to be zero, that is $\theta_0 = 0$. The scalars u_t and e_t are disturbance terms of the structural and the first stage equations, respectively. z_t is a $q \times 1$ vector of instruments which are chosen from $q_{max} \in \{8, 16\}$ candidates via some criterion, such as $RMSC(c)$. For the fixed true value of $\theta_0 = 0$ and for certain specifications of π, random samples are generated under the assumption that $[u_t, e_t, z_t']' \sim NID(0_{(q_{max}+2)\times 1}, \Sigma)$ where Σ is a matrix whose diagonal elements are all equal to one and whose only non-zero off diagonal elements are the $(1,2)$ and $(2,1)$ entries which are equal to σ_{ue}. We choose σ_{ue} from the set $\sigma_{ue} \in \{0.1, 0.5, 0.9\}$. The simulation is based on 10,000 Monte Carlo draws for sample size $T = 100$. Following the conventional practice in econometrics, we also use the $T \times 1$ matrix X and $T \times q_{max}$ matrix Z whose t^{th} rows are x_t and z_t', respectively.

The choice of $q_{max} \times 1$ first stage parameter vector π is made so that the concentration parameter in population satisfies a certain value. That is, the following relationship is satisfied by the choice of π.

$$\frac{\lambda'\lambda}{q} = \frac{\pi' E[Z'Z]\pi}{q_{max}\sigma_e^2} = \frac{T}{q_{max}} \sum_{i=1}^{q_{max}} \pi_i^2 \tag{5.3}$$

The second equality holds for this certain data generating process in (5.1) and (5.2). For the specifications that determines each entry of π, $\{\pi_i : i = 1, 2, \ldots, q_{max}\}$, we consider the following three distinct cases.

1. **$\mathbf{ID}_\pi = \mathbf{I}$:** In this specification, only the first instrument is relevant and so only the first element of π is non-zero. π is determined by

$$\pi = (\gamma_1, 0'_{(q_{max}-1)\times 1})' \tag{5.4}$$

It is worth noting that given the first instrument all the other instruments are redundant and so $c_r = (1, 0'_{(q_{max}-1)\times 1})'$;

2. **$\text{ID}_\pi = \text{II}$:** In this specification, all the instruments are equally important and so all the elements of π are equal. π is determined by

$$\pi = (\gamma_2, \gamma_2, \ldots, \gamma_2)' \tag{5.5}$$

In this case, it is clear that $c_r = \iota_{q_{max}\times 1}$.

3. **$\text{ID}_\pi = \text{III}$:** In this specification, we assume that some prior information on the order of importance of instruments are known and so we have that $\pi_1 \geq \pi_2 \geq \ldots \geq \pi_{q_{max}}$. Following Donald and Newey (2001), we determine the i^{th} element of π as follows.

$$\pi_i = \gamma_3\Big(1 - \frac{i}{q_{max}+1}\Big)^4 \quad \text{for} \quad i = 1, \ldots, q_{max} \tag{5.6}$$

Again, we have that $c_r = \iota_{q_{max}\times 1}$.

The values of constants γ_1, γ_2 and γ_3 are chosen so that (5.3) holds for the set $\frac{\lambda'\lambda}{q} \in \{1, 10, 25, 50\}$. The concentration parameter is widely accepted as a unitless measure for weak identification. When the value of concentration parameter is low, we suspect that we may have weak identification problem. As it will be discussed later, we also add test for detecting weak identification in our simulation design.

Within this framework the 2SLS estimator is asymptotically equivalent to the GMM estimator with the optimal weighting matrix.[1] For the 2SLS estimator, we

[1] Recall 2SLS is the GMM estimator based on $E[z_t u_t] = 0$ with weighting matrix $W_T = (T^{-1}Z'Z)^{-1}$.

use the following consistent estimator of the asymptotic variance-covariance matrix to compute the relevant moment selection criterion.

$$\hat{V}_{\theta,T}(c) \;=\; \hat{\sigma}_u^2(c)\left[T^{-1}\sum_{t=1}^{T} x_t z_t(c)'\left(T^{-1}\sum_{t=1}^{T} z_t(c)z_t(c)'\right)^{-1} T^{-1}\sum_{t=1}^{T} z_t(c)x_t\right]^{-1} \tag{5.7}$$

where $\hat{\sigma}_u^2(c) \;=\; T^{-1}\sum_{t=1}^{T}\hat{u}_t^2(c)$, $\hat{u}_t(c) = y_t - \hat{\theta}_T(c)x_t$ and $\hat{\theta}_T(c)$ is the 2SLS estimator of θ_0 based on $E[z_t(c)(y_t - x_t\theta)] = 0$. Note that these variables are indexed by c denoting that they are based on $z_t(c)$ which is the instrument vector chosen by the selection vecotr c. The moment selection procedure is implemented with the penalty term associated with the BIC-type criterion, and so

$$RMSC(c) \;=\; \ln\left[\left|\hat{V}_{\theta,T}(c)\right|\right] \;+\; \frac{(c'c-1)\ln(T^{1/2})}{T^{1/2}}. \tag{5.8}$$

Now we express the chosen selection vector as

$$\hat{c}_T = \text{Argmin}_{c\in C}\left\{RMSC(c) \text{ in } (5.8)\right\} \tag{5.9}$$

When conducting the relevant moment selection procedure via $RMSC(c)$ in (5.8), we consider two distinct specifications for constructing the candidate set C. First, we consider the case where $RMSC(c)$ is minimized over the following $q_C \in \{8, 16\}$ choices of c: $c = (1'_{q\times 1}0'_{(q_{max}-q)\times 1})'$, $q = 1, 2, ..., q_{max}$. Notice that in this sequential design $q_C = q_{max}$, where q_C is the cardinality of the candidate set C. Second, we consider the case where $RMSC(c)$ is minimized over all $q_C = 255$ possible combinations[2] of $q_{max} = 8$ instruments. For the case of $\mathbf{ID}_\pi = \mathbf{I}$ (only one relevant instrument), we group these 255 possibilities into four cases: $1R$, $1R/I^*$, I and All, where $1R$ denotes $c = (1, 0_7')'$, indicating the that the selection vector consists of only one relevant instrument; $1R/I^*$ denotes $c = (1, a')'$, $a \neq 0_7$ or ι_7, implying that the

[2]Since we must have $|c| \geq p$, this number should be smaller than $2^8 = 256$. In our simulation design this number is 255 because $p = 1$.

selection vector consists of the relevant and some but not all irrelevant instruments; I denotes $c = (0, b')'$, $b \neq 0_7$, meaning that the selection vector consists of no relevant instrument and some or all irrelevant instruments; and All denotes $c = \iota_8$, that is, the selection vector consists of all the instruments. For the case of $\mathbf{ID}_\pi = \mathbf{II}$ (equally important instrument), we group the 255 cases into eight categories: $1Z, 2Z, \ldots, 7Z$ and $8Z$, where xZ implies that $|c| = x$ for $x = 1, 2, \ldots, 8$, meaning x instruments are included in the selection vector. For the case of $\mathbf{ID}_\pi = \mathbf{III}$ (ordered instrument by the importance), we group these 255 candidates into nine cases: $1S, 2S, \ldots, 8S$ and $Others$ where xS denotes $c = (\iota_x', 0_{8-x}')'$, for $x = 1, \ldots, 8$, implying the selection vector consists of the first x important instruments; $Others$ denotes all the other c's.

In this simulation, we also conduct a test for detecting weak identification as a pretest prior to applying $RMSC(c)$. To this end, we use the test proposed by Stock and Yogo (2002). They defined weak identification as the situation where $\lambda'\lambda/q$ is small enough that the inference based on the standard t-statistic or Wald statistic is misleading. For the test based on this definition of weak identification, they consider four types of tolerances for the deviation from the usual standards of inference: 2SLS bias, 2SLS size, Fuller-k bias and LIML size. Among these four we will consider the tests based on 2SLS bias and size. They also tabulate the critical value for both tests based on the weak identification asymptotics. See Stock and Yogo (2002, Tables 1–4). In our simulation design, the test statistic is:

$$F_{SY} = \frac{X'Z(Z'Z)^{-1}Z'X}{q\hat{\sigma}_e^2} \tag{5.10}$$

where $\hat{\sigma}_e^2 = X'(I - Z(Z'Z)^{-1}Z')X/(T - q)$. Since there is only one endogenous regressor, this test statistic is simply the "first stage F-statistic" for testing joint insignificance of the slope coefficients in the first stage equation. When this test statistic is equal to or greater than proper critical value, we reject the null hypothesis that the instruments are weak. Now the selection vector chosen by our moment selection procedure can be expressed as

$$\check{c}_{T,\varpi} = \begin{cases} \iota_{q_{max}} & \text{, if } H_0 \text{ is not rejected.} \\ \hat{c}_T \text{ in (5.9)} & \text{, if } H_0 \text{ is rejected.} \end{cases} \tag{5.11}$$

,where $\varpi = \{b, s\}$ denotes Stock and Yogo's (2002) weak identification test based on 2SLS bias and size, respectively. As for the tolerance values, we use $b = 0.05$ for the test based on bias and $r = 0.10$ for the test based on size.

5.2 Simulation Results

In this section, we report and summarize the simulation results.

Table 5.1 reports the median bias of the 2SLS estimator and the empirical coverage rate and the width of 95% confidence interval for both Wald and Anderson–Rubin tests[3] when always using all $q_{max} = 8$ instrument candidates. Wald test exhibits serious distortions in the empirical coverage rates especially with lower $\frac{\lambda'\lambda}{q}$ and higher endogeneity of the regressors. As the concentration parameter value grows, these distortions tend to disappear. On the other hand, Anderson–Rubin test yields the empirical coverage rates pretty close to the nominal value 0.95 through out all the cases. This discrepancy between the reported empirical coverage rates of Anderson–rubin statistic and its nominal value is due to the use of asymptotic χ^2 distribution for calculating the critical value of the test. Staiger and Stock (1997) suggest to use the exact F distribution, since this is more conservative than the asymptotic χ^2

[3]Hahn and Inoue (2002) explain how to calculate the width of Anderson–Rubin confidence interval.

distribution. Even though it is not reported here, our observations confirm Staiger and Stock's (1997) suggestion. It should be also noted that the 95% confidence interval of Anderson–Rubin test is longer than that of Wald test.

Table 5.2 reports the empirical selection probabilities of $RMSC(c)$ over the sequential choice of 8 instruments. The median bias, empirical coverage rates and width of confidence interval for Wald and Anderson–Rubin tests for the post–selection estimator are also reported. In case $\mathbf{ID}_\pi = \mathbf{I}$, we observe improvements in almost all areas. Even though there is a slight loss in the length of confidence interval of Wald test, the size distortion gets less severe. The width of confidence interval of Anderson–Rubin test shrinks. From $\frac{\lambda'\lambda}{q} = 10$, $RMSC(c)$ is very successful to choose the relevant instruments only. In the case of $\mathbf{ID}_\pi = \mathbf{III}$, we observe similar improvements. One particular observation is that $\frac{\lambda'\lambda}{q} = 25$, $RMSC(c)$ selects no more than the first three important instruments. Interestingly, in case $\mathbf{ID}_\pi = \mathbf{II}$ where all the instruments are equally important, it seems that $RMSC(c)$ does not help. The median bias of the post–selection estimator is larger than the results reported in Table 5.1 and the width of confidence intervals of both Wald and Anderson–Rubin tests are wider. Another interesting observation is the size distortion of the Anderson–Rubin test. This distortion is observed with weak instruments and higher endogeneity through out all the specifications of π.

The results reported in Tables 5.3 and 5.4 are associated with Stock and Yogo's (2002) pretest based on 2SLS bias and size, respectively. The test threshold values are $b = 0.05$ and $r = 0.10$. For both tests, we use all 8 instruments when the pretest does not reject the null hypothesis that instruments are weak; we conduct $RMSC(c)$ over 8 sequential choices of instruments when the null hypothesis is rejected. When $\frac{\lambda'\lambda}{q} = 1$, both tests reject the null with probability one and so the results in this category are basically same as the those in Table 5.1. The slight differences are due to the random number generating process of the computer software. By this reason, it is natural that the size distortions in Anderson–Rubin test observed in Table 5.2 disappear. Comparing the results in Tables 5.3 and 5.4, we expect that the test based on 2SLS size is more conservative than the test based on 2SLS bias in the sense that it tends to accept the false null hypothesis more often. This means the weak identification based on 2SLS size has a lager type II error and in turn smaller power. All the other results are very similar to the previous findings.

In the next three Tables 5.5–5.7, we report the results using $RMSC(c)$ associated with AIC–type penalty, which is given by $\frac{2(c'c-1)}{T^{1/2}}$. Even though, this AIC–type penalty does not satisfy Assumption 2.7 and so the $RMSC(c)$ associated with this type penalty is not consistent, all the results are pretty much the same as previous ones. It is worth noting that since AIC–type penalty is more lenient than BIC–type penalty, $RMSC(c)$ tends to select larger number of instruments. Especially when $\mathbf{ID}_\pi = \mathbf{II}$, empirical selection probability of choosing 6, 7 anad 8 instruments is now higher.

We also consider the behavior of $RMSC(c)$ with BIC–type penalty term over all 255 possible combinations of 8 instruments. The results are reported in Tables 5.8–5.10. The results suggest the validity of the relatively simple sequential strategy: for the cases of $\mathbf{ID}_\pi = \mathbf{I}$ and $\mathbf{ID}_\pi = \mathbf{III}$ the behavior of $RMSC(c)$ very similar to its counterpart with sequential strategy; however, we observe slight deteriorations in the quality of the small sample behavior of $RMSC(c)$ when $\mathbf{ID}_\pi = \mathbf{II}$.

The results above suggest that the proposed entropy based moment selection procedure can identify the relevant instruments with high probability and improve the quality of subsequent inferences in cases where $\mathbf{ID}_\pi = \mathbf{I}$ and $\mathbf{ID}_\pi = \mathbf{III}$. Especially, in case $\mathbf{ID}_\pi = \mathbf{I}$, $RMSC(c)$ seems very successful without the Stock and Yogo (2002) pretest. However, in the case where $\mathbf{ID}_\pi = \mathbf{II}$, that is, all the candidates for instruments are equally important, it seems that $RMSC(c)$ does not help or does some harms. We conjecture this deterioration may be due to the small number of possible candidates in set C. The rest of the tables reports the simulation results with sequential design of $q_{max} = 16$ possible choices of instruments.

Comparing the results in Table 5.11 to those in Table 5.1, we observe improvements in width of 95% confidence intervals of Anderson–Rubin test. Even though the 95% confidence intervals of Wald test also reduce, Wald test displays more severe size distortion with high endogeneity of regressors, that is, $\sigma_{ue} = \{0.5, 0.9\}$. When we conduct $RMSC(c)$ without any test for detecting weak identification, we observe the improvement in the performance of post–selection inferences when $\mathbf{ID}_\pi = \mathbf{I}$ and degradation when $\mathbf{ID}_\pi = \mathbf{II}$, which agree with the former results. Unlike the case of $q_{max} = 8$, however, we also observe deterioration in the results when $\mathbf{ID}_\pi = \mathbf{III}$. These results are reported in Table 5.12. Next we conduct Stock–Yogo tests prior to $RMSC(c)$ and the results are in Tables 5.13 and 5.14. This sequential strategy of Stock–Yogo pretest and $RMSC(c)$ seems to help. We now observe improvements in the quality of post-selection inference when $\mathbf{ID}_\pi = \mathbf{III}$. One other noticeable fact in this case is that $RMSC(c)$ tends to pick at most four instruments among sixteen candidates when Stock–Yogo pretest is conducted. On the other hand, we do not have any improvements when $\mathbf{ID}_\pi = \mathbf{II}$. Through out all the cases with $q_{max} = 16$, some size distortions of Anderson–Rubin test are detected. These distortions are associated with lower values of concentration parameter and higher degrees of endogeneity of the regressor.

Figures 5.1–5.3 provide the plots of kernel estimates of the parameter estimates, $\theta_T(\hat{c}_T)$ and $\theta_T(\check{c}_{T,b})$, as well as the associated t-statistics. In order to save space, only three representative cases from each of $\mathbf{ID}_\pi = \mathbf{I}$, $\mathbf{II}$ and $\mathbf{III}$ are plotted. The detailed specifications of these three plots are provided in the captions of the figures. For the kernel density estimates of the parameter estimates, the reference distribution is not so obvious because the parameter estimate of each Mente Carlo draw has different limiting distribution. On the other hand, the associated t-statistic in each replications of simulation has standard normal distribution as its benchmark. For $\mathbf{ID}_\pi = \mathbf{I}$, it seems that even the $RMSC(c)$ without any pretest helps to reduce the bias (see Figure 5.1(a)). The advantage of $RMSC(c)$ can also be seen in Figure 5.1(b), kernel estimates for t-statistics with $RMSC(c)$ is closer to the reference density, $N(0,1)$. For $\mathbf{ID}_\pi = \mathbf{III}$, we have similar evidence for the usefulness of $RMSC(c)$, but in this case we need $q_{max} = 16$ and Stock–Yogo pretest for noticeable improvements (see Figure 5.2(a) and 5.2(b)). For $\mathbf{ID}_\pi = \mathbf{II}$, however, it does not seem that our moment selection procedure provides any advantages (see Figure 5.3(a) and 5.3(b)).

Table 5.1: Median Bias and Empirical Coverage Rates: $T = 100$ and all 8 Z's

T	$\frac{\lambda'\lambda}{q}$	σ_{ue}	$\mathbf{ID}_\pi$[a]	*Med Bias*	**Wald**	*width*	**A-R**	*width*
100	1	0.1	I	0.046	0.967	1.023	0.934	∞
100	1	0.1	II	0.047	0.966	1.021	0.938	∞
100	1	0.1	III	0.053	0.966	1.020	0.940	∞
100	1	0.5	I	0.239	0.785	0.928	0.938	∞
100	1	0.5	II	0.243	0.778	0.921	0.934	∞
100	1	0.5	III	0.244	0.783	0.924	0.940	∞
100	1	0.9	I	0.436	0.340	0.651	0.934	∞
100	1	0.9	II	0.431	0.343	0.653	0.937	∞
100	1	0.9	III	0.437	0.341	0.649	0.934	∞
100	10	0.1	I	0.008	0.948	0.420	0.934	0.708
100	10	0.1	II	0.009	0.949	0.420	0.936	0.708
100	10	0.1	III	0.008	0.950	0.420	0.935	0.708
100	10	0.5	I	0.040	0.919	0.411	0.938	0.719
100	10	0.5	II	0.041	0.915	0.410	0.938	0.710
100	10	0.5	III	0.042	0.917	0.409	0.936	0.706
100	10	0.9	I	0.071	0.854	0.393	0.937	0.732
100	10	0.9	II	0.073	0.850	0.392	0.937	0.730
100	10	0.9	III	0.074	0.844	0.390	0.936	0.723
100	25	0.1	I	0.003	0.951	0.272	0.937	0.431
100	25	0.1	II	0.004	0.949	0.272	0.936	0.431
100	25	0.1	III	0.004	0.949	0.271	0.936	0.432
100	25	0.5	I	0.017	0.934	0.269	0.937	0.430
100	25	0.5	II	0.017	0.934	0.269	0.936	0.431
100	25	0.5	III	0.018	0.936	0.268	0.936	0.432
100	25	0.9	I	0.032	0.896	0.263	0.936	0.433
100	25	0.9	II	0.032	0.906	0.263	0.936	0.434
100	25	0.9	III	0.031	0.901	0.263	0.939	0.434
100	50	0.1	I	0.001	0.944	0.193	0.934	0.301
100	50	0.1	II	0.001	0.945	0.194	0.939	0.303
100	50	0.1	III	0.003	0.950	0.194	0.936	0.302
100	50	0.5	I	0.009	0.937	0.193	0.940	0.301
100	50	0.5	II	0.010	0.940	0.192	0.934	0.300
100	50	0.5	III	0.009	0.939	0.192	0.941	0.303
100	50	0.9	I	0.016	0.923	0.190	0.936	0.303
100	50	0.9	II	0.015	0.928	0.191	0.936	0.304
100	50	0.9	III	0.015	0.926	0.191	0.937	0.302

[a] $\mathbf{ID}_\pi$ denotes the specification how the first stage parameter vector is constructed. 'I' means only the first instrument is relevant, 'II' means all the instruments are equally important, and 'III' means the instruments are in order of their importance.

Table 5.2: *RMSC(c)*: $T = 100$, BIC penalty and sequence of 8 Z's

	$\frac{\lambda'\lambda}{q}=1$			$\frac{\lambda'\lambda}{q}=10$			$\frac{\lambda'\lambda}{q}=25$			$\frac{\lambda'\lambda}{q}=50$		

Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]

ID($\mathcal{Z}$)[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.664	0.566	0.331	0.999	0.999	0.989	1.000	1.000	1.000	1.000	1.000	1.000
2	0.135	0.161	0.142	0.001	0.001	0.010	0.000	0.000	0.000	0.000	0.000	0.000
3	0.076	0.092	0.116	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
4	0.048	0.064	0.111	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.031	0.047	0.094	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.022	0.031	0.075	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.013	0.021	0.070	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.010	0.018	0.061	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.028	0.148	0.325	-.000	0.000	0.001	-.000	-.000	0.002	-.000	-.001	0.000
Wald	0.984	0.914	0.592	0.953	0.950	0.943	0.949	0.945	0.948	0.946	0.946	0.941
width	1.293	1.169	0.815	0.438	0.437	0.434	0.276	0.277	0.275	0.195	0.195	0.195
A-R	0.953	0.931	0.875	0.949	0.949	0.937	0.948	0.942	0.948	0.947	0.946	0.941
width	2.027	1.975	2.045	0.452	0.453	0.455	0.280	0.282	0.282	0.197	0.197	0.198

Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]

ID($\mathcal{Z}$)[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.069	0.076	0.089	0.026	0.033	0.065	0.014	0.018	0.031	0.011	0.011	0.018
2	0.133	0.131	0.150	0.109	0.123	0.153	0.099	0.105	0.131	0.089	0.093	0.105
3	0.167	0.168	0.163	0.202	0.207	0.191	0.209	0.221	0.211	0.219	0.228	0.218
4	0.170	0.169	0.160	0.244	0.223	0.192	0.272	0.259	0.231	0.300	0.278	0.259
5	0.149	0.154	0.137	0.203	0.183	0.163	0.222	0.211	0.189	0.234	0.230	0.218
6	0.129	0.121	0.118	0.124	0.130	0.116	0.129	0.124	0.121	0.114	0.115	0.121
7	0.096	0.099	0.097	0.065	0.070	0.074	0.045	0.049	0.058	0.030	0.038	0.049
8	0.086	0.082	0.087	0.027	0.030	0.048	0.010	0.014	0.028	0.004	0.007	0.013

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.059	0.283	0.504	0.016	0.076	0.129	0.010	0.038	0.066	0.003	0.021	0.037
Wald	0.981	0.827	0.331	0.966	0.907	0.785	0.959	0.932	0.864	0.956	0.940	0.898
width	1.242	1.119	0.761	0.539	0.525	0.487	0.354	0.350	0.338	0.256	0.254	0.247
A-R	0.947	0.930	0.887	0.948	0.939	0.923	0.944	0.939	0.928	0.940	0.941	0.932
width	7.301	7.486	10.527	0.819	0.806	0.783	0.510	0.507	0.495	0.362	0.359	0.353

Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]

ID($\mathcal{Z}$)[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.297	0.296	0.222	0.310	0.342	0.376	0.268	0.300	0.345	0.221	0.248	0.301
2	0.356	0.315	0.209	0.651	0.595	0.506	0.726	0.680	0.613	0.778	0.749	0.683
3	0.176	0.168	0.148	0.039	0.061	0.109	0.006	0.019	0.042	0.001	0.003	0.016
4	0.079	0.092	0.117	0.000	0.002	0.008	0.000	0.000	0.000	0.000	0.000	0.000
5	0.043	0.051	0.097	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
6	0.022	0.038	0.079	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.016	0.022	0.069	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.011	0.017	0.060	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.036	0.184	0.347	0.007	0.024	0.038	0.002	0.010	0.018	0.001	0.006	0.009
Wald	0.984	0.901	0.563	0.956	0.945	0.917	0.949	0.943	0.933	0.949	0.942	0.937
width	1.263	1.150	0.807	0.466	0.459	0.447	0.295	0.295	0.291	0.208	0.209	0.208
A-R	0.949	0.934	0.884	0.948	0.946	0.941	0.946	0.946	0.942	0.943	0.947	0.945
width	2.422	2.376	2.361	0.539	0.534	0.528	0.336	0.337	0.333	0.237	0.238	0.237

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $\mathrm{ID}_Z{=}q$ corresponds to the moment selection vector $c = (\iota_q' 0_{(q_{max}-q)}')'$, for $q = 1, \ldots, q_{max}$.

Table 5.3: *RMSC(c)*: $b = 0.05$, $T = 100$, BIC penalty and sequence of 8 Z's

$b = 0.05$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		

Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]

ID($\mathcal{Z}$)[b]\\σ_{ue}	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.011	0.013	0.013	0.840	0.841	0.840	1.000	1.000	1.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.989	0.987	0.987	0.160	0.159	0.160	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.050	0.243	0.432	0.008	0.040	0.074	-.002	0.003	0.005	-.001	-.000	-.000
Wald	0.969	0.787	0.345	0.949	0.921	0.847	0.950	0.949	0.951	0.944	0.945	0.943
width	1.024	0.924	0.656	0.421	0.412	0.391	0.276	0.275	0.275	0.195	0.195	0.195
A-R	0.936	0.939	0.938	0.932	0.933	0.937	0.948	0.946	0.957	0.945	0.944	0.945
width	∞	∞	∞	0.703	0.708	0.728	0.280	0.280	0.281	0.197	0.197	0.198

Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]

ID($\mathcal{Z}$)[b]\\σ_{ue}	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.017	0.025	0.010	0.011	0.015
2	0.000	0.000	0.000	0.001	0.001	0.003	0.080	0.088	0.105	0.086	0.093	0.110
3	0.000	0.000	0.000	0.003	0.003	0.002	0.186	0.185	0.189	0.210	0.220	0.217
4	0.000	0.000	0.000	0.002	0.003	0.003	0.237	0.218	0.199	0.292	0.281	0.261
5	0.000	0.000	0.000	0.002	0.002	0.002	0.192	0.183	0.163	0.244	0.237	0.216
6	0.000	0.000	0.000	0.001	0.001	0.001	0.096	0.102	0.097	0.122	0.115	0.122
7	0.000	0.000	0.000	0.001	0.000	0.001	0.032	0.037	0.043	0.033	0.036	0.046
8	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.008	0.020	0.004	0.008	0.014
weak[c]	1.000	1.000	1.000	0.991	0.990	0.988	0.163	0.164	0.159	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.056	0.237	0.436	0.008	0.039	0.075	0.008	0.032	0.059	0.003	0.021	0.036
Wald	0.967	0.794	0.343	0.952	0.915	0.843	0.959	0.933	0.856	0.959	0.940	0.904
width	1.025	0.931	0.652	0.421	0.413	0.391	0.338	0.335	0.325	0.255	0.253	0.248
A-R	0.935	0.940	0.936	0.942	0.941	0.937	0.943	0.941	0.926	0.943	0.941	0.935
width	∞	∞	∞	0.706	0.713	0.724	0.499	0.498	0.490	0.362	0.359	0.355

Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]

ID($\mathcal{Z}$)[b]\\σ_{ue}	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.005	0.005	0.005	0.225	0.262	0.290	0.223	0.254	0.307
2	0.000	0.000	0.000	0.007	0.006	0.005	0.609	0.570	0.511	0.776	0.742	0.677
3	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.012	0.029	0.001	0.005	0.016
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.989	0.988	0.989	0.164	0.157	0.170	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.050	0.242	0.434	0.006	0.040	0.074	0.003	0.012	0.018	0.001	0.004	0.010
Wald	0.969	0.787	0.340	0.949	0.919	0.844	0.951	0.945	0.929	0.948	0.944	0.940
width	1.020	0.928	0.650	0.420	0.413	0.391	0.292	0.292	0.291	0.208	0.209	0.208
A-R	0.938	0.937	0.941	0.935	0.937	0.942	0.947	0.949	0.941	0.946	0.947	0.946
width	∞	∞	∞	0.710	0.714	0.727	0.338	0.337	0.335	0.237	0.237	0.238

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $\text{ID}_Z = q$ corresponds to the moment selection vector $c = (\iota_q' 0_{(q_{max}-q)}')'$, for $q = 1, \ldots, q_{max}$.

[c] Stock and Yogo test based on 2SLS bias ($b = 0.05$) can not reject H_0 that instruments are weak. $c = \iota_8$.

Table 5.4: *RMSC(c)*: $r = 0.10$, $T = 100$, BIC penalty and sequence of 8 Z's

$r = 0.10$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		

Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]

ID($\mathcal{Z}$)[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.132	0.133	0.130	0.962	0.962	0.959
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.869	0.867	0.870	0.038	0.038	0.042

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.047	0.244	0.432	0.009	0.041	0.075	0.002	0.014	0.028	0.001	0.001	0.000
Wald	0.966	0.782	0.337	0.954	0.921	0.847	0.948	0.937	0.919	0.947	0.948	0.949
width	1.024	0.928	0.651	0.420	0.411	0.392	0.272	0.270	0.264	0.195	0.195	0.195
A-R	0.941	0.937	0.934	0.933	0.936	0.938	0.939	0.937	0.941	0.947	0.950	0.953
width	∞	∞	∞	0.710	0.712	0.725	0.421	0.424	0.434	0.197	0.197	0.198

Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]

ID($\mathcal{Z}$)[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.004	0.008	0.009	0.017
2	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.015	0.016	0.084	0.086	0.100
3	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.034	0.035	0.211	0.217	0.208
4	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.037	0.033	0.288	0.276	0.261
5	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.029	0.025	0.223	0.226	0.206
6	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.012	0.013	0.110	0.108	0.120
7	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.006	0.031	0.037	0.042
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005	0.006	0.011
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.861	0.868	0.866	0.038	0.035	0.035

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.043	0.243	0.434	0.008	0.042	0.073	0.006	0.018	0.033	0.004	0.020	0.035
Wald	0.966	0.784	0.342	0.950	0.919	0.850	0.946	0.933	0.902	0.958	0.936	0.901
width	1.026	0.933	0.649	0.420	0.410	0.392	0.280	0.278	0.272	0.252	0.250	0.246
A-R	0.942	0.936	0.937	0.938	0.939	0.934	0.936	0.937	0.935	0.938	0.938	0.930
width	∞	∞	∞	0.709	0.713	0.723	0.444	0.443	0.442	0.358	0.359	0.353

Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]

ID($\mathcal{Z}$)[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.047	0.054	0.209	0.249	0.298
2	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.083	0.078	0.749	0.712	0.653
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.001	0.003	0.012
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.863	0.870	0.866	0.042	0.037	0.037

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.050	0.243	0.435	0.008	0.042	0.074	0.003	0.016	0.029	0.002	0.005	0.011
Wald	0.966	0.774	0.343	0.949	0.921	0.849	0.948	0.933	0.916	0.952	0.947	0.940
width	1.030	0.926	0.653	0.420	0.411	0.391	0.274	0.272	0.266	0.208	0.208	0.208
A-R	0.937	0.935	0.941	0.934	0.936	0.935	0.939	0.937	0.935	0.952	0.951	0.947
width	∞	∞	∞	0.709	0.712	0.728	0.420	0.425	0.434	0.238	0.238	0.236

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $\mathrm{ID}_Z=q$ corresponds to the moment selection vector $c = (\iota_q' 0_{(q_{max}-q)}')'$, for $q = 1, \ldots, q_{max}$.

[c] Stock and Yogo test based on 2SLS size ($r = 0.10$) can not reject H_0 that instruments are weak. $c = \iota_8$.

Table 5.5: *RMSC(c)*: $T = 100$, AIC penalty and sequence of 8 Z's

	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		
Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_{\pi} = \mathbf{I}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.609	0.511	0.277	0.999	0.997	0.980	1.000	1.000	1.000	1.000	1.000	1.000
2	0.140	0.163	0.131	0.001	0.003	0.016	0.000	0.000	0.000	0.000	0.000	0.000
3	0.086	0.098	0.114	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000
4	0.059	0.073	0.113	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
5	0.040	0.057	0.106	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.030	0.042	0.089	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.020	0.031	0.087	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.016	0.026	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.033	0.164	0.348	-.000	0.000	0.002	-.000	-.000	0.002	-.000	-.001	0.000
Wald	0.984	0.905	0.549	0.953	0.950	0.942	0.949	0.945	0.948	0.946	0.946	0.941
width	1.266	1.138	0.785	0.438	0.437	0.434	0.276	0.277	0.275	0.195	0.195	0.195
A-R	0.952	0.929	0.870	0.949	0.948	0.932	0.948	0.942	0.948	0.947	0.946	0.941
width	2.037	1.992	2.140	0.452	0.453	0.454	0.280	0.282	0.282	0.197	0.197	0.198
Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_{\pi} = \mathbf{II}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.052	0.056	0.070	0.013	0.019	0.043	0.007	0.009	0.017	0.004	0.005	0.007
2	0.106	0.107	0.127	0.070	0.082	0.118	0.057	0.061	0.086	0.045	0.047	0.062
3	0.145	0.147	0.149	0.156	0.164	0.164	0.148	0.161	0.164	0.143	0.158	0.157
4	0.162	0.162	0.154	0.216	0.203	0.179	0.233	0.228	0.217	0.253	0.248	0.228
5	0.156	0.156	0.141	0.212	0.197	0.171	0.242	0.236	0.202	0.270	0.260	0.239
6	0.143	0.138	0.131	0.168	0.159	0.139	0.183	0.172	0.156	0.187	0.173	0.172
7	0.119	0.123	0.114	0.105	0.111	0.104	0.096	0.093	0.097	0.078	0.081	0.095
8	0.117	0.111	0.114	0.060	0.064	0.081	0.034	0.041	0.062	0.018	0.028	0.039
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.058	0.282	0.502	0.015	0.076	0.128	0.010	0.037	0.064	0.003	0.022	0.037
Wald	0.980	0.817	0.315	0.964	0.900	0.776	0.957	0.928	0.857	0.955	0.937	0.893
width	1.194	1.078	0.736	0.511	0.498	0.464	0.335	0.331	0.320	0.241	0.239	0.233
A-R	0.946	0.931	0.886	0.946	0.936	0.924	0.941	0.938	0.926	0.940	0.936	0.930
width	6.496	6.466	8.793	0.788	0.777	0.757	0.490	0.488	0.477	0.348	0.345	0.340
Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_{\pi} = \mathbf{III}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.254	0.255	0.188	0.230	0.269	0.320	0.171	0.207	0.272	0.120	0.148	0.209
2	0.342	0.301	0.184	0.702	0.637	0.524	0.812	0.755	0.656	0.877	0.842	0.759
3	0.187	0.173	0.145	0.067	0.090	0.140	0.017	0.037	0.071	0.004	0.010	0.032
4	0.090	0.102	0.119	0.001	0.005	0.014	0.000	0.000	0.001	0.000	0.000	0.000
5	0.055	0.060	0.105	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
6	0.031	0.049	0.090	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
7	0.024	0.033	0.086	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.018	0.027	0.084	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.040	0.194	0.363	0.007	0.024	0.039	0.001	0.009	0.018	0.001	0.005	0.008
Wald	0.983	0.892	0.526	0.956	0.946	0.916	0.948	0.942	0.932	0.950	0.943	0.938
width	1.235	1.121	0.780	0.461	0.453	0.442	0.292	0.291	0.288	0.206	0.207	0.206
A-R	0.949	0.931	0.879	0.947	0.947	0.940	0.945	0.946	0.940	0.942	0.947	0.947
width	2.428	2.386	2.413	0.540	0.537	0.528	0.338	0.339	0.334	0.238	0.239	0.238

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $\mathrm{ID}_Z=q$ corresponds to the moment selection vector $c = (\iota_q' 0_{q_{max}-q}')'$, for $q = 1, \ldots, q_{max}$.

Table 5.6: *RMSC(c)*: $b = 0.05$, $T = 100$, AIC penalty and sequence of 8 Z's

$b = 0.05$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		

Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.011	0.013	0.013	0.840	0.841	0.840	1.000	1.000	1.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.989	0.987	0.987	0.160	0.159	0.160	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.050	0.243	0.432	0.008	0.040	0.074	-.002	0.003	0.005	-.001	-.000	-.000
Wald	0.969	0.787	0.345	0.949	0.921	0.847	0.950	0.949	0.951	0.944	0.945	0.943
width	1.024	0.924	0.656	0.421	0.412	0.391	0.276	0.275	0.275	0.195	0.195	0.195
A-R	0.936	0.939	0.938	0.932	0.933	0.937	0.948	0.946	0.957	0.945	0.944	0.945
width	∞	∞	∞	0.703	0.708	0.728	0.280	0.280	0.281	0.197	0.197	0.198

Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.001	0.005	0.007	0.014	0.005	0.005	0.006
2	0.000	0.000	0.000	0.001	0.001	0.002	0.047	0.052	0.069	0.046	0.050	0.063
3	0.000	0.000	0.000	0.003	0.002	0.002	0.133	0.132	0.146	0.145	0.151	0.163
4	0.000	0.000	0.000	0.002	0.002	0.003	0.206	0.196	0.180	0.243	0.244	0.233
5	0.000	0.000	0.000	0.002	0.002	0.002	0.213	0.203	0.181	0.269	0.266	0.234
6	0.000	0.000	0.000	0.001	0.002	0.001	0.143	0.147	0.129	0.188	0.177	0.176
7	0.000	0.000	0.000	0.001	0.001	0.001	0.071	0.070	0.077	0.083	0.083	0.087
8	0.000	0.000	0.000	0.000	0.000	0.001	0.020	0.028	0.045	0.021	0.025	0.039
weak[c]	1.000	1.000	1.000	0.991	0.990	0.988	0.163	0.164	0.159	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.056	0.237	0.436	0.009	0.039	0.075	0.008	0.033	0.058	0.004	0.020	0.036
Wald	0.967	0.794	0.343	0.952	0.914	0.843	0.958	0.929	0.849	0.958	0.937	0.899
width	1.025	0.931	0.652	0.420	0.413	0.391	0.324	0.321	0.312	0.240	0.239	0.234
A-R	0.935	0.940	0.936	0.942	0.941	0.937	0.941	0.938	0.924	0.941	0.941	0.935
width	∞	∞	∞	0.706	0.712	0.723	0.483	0.482	0.474	0.347	0.346	0.342

Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.003	0.004	0.004	0.144	0.180	0.226	0.116	0.156	0.211
2	0.000	0.000	0.000	0.008	0.007	0.006	0.683	0.638	0.553	0.880	0.834	0.754
3	0.000	0.000	0.000	0.000	0.001	0.001	0.009	0.025	0.051	0.004	0.010	0.034
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.989	0.988	0.989	0.164	0.157	0.170	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.050	0.242	0.434	0.006	0.040	0.074	0.003	0.011	0.018	0.001	0.003	0.009
Wald	0.969	0.787	0.340	0.949	0.919	0.843	0.950	0.946	0.929	0.948	0.943	0.939
width	1.020	0.928	0.650	0.420	0.413	0.391	0.289	0.289	0.288	0.206	0.206	0.206
A-R	0.938	0.937	0.941	0.935	0.937	0.942	0.946	0.949	0.941	0.947	0.947	0.947
width	∞	∞	∞	0.710	0.714	0.727	0.340	0.339	0.336	0.238	0.239	0.238

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $\text{ID}_Z{=}q$ corresponds to the moment selection vector $c = (\iota_q' 0_{q_{max}-q}')'$, for $q = 1, \ldots, q_{max}$.

[c] Stock and Yogo test based on 2SLS bias ($b = 0.05$) can not reject H_0 that instruments are weak. $c = \iota_8$.

Table 5.7: *RMSC(c)*: $r = 0.10$, $T = 100$, AIC penalty and sequence of 8 Z's

$r = 0.10$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		

Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.129	0.136	0.137	0.960	0.958	0.962
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.871	0.864	0.864	0.040	0.042	0.038

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.049	0.242	0.435	0.010	0.039	0.076	0.003	0.015	0.028	0.001	-.000	0.001
Wald	0.968	0.780	0.335	0.951	0.921	0.845	0.948	0.942	0.919	0.946	0.947	0.946
width	1.029	0.925	0.645	0.420	0.410	0.390	0.272	0.269	0.264	0.195	0.195	0.195
A-R	0.941	0.931	0.934	0.936	0.939	0.939	0.937	0.936	0.941	0.947	0.949	0.952
width	∞	∞	∞	0.707	0.711	0.718	0.419	0.419	0.430	0.197	0.197	0.197

Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.004	0.008
2	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.009	0.011	0.048	0.048	0.061
3	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.025	0.024	0.145	0.151	0.157
4	0.000	0.000	0.000	0.000	0.000	0.000	0.036	0.036	0.031	0.246	0.234	0.223
5	0.000	0.000	0.000	0.000	0.000	0.000	0.034	0.034	0.029	0.253	0.251	0.229
6	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.019	0.018	0.176	0.169	0.164
7	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.008	0.009	0.074	0.081	0.084
8	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.004	0.015	0.024	0.033
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.863	0.868	0.871	0.040	0.038	0.041

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.055	0.244	0.433	0.006	0.039	0.075	0.002	0.018	0.032	0.005	0.020	0.034
Wald	0.964	0.789	0.339	0.952	0.913	0.843	0.947	0.927	0.899	0.955	0.936	0.896
width	1.029	0.928	0.651	0.421	0.410	0.390	0.279	0.277	0.270	0.239	0.238	0.234
A-R	0.935	0.940	0.937	0.937	0.935	0.937	0.933	0.939	0.931	0.940	0.940	0.933
width	∞	∞	∞	0.705	0.712	0.724	0.440	0.440	0.439	0.346	0.345	0.343

Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.036	0.039	0.124	0.154	0.212
2	0.000	0.000	0.000	0.000	0.000	0.000	0.108	0.098	0.087	0.839	0.799	0.725
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.005	0.002	0.008	0.024
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.861	0.864	0.868	0.036	0.038	0.039

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.051	0.235	0.436	0.008	0.042	0.074	0.003	0.016	0.029	0.000	0.004	0.010
Wald	0.963	0.798	0.339	0.949	0.919	0.847	0.950	0.938	0.913	0.950	0.944	0.941
width	1.028	0.928	0.648	0.418	0.410	0.391	0.274	0.272	0.265	0.206	0.206	0.206
A-R	0.935	0.942	0.938	0.938	0.937	0.939	0.940	0.939	0.933	0.946	0.950	0.952
width	∞	∞	∞	0.709	0.709	0.729	0.420	0.424	0.433	0.238	0.238	0.238

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $\mathrm{ID}_Z = q$ corresponds to the moment selection vector $c = (\iota_q' 0_{q_{max}-q}')'$, for $q = 1, \ldots, q_{max}$.

[c] Stock and Yogo test based on 2SLS size ($r = 0.10$) can not reject H_0 that instruments are weak. $c = \iota_8$.

Table 5.8: *RMSC(c)*: $T = 100$, BIC penalty and all combination of 8 Z's

	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		
Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
$1R$	0.329	0.249	0.090	0.999	0.988	0.926	1.000	1.000	0.999	1.000	1.000	1.000
$1R/I*$	0.605	0.656	0.504	0.001	0.012	0.074	0.000	0.000	0.001	0.000	0.000	0.000
I	0.066	0.095	0.406	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
All	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.046	0.234	0.532	0.000	0.001	0.008	-.000	-.000	0.002	-.000	-.001	0.000
Wald	0.986	0.877	0.273	0.953	0.950	0.940	0.949	0.945	0.948	0.946	0.946	0.941
width	1.180	1.066	0.740	0.438	0.437	0.431	0.276	0.277	0.275	0.195	0.195	0.195
A-R	0.954	0.893	0.734	0.949	0.945	0.892	0.948	0.942	0.947	0.947	0.946	0.941
width	1.854	1.655	1.497	0.451	0.452	0.442	0.280	0.282	0.282	0.197	0.197	0.198
Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]												
ID_Z[c]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1Z	0.064	0.065	0.119	0.006	0.008	0.022	0.003	0.003	0.007	0.002	0.001	0.003
2Z	0.448	0.464	0.507	0.230	0.268	0.387	0.168	0.187	0.257	0.148	0.158	0.184
3Z	0.428	0.418	0.339	0.623	0.607	0.526	0.654	0.652	0.620	0.670	0.664	0.656
4Z	0.058	0.052	0.034	0.138	0.115	0.064	0.172	0.153	0.114	0.177	0.174	0.156
5Z	0.001	0.001	0.001	0.003	0.002	0.001	0.003	0.004	0.003	0.003	0.003	0.003
6Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.068	0.325	0.597	0.028	0.134	0.217	0.017	0.075	0.127	0.009	0.045	0.078
Wald	0.989	0.810	0.175	0.972	0.835	0.545	0.967	0.878	0.672	0.962	0.907	0.764
width	1.204	1.074	0.709	0.534	0.515	0.463	0.356	0.349	0.331	0.259	0.254	0.246
A-R	0.956	0.916	0.748	0.947	0.925	0.877	0.949	0.934	0.894	0.945	0.936	0.911
width	2.374	2.158	1.650	0.731	0.720	0.690	0.471	0.474	0.467	0.342	0.341	0.338
Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]												
ID_Z[d]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1S	0.123	0.111	0.052	0.291	0.306	0.294	0.265	0.291	0.322	0.220	0.246	0.292
2S	0.177	0.137	0.042	0.639	0.574	0.447	0.723	0.674	0.601	0.778	0.747	0.679
3S	0.036	0.029	0.007	0.024	0.034	0.046	0.004	0.011	0.024	0.001	0.002	0.010
4S	0.001	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
5S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Others	0.662	0.721	0.898	0.046	0.086	0.213	0.008	0.023	0.052	0.001	0.005	0.019
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.053	0.275	0.581	0.007	0.030	0.055	0.002	0.010	0.020	0.001	0.006	0.010
Wald	0.986	0.853	0.212	0.957	0.943	0.890	0.949	0.943	0.929	0.949	0.943	0.937
width	1.189	1.074	0.724	0.466	0.458	0.442	0.295	0.295	0.291	0.208	0.209	0.208
A-R	0.955	0.904	0.745	0.947	0.939	0.903	0.946	0.943	0.935	0.943	0.946	0.942
width	2.082	1.893	1.574	0.538	0.530	0.509	0.336	0.336	0.331	0.237	0.238	0.237

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $1R$ denotes $c = (1, 0_7')'$; $1R/I^*$ denotes $c = (1, a')'$, $a \neq 0_7$ or ι_7; I denotes $c = (0, b')'$, $b \neq 0_7$; and *All* denotes $c = \iota_8$.

[c] xZ denotes $|c| = x$, for $x = 1, \ldots, 8$.

[d] xS denotes $c = (\iota_x', 0_{8-x}')'$, for $x = 1, \ldots, 8$; *Others* denotes all the other c's.

Table 5.9: *RMSC(c)*: $b = 0.05$, $T = 100$, BIC penalty and all combination of 8 Z's

$b = 0.05$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		
Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
$1R$	0.000	0.000	0.000	0.011	0.013	0.013	0.840	0.841	0.840	1.000	1.000	1.000
$1R/I*$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
I	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
All	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.989	0.987	0.987	0.160	0.159	0.160	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.050	0.243	0.432	0.008	0.040	0.074	-.002	0.003	0.005	-.001	-.000	-.000
Wald	0.969	0.787	0.345	0.949	0.921	0.847	0.950	0.949	0.951	0.944	0.945	0.943
width	1.024	0.924	0.656	0.421	0.412	0.391	0.276	0.275	0.275	0.195	0.195	0.195
A-R	0.936	0.939	0.938	0.932	0.933	0.937	0.948	0.946	0.957	0.945	0.944	0.945
width	∞	∞	∞	0.703	0.708	0.728	0.280	0.280	0.281	0.197	0.197	0.198
Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]												
ID_Z[d]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1Z	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.005	0.001	0.002	0.002
2Z	0.000	0.000	0.000	0.002	0.003	0.004	0.140	0.160	0.207	0.149	0.155	0.174
3Z	0.000	0.000	0.000	0.007	0.007	0.008	0.558	0.549	0.534	0.665	0.666	0.672
4Z	0.000	0.000	0.000	0.001	0.000	0.000	0.136	0.123	0.093	0.181	0.174	0.149
5Z	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.002	0.003	0.004	0.003
6Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.991	0.990	0.988	0.163	0.164	0.159	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.056	0.237	0.436	0.009	0.039	0.075	0.014	0.064	0.114	0.009	0.044	0.077
Wald	0.967	0.794	0.343	0.952	0.913	0.841	0.967	0.888	0.686	0.966	0.908	0.776
width	1.025	0.931	0.652	0.421	0.413	0.391	0.342	0.336	0.321	0.259	0.255	0.247
A-R	0.935	0.940	0.936	0.942	0.940	0.934	0.943	0.932	0.890	0.946	0.940	0.919
width	∞	∞	∞	0.706	0.712	0.723	0.468	0.468	0.461	0.341	0.340	0.340
Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]												
ID_Z[e]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1S	0.000	0.000	0.000	0.005	0.005	0.004	0.222	0.255	0.269	0.223	0.252	0.300
2S	0.000	0.000	0.000	0.006	0.006	0.005	0.606	0.566	0.502	0.775	0.740	0.672
3S	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.007	0.017	0.001	0.003	0.010
4S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Others	0.000	0.000	0.000	0.000	0.001	0.002	0.006	0.016	0.042	0.001	0.005	0.017
weak[c]	1.000	1.000	1.000	0.989	0.988	0.989	0.164	0.157	0.170	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.050	0.242	0.434	0.006	0.040	0.074	0.003	0.012	0.020	0.001	0.004	0.010
Wald	0.969	0.787	0.340	0.949	0.919	0.843	0.950	0.944	0.926	0.948	0.943	0.939
width	1.020	0.928	0.650	0.420	0.413	0.391	0.292	0.292	0.290	0.208	0.209	0.208
A-R	0.938	0.937	0.941	0.935	0.937	0.941	0.946	0.947	0.934	0.946	0.947	0.944
width	∞	∞	∞	0.710	0.714	0.727	0.338	0.337	0.333	0.237	0.237	0.237

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $1R$ denotes $c = (1, 0_7')'$; $1R/I^*$ denotes $c = (1, a')'$, $a \neq 0_7$ or ι_7; I denotes $c = (0, b')'$, $b \neq 0_7$; and *All* denotes $c = \iota_8$.

[c] Stock and Yogo's pretest based on 2SLS bias can not reject the H_0 that the instruments are weak. $c = \iota_8$ is used.

[d] xZ denotes $|c| = x$, for $x = 1, \ldots, 8$.

[e] xS denotes $c = (\iota_x', 0_{8-x}')'$, for $x = 1, \ldots, 8$; *Others* denotes all the other c's.

Table 5.10: *RMSC(c)*: $r = 0.10$, $T = 100$, BIC penalty and all combination of 8 Z's

$r = 0.10$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		
Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]												
ID_Z[b] $\diagdown \sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
$1R$	0.000	0.000	0.000	0.000	0.000	0.000	0.129	0.136	0.137	0.960	0.958	0.962
$1R/I*$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
I	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
All	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.871	0.864	0.864	0.040	0.042	0.038
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.049	0.242	0.435	0.010	0.039	0.076	0.003	0.015	0.028	0.001	-.000	0.001
Wald	0.968	0.780	0.335	0.951	0.921	0.845	0.948	0.942	0.919	0.946	0.947	0.946
width	1.029	0.925	0.645	0.420	0.410	0.390	0.272	0.269	0.264	0.195	0.195	0.195
A-R	0.941	0.931	0.934	0.936	0.939	0.939	0.937	0.936	0.941	0.947	0.949	0.952
width	∞	∞	∞	0.707	0.711	0.718	0.419	0.419	0.430	0.197	0.197	0.197
Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]												
ID_Z[d] $\diagdown \sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.003
2Z	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.027	0.033	0.143	0.147	0.176
3Z	0.000	0.000	0.000	0.000	0.000	0.000	0.095	0.090	0.085	0.643	0.644	0.638
4Z	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.015	0.010	0.171	0.167	0.140
5Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.003	0.003
6Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8Z	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.863	0.868	0.871	0.040	0.038	0.041
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.055	0.244	0.433	0.006	0.039	0.075	0.004	0.022	0.034	0.009	0.044	0.073
Wald	0.964	0.789	0.339	0.952	0.913	0.843	0.947	0.919	0.866	0.962	0.906	0.771
width	1.029	0.928	0.651	0.421	0.410	0.390	0.281	0.278	0.271	0.256	0.252	0.246
A-R	0.935	0.940	0.937	0.937	0.935	0.937	0.932	0.933	0.918	0.944	0.940	0.914
width	∞	∞	∞	0.705	0.712	0.724	0.436	0.436	0.436	0.338	0.340	0.340
Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]												
ID_Z[e] $\diagdown \sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1S	0.000	0.000	0.000	0.000	0.000	0.000	0.047	0.049	0.048	0.219	0.249	0.294
2S	0.000	0.000	0.000	0.000	0.000	0.000	0.092	0.085	0.078	0.744	0.707	0.645
3S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.007
4S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8S	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Others	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.001	0.004	0.015
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.861	0.864	0.868	0.036	0.038	0.039
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.051	0.235	0.436	0.008	0.042	0.074	0.003	0.016	0.029	0.001	0.006	0.012
Wald	0.963	0.798	0.339	0.949	0.919	0.847	0.950	0.938	0.912	0.950	0.944	0.940
width	1.028	0.928	0.648	0.418	0.410	0.391	0.274	0.272	0.265	0.208	0.208	0.208
A-R	0.935	0.942	0.938	0.938	0.937	0.939	0.940	0.939	0.931	0.945	0.949	0.948
width	∞	∞	∞	0.709	0.709	0.729	0.420	0.424	0.433	0.237	0.237	0.236

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $1R$ denotes $c = (1, 0_7')'$; $1R/I^*$ denotes $c = (1, a')'$, $a \neq 0_7$ or ι_7; I denotes $c = (0, b')'$, $b \neq 0_7$; and *All* denotes $c = \iota_8$.

[c] Stock and Yogo's pretest based on 2SLS size can not reject the H_0 that the instruments are weak. $c = \iota_8$ is used.

[d] xZ denotes $|c| = x$, for $x = 1, \ldots, 8$.

[e] xS denotes $c = (\iota_x', 0_{8-x}')'$, for $x = 1, \ldots, 8$; *Others* denotes all the other c's.

Table 5.11: Median Bias and Empirical Coverage Rates: $T = 100$ and all 16 Z's

T	$\frac{\lambda'\lambda}{q}$	σ_{ue}	$\mathbf{ID}_{\pi}$ [a]	*Med Bias*	**Wald**	*width*	**A-R**	*width*
100	1	0.1	I	0.046	0.949	0.704	0.922	3.725
100	1	0.1	II	0.051	0.946	0.704	0.925	3.672
100	1	0.1	III	0.053	0.945	0.703	0.923	3.671
100	1	0.5	I	0.245	0.646	0.638	0.927	3.763
100	1	0.5	II	0.245	0.648	0.639	0.921	3.820
100	1	0.5	III	0.244	0.648	0.638	0.923	3.814
100	1	0.9	I	0.443	0.117	0.458	0.923	5.007
100	1	0.9	II	0.444	0.118	0.458	0.925	4.933
100	1	0.9	III	0.443	0.119	0.459	0.925	5.090
100	10	0.1	I	0.009	0.947	0.295	0.927	0.565
100	10	0.1	II	0.010	0.944	0.295	0.923	0.562
100	10	0.1	III	0.009	0.947	0.295	0.927	0.568
100	10	0.5	I	0.044	0.897	0.289	0.925	0.565
100	10	0.5	II	0.044	0.897	0.290	0.924	0.563
100	10	0.5	III	0.043	0.893	0.290	0.921	0.561
100	10	0.9	I	0.079	0.767	0.275	0.927	0.567
100	10	0.9	II	0.078	0.763	0.275	0.923	0.574
100	10	0.9	III	0.078	0.770	0.275	0.923	0.570
100	25	0.1	I	0.003	0.947	0.191	0.926	0.348
100	25	0.1	II	0.004	0.943	0.191	0.923	0.346
100	25	0.1	III	0.004	0.945	0.191	0.926	0.345
100	25	0.5	I	0.017	0.926	0.190	0.928	0.348
100	25	0.5	II	0.018	0.928	0.190	0.924	0.346
100	25	0.5	III	0.017	0.928	0.190	0.923	0.347
100	25	0.9	I	0.033	0.873	0.186	0.924	0.346
100	25	0.9	II	0.033	0.873	0.186	0.926	0.349
100	25	0.9	III	0.033	0.870	0.187	0.928	0.348
100	50	0.1	I	0.002	0.947	0.136	0.922	0.241
100	50	0.1	II	0.002	0.952	0.137	0.922	0.243
100	50	0.1	III	0.002	0.947	0.137	0.923	0.242
100	50	0.5	I	0.010	0.934	0.136	0.926	0.244
100	50	0.5	II	0.009	0.932	0.136	0.925	0.243
100	50	0.5	III	0.009	0.932	0.136	0.927	0.242
100	50	0.9	I	0.017	0.915	0.135	0.923	0.245
100	50	0.9	II	0.018	0.910	0.135	0.923	0.243
100	50	0.9	III	0.016	0.908	0.135	0.925	0.244

[a] $\mathbf{ID}_{\pi}$ denotes the specification how the first stage parameter vector is constructed. 'I' means only the first instrument is relevant, 'II' means all the instruments are equally important, and 'III' means the instruments are in order of their importance.

Table 5.12: *RMSC(c)*: $T = 100$ and sequence of 16 Z's

	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		

Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_{\pi} = \mathbf{I}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.882	0.791	0.583	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
2	0.070	0.101	0.140	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
3	0.026	0.048	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.012	0.026	0.059	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.005	0.014	0.044	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.003	0.009	0.034	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.001	0.005	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.001	0.002	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.001	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.001	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.010	0.052	0.138	0.000	-0.001	-0.001	-0.000	0.000	0.000	-0.000	-0.001	0.001
Wald	0.974	0.941	0.805	0.948	0.947	0.948	0.950	0.945	0.949	0.946	0.948	0.946
width	0.979	0.930	0.790	0.310	0.308	0.307	0.195	0.195	0.195	0.138	0.138	0.137
A-R	0.947	0.934	0.885	0.947	0.944	0.947	0.950	0.946	0.948	0.947	0.949	0.945
width	1.155	1.126	1.088	0.316	0.315	0.316	0.197	0.197	0.198	0.139	0.139	0.139

Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_{\pi} = \mathbf{II}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.068	0.069	0.089	0.041	0.043	0.059	0.028	0.032	0.041	0.028	0.027	0.029
2	0.126	0.122	0.132	0.120	0.127	0.144	0.112	0.117	0.128	0.109	0.118	0.117
3	0.153	0.160	0.155	0.193	0.198	0.189	0.199	0.208	0.198	0.204	0.205	0.209
4	0.166	0.161	0.157	0.214	0.211	0.194	0.238	0.224	0.215	0.239	0.231	0.230
5	0.152	0.140	0.130	0.183	0.172	0.158	0.196	0.192	0.181	0.198	0.197	0.193
6	0.118	0.120	0.109	0.128	0.124	0.112	0.126	0.123	0.119	0.132	0.124	0.123
7	0.088	0.088	0.084	0.070	0.067	0.069	0.066	0.067	0.068	0.060	0.066	0.064
8	0.057	0.058	0.054	0.035	0.037	0.041	0.025	0.027	0.031	0.022	0.022	0.022
9	0.034	0.037	0.037	0.011	0.014	0.021	0.008	0.009	0.013	0.007	0.007	0.009
10	0.019	0.023	0.025	0.004	0.005	0.007	0.002	0.003	0.004	0.001	0.001	0.003
11	0.010	0.013	0.013	0.001	0.001	0.004	0.000	0.000	0.001	0.000	0.000	0.001
12	0.005	0.006	0.008	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000
13	0.002	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates

Med Bias	0.052	0.280	0.496	0.013	0.065	0.112	0.006	0.032	0.054	0.003	0.016	0.028
Wald	0.983	0.825	0.309	0.972	0.925	0.821	0.966	0.945	0.894	0.964	0.954	0.928
width	1.176	1.054	0.739	0.515	0.503	0.471	0.337	0.332	0.322	0.241	0.240	0.236
A-R	0.945	0.931	0.888	0.947	0.938	0.926	0.943	0.942	0.939	0.943	0.940	0.943
width	7.855	7.389	8.663	0.843	0.827	0.805	0.520	0.516	0.507	0.368	0.366	0.362

(Table 5.12 continued)

	$\frac{\lambda'\lambda}{q}=1$			$\frac{\lambda'\lambda}{q}=10$			$\frac{\lambda'\lambda}{q}=25$			$\frac{\lambda'\lambda}{q}=50$		
Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]												
ID_Z[b] $\diagdown \sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.168	0.178	0.206	0.056	0.083	0.134	0.034	0.041	0.069	0.023	0.027	0.044
2	0.344	0.336	0.242	0.588	0.550	0.476	0.639	0.634	0.577	0.684	0.670	0.635
3	0.273	0.246	0.200	0.333	0.328	0.316	0.322	0.312	0.329	0.290	0.299	0.311
4	0.135	0.138	0.135	0.022	0.038	0.068	0.005	0.012	0.024	0.002	0.005	0.011
5	0.051	0.060	0.088	0.000	0.001	0.006	0.000	0.000	0.001	0.000	0.000	0.000
6	0.019	0.025	0.049	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.007	0.009	0.032	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.002	0.004	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.001	0.002	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.001	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.024	0.133	0.241	0.006	0.019	0.037	0.001	0.010	0.015	0.001	0.005	0.008
Wald	0.977	0.902	0.699	0.957	0.938	0.910	0.951	0.942	0.933	0.952	0.949	0.941
width	0.995	0.928	0.748	0.354	0.347	0.340	0.225	0.223	0.222	0.160	0.159	0.159
A-R	0.949	0.930	0.892	0.943	0.946	0.933	0.945	0.938	0.941	0.944	0.944	0.942
width	1.605	1.530	1.440	0.437	0.430	0.422	0.274	0.271	0.271	0.193	0.192	0.191

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] $\text{ID}_Z=q$ corresponds to the moment selection vector $c = (\iota_q' 0_{q_{max}-q}')'$, for $q = 1, \ldots, q_{max}$.

Table 5.13: *RMSC(c)*: $b = 0.05$, $T = 100$ and sequence of 16 Z's

$b = 0.05$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		
Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.003	0.004	0.004	0.804	0.811	0.806	1.000	1.000	1.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5 – 16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.997	0.996	0.996	0.196	0.189	0.194	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.047	0.247	0.441	0.009	0.042	0.078	0.001	0.004	0.007	0.000	0.001	-0.000
Wald	0.946	0.646	0.115	0.943	0.896	0.773	0.946	0.949	0.949	0.947	0.946	0.951
width	0.705	0.636	0.459	0.295	0.290	0.275	0.195	0.194	0.194	0.138	0.138	0.138
A-R	0.926	0.920	0.926	0.924	0.924	0.924	0.942	0.950	0.955	0.947	0.946	0.951
width	3.777	3.806	5.044	0.558	0.562	0.570	0.198	0.198	0.197	0.139	0.139	0.139
Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.026	0.029	0.026	0.028	0.028
2	0.000	0.000	0.000	0.000	0.001	0.000	0.093	0.098	0.108	0.115	0.117	0.118
3	0.000	0.000	0.000	0.001	0.001	0.001	0.166	0.161	0.164	0.205	0.203	0.213
4	0.000	0.000	0.000	0.001	0.001	0.001	0.185	0.190	0.178	0.231	0.228	0.235
5	0.000	0.000	0.000	0.000	0.001	0.001	0.157	0.149	0.138	0.208	0.198	0.183
6	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.101	0.096	0.127	0.128	0.127
7	0.000	0.000	0.000	0.000	0.000	0.001	0.054	0.048	0.051	0.060	0.067	0.063
8	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.021	0.026	0.021	0.025	0.026
9	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.004	0.009	0.006	0.007	0.006
10	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003	0.001	0.001	0.002
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000
12 – 16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.997	0.997	0.997	0.194	0.200	0.197	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.049	0.246	0.443	0.008	0.041	0.079	0.005	0.025	0.045	0.004	0.016	0.028
Wald	0.945	0.647	0.114	0.943	0.904	0.769	0.960	0.942	0.893	0.965	0.953	0.927
width	0.703	0.640	0.456	0.295	0.290	0.275	0.308	0.304	0.291	0.241	0.239	0.236
A-R	0.925	0.928	0.925	0.924	0.929	0.924	0.942	0.943	0.932	0.945	0.939	0.941
width	3.656	4.017	4.773	0.563	0.569	0.566	0.479	0.480	0.475	0.366	0.364	0.361
Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]												
ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.001	0.025	0.037	0.058	0.023	0.027	0.045
2	0.000	0.000	0.000	0.003	0.002	0.001	0.533	0.518	0.480	0.681	0.670	0.633
3	0.000	0.000	0.000	0.001	0.001	0.001	0.238	0.247	0.254	0.294	0.300	0.312
4	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.007	0.016	0.002	0.004	0.010
5 – 16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	0.996	0.997	0.998	0.202	0.190	0.192	0.000	0.000	0.000
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.057	0.244	0.444	0.007	0.042	0.077	0.002	0.010	0.018	0.000	0.005	0.008
Wald	0.946	0.652	0.114	0.946	0.893	0.766	0.951	0.945	0.931	0.951	0.949	0.941
width	0.704	0.640	0.459	0.295	0.289	0.276	0.218	0.217	0.215	0.160	0.160	0.158
A-R	0.921	0.929	0.926	0.928	0.929	0.928	0.942	0.943	0.943	0.948	0.944	0.946
width	3.636	3.977	5.135	0.563	0.565	0.572	0.275	0.274	0.272	0.193	0.193	0.192

[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] ID_Z=q corresponds to the moment selection vector $c = (\iota_q' 0_{q_{max}-q}')'$, for $q = 1, \ldots, q_{max}$.

[c] Stock and Yogo test based on 2SLS bias ($b = 0.05$) can not reject H_0 that instruments are weak. $c = \iota_{16}$.

Table 5.14: *RMSC(c)*: $r = 0.10$, $T = 100$ and sequence of 16 Z's

$r = 0.10$	$\frac{\lambda'\lambda}{q} = 1$			$\frac{\lambda'\lambda}{q} = 10$			$\frac{\lambda'\lambda}{q} = 25$			$\frac{\lambda'\lambda}{q} = 50$		

Part I. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{I}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.446	0.440	0.448
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5 – 16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	1.000	0.554	0.560	0.552
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.047	0.245	0.443	0.009	0.042	0.077	0.004	0.018	0.033	0.001	0.005	0.010
Wald	0.946	0.649	0.114	0.942	0.894	0.766	0.948	0.924	0.872	0.945	0.941	0.931
width	0.704	0.641	0.458	0.295	0.289	0.276	0.192	0.190	0.186	0.137	0.137	0.136
A-R	0.927	0.923	0.927	0.920	0.922	0.927	0.925	0.928	0.923	0.937	0.941	0.944
width	3.712	3.900	4.796	0.562	0.562	0.569	0.345	0.346	0.346	0.153	0.157	0.157

Part II. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{II}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.012	0.013
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.054	0.053	0.058
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.095	0.088
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.106	0.106	0.104
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.090	0.091	0.087
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053	0.050	0.052
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.023	0.025
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.008	0.011
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.003
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12 – 16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.560	0.560	0.558
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.047	0.248	0.443	0.009	0.044	0.079	0.003	0.016	0.034	0.002	0.011	0.019
Wald	0.948	0.642	0.117	0.949	0.895	0.767	0.948	0.928	0.870	0.954	0.944	0.923
width	0.703	0.637	0.459	0.296	0.289	0.276	0.192	0.190	0.186	0.159	0.159	0.159
A-R	0.925	0.924	0.928	0.925	0.924	0.926	0.926	0.921	0.921	0.936	0.930	0.939
width	3.596	3.803	5.059	0.565	0.565	0.565	0.347	0.346	0.345	0.287	0.289	0.292

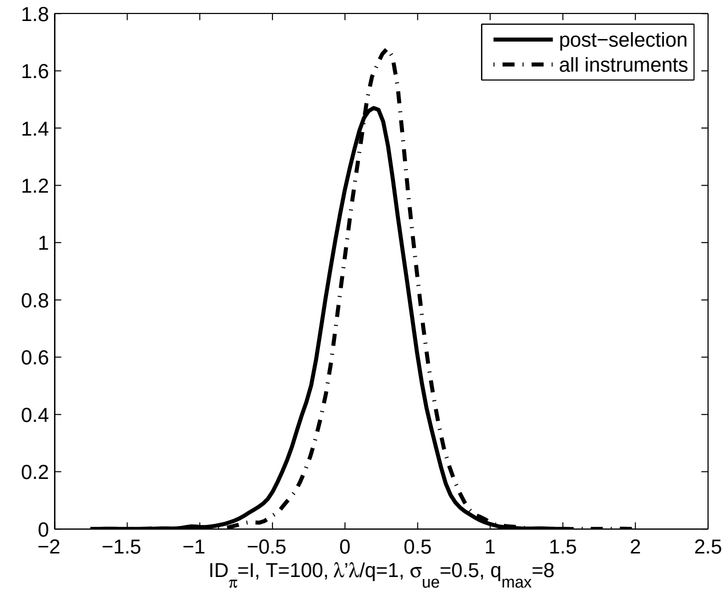
Part III. *Empirical Selection Probabilities:* $\mathbf{ID}_\pi = \mathbf{III}$[a]

ID_Z[b]$\diagdown\sigma_{ue}$	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*	*0.1*	*0.5*	*0.9*
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.014	0.022
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.319	0.312	0.300
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.108	0.109	0.113
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002
5 – 16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
weak[c]	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.561	0.565	0.563
Sampling Properties of Post-Selection Estimator: Median Bias & Empirical Coverage Rates												
Med Bias	0.048	0.248	0.443	0.009	0.043	0.078	0.004	0.019	0.033	0.002	0.007	0.012
Wald	0.945	0.644	0.120	0.945	0.890	0.765	0.947	0.924	0.875	0.946	0.943	0.931
width	0.706	0.639	0.458	0.296	0.290	0.276	0.191	0.190	0.186	0.146	0.146	0.144
A-R	0.921	0.924	0.927	0.926	0.929	0.921	0.925	0.923	0.925	0.933	0.940	0.942
width	3.753	3.825	4.920	0.562	0.566	0.569	0.346	0.348	0.347	0.204	0.205	0.204

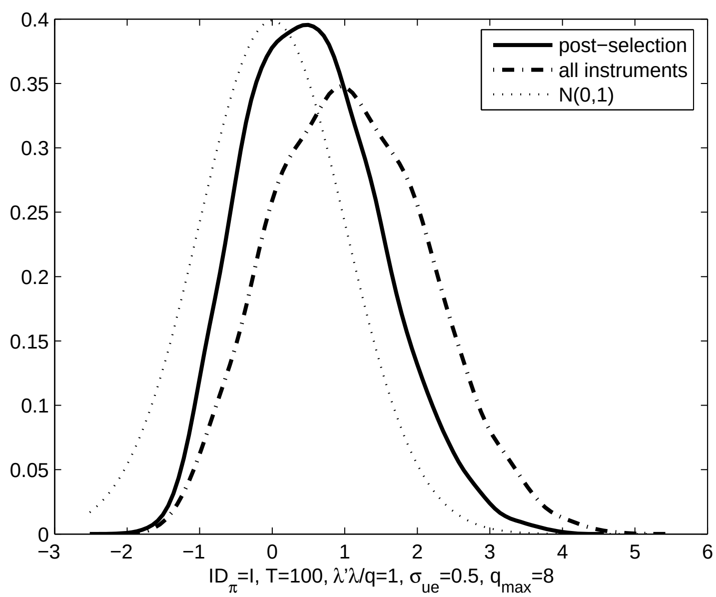
[a] I denotes only the first instrument is relevant; II denotes all instruments are equally important; and III denotes instruments are in order of their importance.

[b] ID_Z=q corresponds to the moment selection vector $c = (\iota_q' 0_{q_{max}-q}')'$, for $q = 1, \ldots, q_{max}$.

[c] Stock and Yogo test based on 2SLS size ($r = 0.10$) can not reject H_0 that instruments are weak. $c = \iota_{16}$.

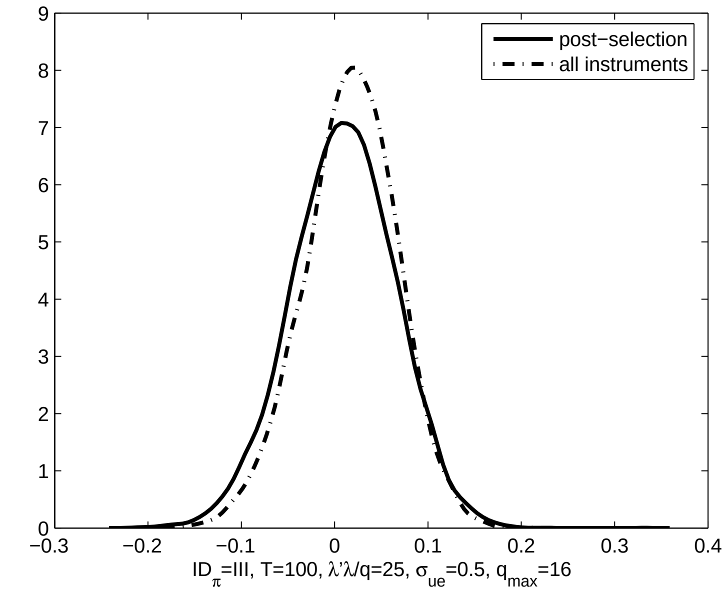


(a) Kernel Estimate of $\hat{\theta}_T(\iota_8)$ & $\hat{\theta}_T(\hat{c}_T)$

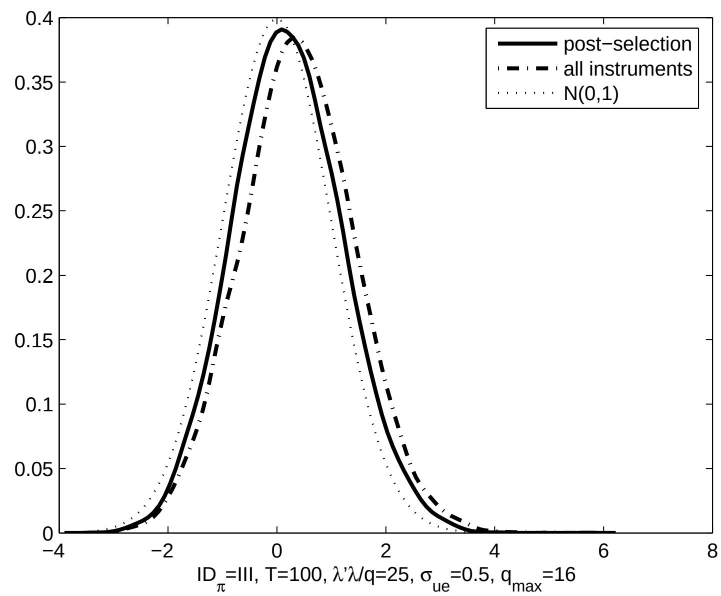


(b) Kernel Estimate of t-statistics

Figure 5.1: Kernel Estimate: All Z's vs $RMSC(c)$ for $\mathbf{ID}_\pi = \mathbf{I}$

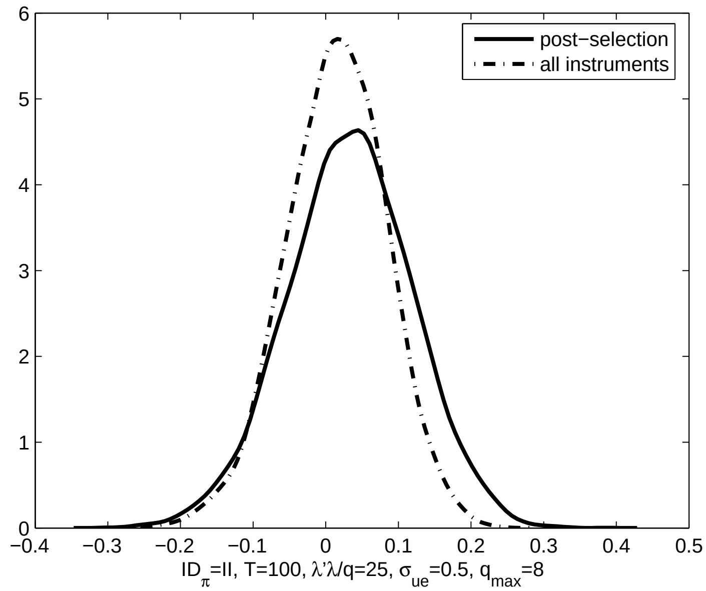


(a) Kernel Estimate of $\hat{\theta}_T(\iota_{16})$ & $\hat{\theta}_T(\check{c}_{T,b})$

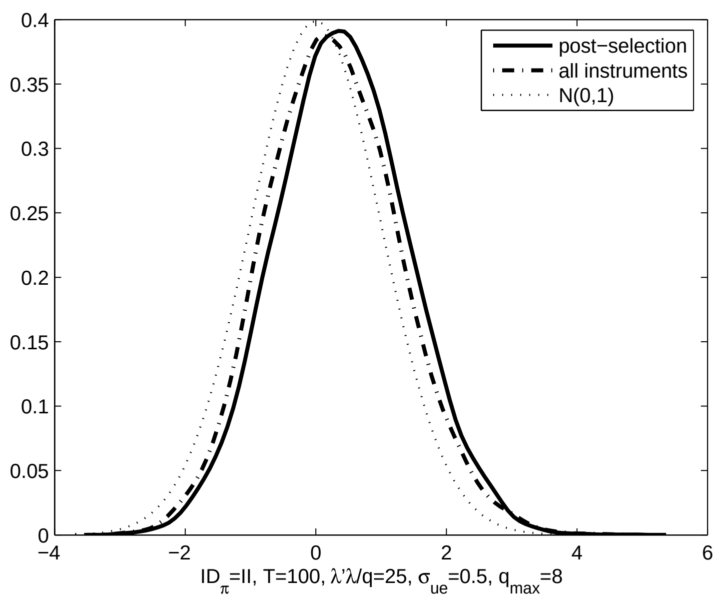


(b) Kernel Estimate of t-statistics

Figure 5.2: Kernel Estimate: All Z's vs $RMSC(c)$ $(b = 0.05)$ for $\mathbf{ID}_\pi = \mathbf{III}$



(a) Kernel Estimate of $\hat{\theta}_T(\iota_8)$ & $\hat{\theta}_T(\check{c}_{T,b})$



(b) Kernel Estimate of t-statistics

Figure 5.3: Kernel Estimate: All Z's vs $RMSC(c)$ ($b = 0.05$) for $\mathbf{ID}_\pi = \mathbf{II}$

Chapter 6

Conclusion

Generalized Method of Moments provides a computationally convenient method for obtaining estimators of the parameters of economic models. Under moderate conditions it can be shown that the resulting estimators are consistent and asymptotically normal. GMM is based on the population moment conditions and it is widely accepted that the information content in the population moment conditions has significant impacts on the quality of the asymptotic approximation to finite sample behavior.

In this dissertation, we focus our attention on a moment selection procedure that leads us to choose relevant (asymptotically efficient and non-redundant) moment conditions when weak identification is a possibility. To this end, we show that the entropy of the limiting distribution of the GMM estimator can provide a useful metric for the information content in the moment conditions and propose a moment selection criterion based on this entropy. Under some regularity conditions, the consistency of the moment selection procedure is established. Differently put, the proposed moment selection criterion tends to choose the selection vector associated with relevant moments with probability one in the limit as $T \to \infty$. We also introduce the concept of near redundancy and compare it to weak identification. The limiting behavior of relevant moment selection criterion is explored and using this behavior we establish the consistency result when weak identification is a possibility. Further, we generalized the discussions above to allow for the nonlinear dynamic GMM models and show that the desirable properties of the relevant moment selection criterion still hold in this more general framework.

Subsequent Monte Carlo simulation simulation studies evaluate the limiting behavior of the proposed moment selection criterion in the various specifications of candidate set, C. Simulation results suggest that $RMSC(c)$ can lead us to pick the relevant moment conditions with high probabilities when some instruments in set C are relevant and the other are redundant given the relevant ones (in the case of $\mathbf{ID}_\pi = \mathbf{I}$). When some prior information on the order of importance of instruments are known (in the case of $\mathbf{ID}_\pi = \mathbf{III}$), we need 16 candidates in set C to observe an improvement in the quality of the post-selection inference. In this case conducting Stock and Yogo's (2002) test for detecting weak identification prior to $RMSC(c)$ plays a key role for the observed improvements. It also noted that $RMSC(c)$ tends to select a parsimonious models, that is, it selects the first three or four important instruments out of sixteen possibilities. We conjecture this is due to the nature of the BIC-type penalty terms. It may impose too harsh penalty on the models with many instruments. Thus, more lenient AIC-type penalty is also considered with 8 possible candidates, but it does not seem that this helps to increase the number of chosen instruments significantly. As mentioned before, it is desirable to develop the optimal penalty term for the purpose of interest. This is out of scope of this dissertation and could open a road to the direction of future research. Interestingly enough, when all the instruments in the set C are equally important (in the case of $\mathbf{ID}_\pi = \mathbf{II}$), we do not observe any advantages of our moment selection criterion. In some cases, a moderate deterioration in the quality of the post-selection inference is found. We believe, however, this is not a major drawback of the proposed moment selection procedure. If researchers have credible belief that all the candidates for the instruments are equally important, they want to use all of them in the estimation and subsequent inferences. As a matter of fact, by doing this, researchers can achieve asymptotic efficiency and non-redundancy. However, it is still desirable to find the cause of deterioration and remedy for it. This is also left for the future research. We also detect some size distortion of Anderson–Rubin test when conducting $RMSC(c)$ under the circumstances where the value of concentration parameter is low and the degree of endogeneity of the regressors is high. Even though it is not reported here, we also find similar distortions for both Kleibergen's (2002) and Moreira's (2003) statistics[1]. This finding is somewhat puzzling because these three Gaussian similar tests are known as fully robust to weak identification. We conjecture the source of these size distortions is the statistical dependence between the $RMSC(c)$ and the test statistics of interest. A combined strategy of Stock and Yogo's (2002) test for detecting weak identification and our relevant moment selection procedure ameliorates these distortions a little bit but they do not disappear. Finding the source of these size distortions and suggesting the invariant test for the post-selection inference will be another branch of future work.

[1]The results will be provided upon requests.

Bibliography

AHMED, N., AND D. GOKHALE (1989): "Entropy expressions and their estimators for multivariate distributions," *IEEE Transactions on Information Theory*, 35, 688–692.

ANDREWS, D. W. K. (1991): "Heteroscedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

——— (1999): "Consistent moment selection procedure for generalized method of moments estimation," *Econometrica*, 67, 543–564.

BREUSCH, T., H. QIAN, P. SCHMIDT, AND D. WYHOWSKI (1999): "Redundancy of moment conditions," *Journal of Econometrics*, 91, 89–111.

DHRYMES, P. J. (2000): *Mathematics for Econometrics*. Springer Verlag, New York, NY, U.S.A., 3rd edn.

DONALD, S. G., AND W. K. NEWEY (2001): "Choosing the number of instruments," *Econometrica*, 69, 1161–1191.

GOLAN, A. (2002): "Information and entropy econometrics–Editor's view," *Journal of Econometrics*, 107, 1–15.

Hahn, J., and A. Inoue (2002): "A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators," *Econometric Reviews*, 21, 309–336.

Hall, A. R. (2005): *Generalized Method of Moments.* Oxford University Press, Oxford, U.K.

Hall, A. R., and A. Inoue (2003): "Information in Empirical Likelihood and Generalized Method of Moments estimation," Discussion paper, Department of Economics, North Carolina State University, Raleigh NC 27695.

Hall, A. R., and F. P. M. Peixe (2003): "A consistent method for the selection of relevant instruments in linear models," *Econometric Review*, forthcoming.

Hall, A. R., G. Rudebusch, and D. Wilcox (1996): "Judging instrument relevance in instrumental variables estimation," *International Economic Review*, 37, 283–298.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press, Princeton, NJ, U.S.A.

Hansen, L. P. (1982): "Large sample properties of Generalized Method of Moments estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., and K. J. Singleton (1982): "Generalized instrumental variables estimation of nonlinear rational expectations models," *Econometrica*, 50(5), 1269–1286.

Hayashi, F. (2000): *Econometrics.* Princeton University Press, Princeton, NJ, U.S.A.

Kleibergen, F. (2002): "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 70, 1781–1803.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Nelson, C., R. Startz, and E. Zivot (2000): "Improved Inference for the Instrumental Variables Estimator," Discussion Paper 1600, Econometric Society, available at http://ideas.repec.org/p/ecm/wc2000/1600.html.

Nelson, C. R., and R. Startz (1990): "The distribution of the instrumental varaibles estimator and its t ratio when the instrument is a poor one," *Journal of Business*, 63, S125–S140.

Rao, C. R. (1973): *Linear Statistical Inference and Its Application.* John Wiley & Sons, New York, NY, U.S.A., 2nd edn.

Searle, S. R. (1982): *Matrix Algebra Useful for Statistics*, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, NY, U.S.A.

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

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

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A survey of weak instruments and weak identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.

Stock, J. H., and M. Yogo (2002): "Testing for weak instruments in linear IV regression," Discussion Paper 0284, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberte/0284.html.

White, H. (1984): *Asymptotic Theory for Econometricians.* Academic Press, INC., New York, NY, U.S.A.

Appendix A

Matlab Codes

Matlab codes used for the simulation in this chapter are provided here. The presented script m–file is for calculating the results in Tables 5.9 and 5.10. The other calculations can be done with a slight modification to this script m–file. Two function m–files follow: one is for calculating $RMSC(c)$; the other for Wald and Anderson–Rubin tests.

A.1 Script m–file for the results in Table 5.9 & 5.10

```
%  Matlab code by Changmock Shin
%
%  Note: In this program, all 255 combinations of instruments considered
%        (Stock & Yogo test, RMSC(c), AR, Wald)

clear all;

fidc=fopen('pretest.out','w');
fprintf(fidc,'====================\n');
fprintf(fidc,' Pretest and RMSC(c)\n');
```

```
fprintf(fidc,'====================\n\n\n');
status=fclose(fidc);


%----------------------------;
% setting some initial values ;
%----------------------------;

randn('state',10);

p      = 1;              % # parameters in structural eq.
theta0 = 0;              % true value of structural eq. parameter

qmax   = 8;              % # possible candidates for instruments

rep    = 10000;          % # repetition
Ts     = 100;            %[100;500];      % 2 samples sizes
sigues = [0.1;0.5;0.9];  % 3 cov(u,e)'s for contemporaneous correaltions
mu2qs  = [1;10;25;50];   % lambda'*lambda/q

tncase=size(Ts,1)*size(sigues,1)*size(mu2qs,1)*3;         % total # cases

p_id   = 1;       % (RMSC penalty)  1:BIC-type   2:AIC-type   3:HQIC-type

%----------------------------------------------------------;
% Stock and Yogo (2001) Critical Values, from Table 1 & 2   ;
%----------------------------------------------------------;

if qmax==8;
    crit_sy = [20.25 11.39  6.69  4.99;...
               33.84 18.54 13.24 10.50];

elseif qmax==16;
    crit_sy = [21.28 11.50  6.39  4.59;...
               52.77 27.99 19.51 15.19];
end;

test_ind=1;  % 1 : based on TSLS bias / 2 : based on TSLS size

th_ind=1;    % 1 : b=0.05 / 2 : b=0.1  / 3 : b=0.2 / 4 : b=0.3
             % 1 : r=0.1  / 2 : r=0.15 / 3 : r=0.2 / 4 : r=0.25

%----------------------------------------------------------;
% constructing the set of all possible selection vectors, c ;
%----------------------------------------------------------;

cbars  = kron(ones(qmax,1),[0 1]); % qmax candidates for instruments
Nstar  = 2^qmax;                   % # combinations including |c| = 0
call   = zeros(qmax,Nstar);        % storage for all selection vectors
```

```
for i1=1:qmax;                      % loop : constructing all combinations
call(i1,:) = kron(ones(1,2^(i1-1)),kron(cbars(i1,:),ones(1,2^(qmax-i1))));
end;

call=call';                         % transpose (just for convenience)
call=call((sum(call,2)>=p)==1,:);  % select rows with |c| >= p

Nstar  = size(call,1);              % total # of possible selection vector


%-----------;
% main loop ;
%-----------;

for test_ind=1:2;

%for th_ind=1:4;

for i2=1:size(Ts,1);                        % do loop w.r.t. Ts

    T=Ts(i2);                               % sample size

    for i3=1:size(mu2qs,1);                 % do loop w.r.t. Rf2s

        mu2q=mu2qs(i3);                     % lambda'lambda/q


        for i4=1:size(sigues,1); % do loop w.r.t. sigues

            sigue=sigues(i4); % contemporaneous covariance
            Sigv=eye(qmax+2); % main diagonal elements = 1
            Sigv(1,2)=sigue;  % non-zero off diagonal elment, cov(u,e)
            Sigv(2,1)=sigue;  % non-zero off diagonal elment, cov(e,u)

            for Psi_ind=1:3;  %----------------------------------------;
                              % indicator: how to construct first-stag ;
                              %  1 : only the first IV relevant        ;
                              %  2 : IV's are equally important        ;
                              %  3 : IV's importance reduces gradually ;
                              %----------------------------------------;

                if Psi_ind==1;
                    Psi=[sqrt(qmax*mu2q/T);zeros(qmax-1,1)];
                    sel_count  = zeros(5,1);        % selection counter
                elseif Psi_ind==2;
                    Psi=sqrt(mu2q/T)*ones(qmax,1);
                    sel_count  = zeros(qmax+1,1);   % selection counter
                else
                    for qi=1:qmax
                        impo(qi)=(1-qi/(qmax+1))^4;
```

```
    end;
    Psi=sqrt(mu2q*qmax/(T*sum(impo.^2)))*impo';
    sel_count  = zeros(qmax*2,1);    % selection counter
end;

post_para  = zeros(rep,1);  % post-selection parameter
post_cover = zeros(2,1);    % post-selcetion coverage rate
W_width    = zeros(rep,1);  % Wald C.I. width
AR_width   = zeros(rep,1);  % AR C.I. width

ncase=3*(size(mu2qs,1)*size(sigues,1)*(i2-1)+...
                 size(sigues,1)*(i3-1)+(i4-1))+Psi_ind;

for r=1:rep;                  % do loop w.r.t. rep

    if mod(r,500)==0          % display repetition #
        fprintf('replication #: %6.0f\n',r);
    end;

    %---------------------------------------------------;
    % generating the random sample:                     ;
    %          v_t' = (u_t,e_t,z_t')' ~ NID(0,Sigv)     ;
    %            y  = X*theta0  + u                     ;
    %            X  = Z1*Psi1 + Z2*Psi2 + e             ;
    %---------------------------------------------------;

    v=randn(T,qmax+2)*chol(Sigv);
    Z = v(:,3:qmax+2);
    X = Z*Psi + v(:,2);
    y = X*theta0 + v(:,1);


    %----------------------------------------------------;
    % Stock & Yogo: test weak instruments (note that n=1) ;
    %----------------------------------------------------;

    Pz=Z*inv(Z'*Z)*Z';
    sige2hat=X'*(eye(T)-Pz)*X/(T-qmax);
    SYstat=X'*Pz*X/(qmax*sige2hat);

    if SYstat<crit_sy(test_ind,th_ind); % IV test (weak IV)

        cindx=size(sel_count,1);  % all IV's considered
        c_hat=ones(1,qmax);
        sel_count(cindx,1)=sel_count(cindx,1) + 1;

    else % IV test (strong IV)

    %------------------------------------------------;
    % calculating Relevant Moment Selection Criterion ;
```

```
                    %--------------------------------------------------;

                    RMSCtemp=zeros(Nstar,1);   % storage of RMSC(c)

                    for j=1:Nstar;             % calculate Nstar RMSC(c)
                    RMSCtemp(j,1)=tsls_rmsc(y,X,Z(:,find(call(j,:))),p_id);
                    end;

                    %---------------------------------------;
                    % post slecetion estimation & inference ;
                    %---------------------------------------;

                    [A,cindx]=min(RMSCtemp);     % pick minimum RMSC(c)

                    c_hat=call(cindx,:);         % choose selection vector

% count the selection
if Psi_ind==1;
% 1st row of sc1 = only the relevant instrument is selected!
% 2nd row of sc1 = relevant inst. + some but not all irrelevant insts
% 3rd row of sc1 = no relevant inst. + some or all irrelevant insts
% 4th row of sc1 = all instruments
sel_count=sel_count + ...
[sum(c_hat(:,1)==1 & sum(c_hat(:,2:8),2)==0);...
 sum(c_hat(:,1)==1 & sum(c_hat(:,2:8),2)~=0 & sum(c_hat(:,2:8),2)~=7);...
 sum(c_hat(:,1)==0 & sum(c_hat(:,2:8),2)>0);...
 sum(sum(c_hat,2)==8);0];
elseif Psi_ind==2;
% Case II : all instruments are equally important.
% i_th row of sc2 = i instruments are selected.
sel_count=sel_count + ...
[sum(sum(c_hat,2)==1);...
 sum(sum(c_hat,2)==2);...
 sum(sum(c_hat,2)==3);...
 sum(sum(c_hat,2)==4);...
 sum(sum(c_hat,2)==5);...
 sum(sum(c_hat,2)==6);...
 sum(sum(c_hat,2)==7);...
 sum(sum(c_hat,2)==8);0];
else
% Case III : instruments are ordered by the importance.
% x_th row, x=1,2,..,8 : first x instruments only (case 1 - 8)
% 9th row  : 1 instrument  but not case 1
% 10th row : 2 instruments but not case 2
% 11th row : 3 instruments but not case 3
% 12th row : 4 instruments but not case 4
% 13th row : 5 instruments but not case 5
% 14th row : 6 instruments but not case 6
% 15th row : 7 instruments but not case 7
sel_count=sel_count + ...
```

```
[sum(c_hat(:,1)==1 & sum(c_hat(:,2:8),2)==0);...
 sum(sum(c_hat(:,1:2),2)==2 & sum(c_hat(:,3:8),2)==0);...
 sum(sum(c_hat(:,1:3),2)==3 & sum(c_hat(:,4:8),2)==0);...
 sum(sum(c_hat(:,1:4),2)==4 & sum(c_hat(:,5:8),2)==0);...
 sum(sum(c_hat(:,1:5),2)==5 & sum(c_hat(:,6:8),2)==0);...
 sum(sum(c_hat(:,1:6),2)==6 & sum(c_hat(:,7:8),2)==0);...
 sum(sum(c_hat(:,1:7),2)==7 & c_hat(:,8)==0);...
 sum(sum(c_hat,2)==8);...
 sum(c_hat(:,1)~=1 & sum(c_hat,2)==1);...
 sum(sum(c_hat(:,1:2),2)~=2 & sum(c_hat,2)==2);...
 sum(sum(c_hat(:,1:3),2)~=3 & sum(c_hat,2)==3);...
 sum(sum(c_hat(:,1:4),2)~=4 & sum(c_hat,2)==4);...
 sum(sum(c_hat(:,1:5),2)~=5 & sum(c_hat,2)==5);...
 sum(sum(c_hat(:,1:6),2)~=6 & sum(c_hat,2)==6);...
 sum(sum(c_hat(:,1:7),2)~=7 & sum(c_hat,2)==7);0];
end;

                    end;    % end of IV test

                    [tslsest,Wtest,ARtest,Wlength,ARlength]= ...
                           tsls_test(y,X,Z(:,find(c_hat)),theta0);

                    post_para(r,1)  = tslsest;
                    post_cover(1,1) = post_cover(1,1)+Wtest;   % Wald test
                    post_cover(2,1) = post_cover(2,1)+ARtest;  % AR test
                    W_width(r,1)    = Wlength;
                    AR_width(r,1)   = ARlength;

                end    % end loop w.r.t. rep

% Write output file

fidc=fopen('pretest.out','a');
fprintf(fidc,'\n\n-----------------------------------------------\n');
fprintf(fidc,'Results for case number:   %5.0f\n',ncase);
fprintf(fidc,'-----------------------------------------------\n');

if Psi_ind==1
    fprintf(fidc,'Only first instrument is relevant.\n');
elseif Psi_ind==2
    fprintf(fidc,'All instruments are equally important.\n');
else
    fprintf(fidc,'Importance of instruments reduces gradually.\n');
end

if test_ind==1
    fprintf(fidc,'Stock and Yogo test based on TSLS bias.\n');
    if th_ind==1
        fprintf(fidc,'Test threshold:            b=0.05\n');
    elseif th_ind==2
```

```
        fprintf(fidc,'Test threshold:            b=0.10\n');
    elseif th_ind==3
        fprintf(fidc,'Test threshold:            b=0.20\n');
    else
        fprintf(fidc,'Test threshold:            b=0.30\n');
    end;
else
    fprintf(fidc,'Stock and Yogo test based on TSLS size.\n');
    if th_ind==1
        fprintf(fidc,'Test threshold:            r=0.10\n');
    elseif th_ind==2
        fprintf(fidc,'Test threshold:            r=0.15\n');
    elseif th_ind==3
        fprintf(fidc,'Test threshold:            r=0.20\n');
    else
        fprintf(fidc,'Test threshold:            r=0.25\n');
    end;
end

fprintf(fidc,'Sample size               %5.0f\n',T);
fprintf(fidc,'lambda''lambda/q          %5.0f\n',mu2q);
fprintf(fidc,'Cov(u,e)                  %5.3f\n',sigue);
fprintf(fidc,'----------------------------------------------\n');
fprintf(fidc,'Empirical Selection Probabilities\n');
fprintf(fidc,'----------------------------------------------\n');
for qi=1:size(sel_count,1)-1;
fprintf(fidc,'%3.0f.              : %6.3f & \n',qi,sel_count(qi,1)/rep);
end;
fprintf(fidc,'%3.0f. IV''s are weak : %6.3f & \n',...
    size(sel_count,1), sel_count(size(sel_count,1),1)/rep);
fprintf(fidc,'----------------------------------------------\n');
fprintf(fidc,'Sampling Porperties of Post-Selection Estimator\n');
fprintf(fidc,'----------------------------------------------\n');
fprintf(fidc,'Med Bias           : %6.3f & \n',median(post_para)-theta0);
fprintf(fidc,'Cov Rate (W)       : %6.3f & \n',post_cover(1,1)/rep);
fprintf(fidc,'CI Width (W)       : %6.3f & \n',median(W_width));
fprintf(fidc,'Cov Rate (AR)      : %6.3f & \n',post_cover(2,1)/rep);
if median(AR_width) > 1000
    fprintf(fidc,'CI Width (AR)      :  \\infi & \n');
else
    fprintf(fidc,'CI Width (AR)      : %6.3f & \n',median(AR_width));
end;
fprintf(fidc,'----------------------------------------------\n\n\n\n\n');
status=fclose(fidc);

% Display results on screen

fprintf('\n\n----------------------------------------------\n');
fprintf('Results for case number:   %5.0f\n',ncase);
fprintf('----------------------------------------------\n');
```

```
if Psi_ind==1
    fprintf('Only first instrument is relevant.\n');
elseif Psi_ind==2
    fprintf('All instruments are equally important.\n');
else
    fprintf('Importance of instruments reduces gradually.\n');
end

if test_ind==1
    fprintf('Stock and Yogo test based on TSLS bias.\n');
    if th_ind==1
        fprintf('Test threshold:            b=0.05\n');
    elseif th_ind==2
        fprintf('Test threshold:            b=0.10\n');
    elseif th_ind==3
        fprintf('Test threshold:            b=0.20\n');
    else
        fprintf('Test threshold:            b=0.30\n');
    end;
else
    fprintf('Stock and Yogo test based on TSLS size.\n');
    if th_ind==1
        fprintf('Test threshold:            r=0.10\n');
    elseif th_ind==2
        fprintf('Test threshold:            r=0.15\n');
    elseif th_ind==3
        fprintf('Test threshold:            r=0.20\n');
    else
        fprintf('Test threshold:            r=0.25\n');
    end;
end

fprintf('Sample size                %5.0f\n',T);
fprintf('lambda''lambda/q           %5.0f\n',mu2q);
fprintf('Cov(u,e)                   %5.3f\n',sigue);
fprintf('-----------------------------------------------\n');
fprintf('Empirical Selection Probabilities\n');
fprintf('-----------------------------------------------\n');
for qi=1:size(sel_count,1)-1;
fprintf('%3.0f.               : %6.3f\n',qi,sel_count(qi,1)/rep);
end;
fprintf('%3.0f. IV''s are weak : %6.3f\n',...
    size(sel_count,1), sel_count(size(sel_count,1),1)/rep);
fprintf('-----------------------------------------------\n');
fprintf('Sampling Porperties of Post-Selection Estimator\n');
fprintf('-----------------------------------------------\n');
fprintf('Med Bias            : %6.3f\n',median(post_para(:,1))-theta0);
fprintf('Cov Rate (W)        : %6.3f\n',post_cover(1,1)/rep);
fprintf('CI Width (W)        : %6.3f\n',median(W_width));
```

```
fprintf('Cov Rate (AR)       : %6.3f\n',post_cover(2,1)/rep);
if median(AR_width) > 1000
    fprintf('CI Width (AR)       : Infinity\n');
else
    fprintf('CI Width (AR)       : %6.3f\n',median(AR_width));
end;
fprintf('-----------------------------------------------\n\n\n\n\n');


% Store data for generating Tables

if test_ind==1;
if Psi_ind==1;
nr=round(ncase/3)+1;
% selection probabilities
for qi=1:size(sel_count,1)
Tbl_pi1b(qi,nr) = sel_count(qi,1)/rep;
end;

% median bias
Tbl_pi1b(size(sel_count,1)+1,nr) = median(post_para)-theta0;

% Wald Test : Empirical Cov Rate & C.I. Width
Tbl_pi1b(size(sel_count,1)+2,nr) = post_cover(1,1)/rep;
Tbl_pi1b(size(sel_count,1)+3,nr) = median(W_width);

% AR Test : Empirical Cov Rate & C.I. Width
Tbl_pi1b(size(sel_count,1)+4,nr) = post_cover(2,1)/rep;
Tbl_pi1b(size(sel_count,1)+5,nr) = median(AR_width);

elseif Psi_ind==2;
nr=round(ncase/3);
% selection probabilities
for qi=1:size(sel_count,1)
Tbl_pi2b(qi,nr) = sel_count(qi,1)/rep;
end;

% median bias
Tbl_pi2b(size(sel_count,1)+1,nr) = median(post_para)-theta0;

% Wald Test : Empirical Cov Rate & C.I. Width
Tbl_pi2b(size(sel_count,1)+2,nr) = post_cover(1,1)/rep;
Tbl_pi2b(size(sel_count,1)+3,nr) = median(W_width);

% AR Test : Empirical Cov Rate & C.I. Width
Tbl_pi2b(size(sel_count,1)+4,nr) = post_cover(2,1)/rep;
Tbl_pi2b(size(sel_count,1)+5,nr) = median(AR_width);

else
nr=ncase/3;
```

```
% selection probabilities
for qi=1:size(sel_count,1)
Tbl_pi3b(qi,nr) = sel_count(qi,1)/rep;
end;

% median bias
Tbl_pi3b(size(sel_count,1)+1,nr) = median(post_para)-theta0;

% Wald Test : Empirical Cov Rate & C.I. Width
Tbl_pi3b(size(sel_count,1)+2,nr) = post_cover(1,1)/rep;
Tbl_pi3b(size(sel_count,1)+3,nr) = median(W_width);

% AR Test : Empirical Cov Rate & C.I. Width
Tbl_pi3b(size(sel_count,1)+4,nr) = post_cover(2,1)/rep;
Tbl_pi3b(size(sel_count,1)+5,nr) = median(AR_width);
end;

else
if Psi_ind==1;
nr=round(ncase/3)+1;
% selection probabilities
for qi=1:size(sel_count,1)
Tbl_pi1s(qi,nr) = sel_count(qi,1)/rep;
end;

% median bias
Tbl_pi1s(size(sel_count,1)+1,nr) = median(post_para)-theta0;

% Wald Test : Empirical Cov Rate & C.I. Width
Tbl_pi1s(size(sel_count,1)+2,nr) = post_cover(1,1)/rep;
Tbl_pi1s(size(sel_count,1)+3,nr) = median(W_width);

% AR Test : Empirical Cov Rate & C.I. Width
Tbl_pi1s(size(sel_count,1)+4,nr) = post_cover(2,1)/rep;
Tbl_pi1s(size(sel_count,1)+5,nr) = median(AR_width);

elseif Psi_ind==2;
nr=round(ncase/3);
% selection probabilities
for qi=1:size(sel_count,1)
Tbl_pi2s(qi,nr) = sel_count(qi,1)/rep;
end;

% median bias
Tbl_pi2s(size(sel_count,1)+1,nr) = median(post_para)-theta0;

% Wald Test : Empirical Cov Rate & C.I. Width
Tbl_pi2s(size(sel_count,1)+2,nr) = post_cover(1,1)/rep;
Tbl_pi2s(size(sel_count,1)+3,nr) = median(W_width);
```

```
% AR Test : Empirical Cov Rate & C.I. Width
Tbl_pi2s(size(sel_count,1)+4,nr) = post_cover(2,1)/rep;
Tbl_pi2s(size(sel_count,1)+5,nr) = median(AR_width);

else
nr=ncase/3;
% selection probabilities
for qi=1:size(sel_count,1)
Tbl_pi3s(qi,nr) = sel_count(qi,1)/rep;
end;

% median bias
Tbl_pi3s(size(sel_count,1)+1,nr) = median(post_para)-theta0;

% Wald Test : Empirical Cov Rate & C.I. Width
Tbl_pi3s(size(sel_count,1)+2,nr) = post_cover(1,1)/rep;
Tbl_pi3s(size(sel_count,1)+3,nr) = median(W_width);

% AR Test : Empirical Cov Rate & C.I. Width
Tbl_pi3s(size(sel_count,1)+4,nr) = post_cover(2,1)/rep;
Tbl_pi3s(size(sel_count,1)+5,nr) = median(AR_width);
end;
end;

            end    % end loop w.r.t. Psi_ind

        end        % end loop w.r.t. sigues

    end            % end loop w.r.t. mu2q

end                % end loop w.r.t. Ts

%end                % end loop w.r.t. test thresholds
end                % end loop w.r.t. test_ind


save pretest.mat Tbl_pi1b Tbl_pi2b Tbl_pi3b Tbl_pi1s Tbl_pi2s Tbl_pi3s;
```

A.2 Function m–file for $RMSC(c)$

```
function rmsc = tsls_rmsc(y,x,Z,id)
%
%  usage: rmsc = tsls_rmac(y,x,Z,id)
%
%    rmsc   : Relevant Moment Selection Criterion using BIC-type penalty
%
[T,k]  = size(Z);
```

```
b      = inv(x'*Z*inv(Z'*Z)*Z'*x)*x'*Z*inv(Z'*Z)*Z'*y;
u      = y-x*b;
sigma2 = u'*u/T;

if id==1       % BIC-type penalty
rmsc  = log(sigma2/(x'*Z*inv(Z'*Z)*Z'*x))+(k-1)*log(sqrt(T))/sqrt(T);

elseif id==2   % AIC-type penalty
rmsc  = log(sigma2/(x'*Z*inv(Z'*Z)*Z'*x))+(k-1)*2/sqrt(T);

elseif id==3   % HQIC-type penalty
bb=2.01;       % some constant such that bb>2
rmsc  = log(sigma2/(x'*Z*inv(Z'*Z)*Z'*x))+...
                                (k-1)*log(log(sqrt(T)))/sqrt(T)*bb;
end;
```

A.3 Function m–file for Wald and Anderson–Rubin Tests

```
function [b,Wtest,ARtest,Wlength,ARlength] = tsls_test(y,x,Z,beta0)
%
%  usage: [b,Wstat,ARstat,Wlength,ARlength] = tsls_test(y,x,Z,beta0)
%
%    b      : 2SLS estimator
%    Wtest  : Wald statistic (Ho: b=beta0)
%    ARtest : Anderson-Rubin statistic (Ho: b=beta0 -> Chi-sq(k)/k)
%    For both Wstat and ARstat, 1: Accept Ho, 0: Reject Ho

[T,k] = size(Z);
b     = inv(x'*Z*inv(Z'*Z)*Z'*x)*x'*Z*inv(Z'*Z)*Z'*y;
p     = size(b,1);
crit1 = chi2inv(0.95,p);
crit2 = chi2inv(0.95,k);
u     = y-x*b;
sig2  = u'*u/T;
% Wstat  = (b-beta0)'*(x'*Z*inv(Z'*Z)*Z'*x)*(b-beta0)/sigma2; % Wald Stat

Wtest = (b-beta0)'*(x'*Z*inv(Z'*Z)*Z'*x)*(b-beta0)/sig2 < crit1;
Wlength = 2*sqrt(sig2*crit1/(x'*Z*inv(Z'*Z)*Z'*x));

% See Stock, Wright and Yogo (JBES, 2002)
Y_      = [y x];
a0      = [beta0;1];
```

```
b0      = [1;-beta0];
Omega   = Y_'*(eye(T)-Z*inv(Z'*Z)*Z')*Y_/(T-k);
ES      = inv(sqrtm(Z'*Z))*Z'*Y_*b0/sqrt(b0'*Omega*b0);

% ARstat  = (ES'*ES)/k;    % Anderson-Rubin Stat
% , or...
% ARstat = ((y-x*beta0)'*(Z*inv(Z'*Z)*Z')*(y-x*beta0)/k)  /...
%          ((y-x*beta0)'*(eye(T)-Z*inv(Z'*Z)*Z')*(y-x*beta0)/(T-k));
ARtest  = (ES'*ES) < crit2;


% See Hahn and Inoue (Econometric Review, 2002)

A = x'*Z*inv(Z'*Z)*Z'*x-crit2*x'*(eye(T)-Z*inv(Z'*Z)*Z')*x/(T-k);
B = y'*Z*inv(Z'*Z)*Z'*x-crit2*y'*(eye(T)-Z*inv(Z'*Z)*Z')*x/(T-k);
C = y'*Z*inv(Z'*Z)*Z'*y-crit2*y'*(eye(T)-Z*inv(Z'*Z)*Z')*y/(T-k);

if A > 0
    if B^2-A*C > 0   ARlength = 2*sqrt(B^2-A*C)/A;
    else             ARlength = 0;
    end
else
    ARlength = realmax;
end;
```