ESSAYS ON MISSPECIFIED MODELS

BY

BING ZUO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Anil K. Bera, Chair
Professor Xiaofeng Shao
Asistant Professor JiHyung Lee
Asistant Professor Eun Yi Chung

Abstract

This thesis identifies the asymptotic properties of generalized empirical likelihood estimators when moment conditions are not correctly specified. Classical generalized empirical likelihood estimators rely on the correct moment conditions, however, those conditions are mostly generated from economic theory and some of them are not testable. Hence, it is needed to understand the property of the estimators and test statistics when moments are misspecified and provide robust estimators and test statistics when moment conditions are misspecified.

Chapter 1, "Robust Inference for Instrumental Variable Models with Locally Non-exogenous Instruments", highlights that conventional tests often fail to give accurate inferences when exogeneity conditions are mildly violated in instrumental variable models. The sizes of those tests can be considerably distorted due to their non-centrally distributed test statistics under the null hypothesis. This paper proposes an adjusted score-type test to correct this size distortion while preserving good discriminatory power. We prove that under the null hypothesis, this adjusted score-type test statistic converges to a central chi-squared distribution and thus is not adversely affected by local non-exogeneity. Furthermore, the Monte Carlo simulations confirm that our newly proposed test has considerable size improvement over the conventional ones, while their power is not very different.

Chapter 2, "Mis-specification-Robust Bootstrap for Empirical Likelihood Estimators ", proposes an adapted bootstrap testing procedure for empirical likelihood estimators. This method extends the bootstrap method in Lee (2014) by using the empirical likelihood weights, which could improve the efficiency if the moment condition model is correctly specified. This proposed bootstrap method is also robust to model misspecification as shown in Lee (2014). The first-order asymptotic validity of the proposed procedure is shown, and multiple Monte Carlo Studies are conducted to support the theoretical findings.

Chapter 3, "Higher Order MSE Comparisons of Generalized Empirical Likelihood Estimators", calculates the higher order asymptotic mean square errors (MSE) of generalized empirical likelihood (GEL) estimators on a simple linear model. It is well known from Newey and Smith (2004) that the Empirical likelihood (EL) estimator has the smallest higher-order asymptotic bias among the GEL estimators; however, in this paper we find that the EL estimator no longer has this property for the criteria of MSE. We propose a data-driven method to achieve the least asymptotic higher-order MSE in the GEL family.

Acknowledgments

I would like to thank all the person who help and support me during my PhD life. Thanks my advisor, Professor Anil K. Bera, for constantly encouraging me, providing guidance to my research, and teaching me to be a better person. Thanks my committee members, Professor Xiaofeng Shao, Professor JiHyung Lee, and Professor Eun Yi Chung, for their generous support and devotion.

I also want to say thanks to my friends who enrich my life and share my emotion at UIUC. Last, I would like to thank my family for inspiring me to finish my Ph.D. Thanks my wife, Cong Zhang, for pushing me to complete my dissertation. Thanks my lovely son, Albert Zuo, for motivating me to work hard and enjoy little things from life.

Table of Contents

Chapter 1

Robust Inference for Instrumental Variable Models with Locally Non-exogenous Instruments

1.1 Introduction

For an instrumental variable (IV) model, the adverse impacts of mildly violated exogeneity conditions on its inference have become a growing concern in recent literature. The study of this topic is often referred to as local non-exogeneity (Caner 2014, for example). A recent development in this field has found that when some of the instruments used do not perfectly satisfy the exogeneity conditions, inference using conventional tests may fail to have the correct size (Guggenberger (2012), Berkowitz, Caner and Fang (2008, 2012), and Caner (2014) among others).

However, few improved methods have been offered to obtain robust size performance when instruments are locally non-exogenous. Therefore, we ask the question: can we develop a size-robust inference method that is not affected by using locally non-exogenous instruments? In the following of this paper, we provide a positive answer to this question.

In this paper, we propose an adjusted score-type test statistic in the framework of the generalized empirical likelihood (GEL) method. We argue that size of the conventional score-type test statistic can be greatly distorted in the presence of locally non-exogenous instruments. As an unfavorable results of size distortion, we may falsely reject the null hypothesis more often than it is supposed to be because the distribution under the null converges away from its usual distribution. But the adjusted version we proposed can be size-robust due to the property that its asymptotic distribution is free from nuisance parameters under the null hypothesis. Furthermore, to improve small sample efficiency, we develope our inference method under the framework of GEL. This estimating technique is often considered as a compelling alternative to generalized method of moments (GMM) due to its property of small sample bias (Newey and Smith (2004)).

To understand why the conventional score-type test statistic fail to have correct size, we show that in the presence of local non-exogeneity its asymptotic distribution under the null hypothesis is a non-central chi-squared distribution. This result is consistent with the literature. Several recent works have also pointed out the invalidity of other conventional testing procedures. For example, Berkowitz, Caner and Fang (2008, 2012) showed that both size of t test and Anderson-Rubin test may greatly diverge from its nominal level ($\alpha = 5\%$) as level of endogeneity in instruments increases. Guggenberger (2012) also compares finite sample performance of various commonly used test statistics, such as Anderson-Rubin test, Moreira's test, and Kleibergen's K test. He ranked these tests according to their robustness to non-exogeneity instruments and found that none of them has consistent size performance as quality of the instruments deteriorates. In this paper, we explicitely develop the asymptotic distribution of conventional score test in the presence of non-exogenous instruments. We find that conventional score test adversely converges to a non-central chi-squared distribution where the magnitude of non-centrality depends on the value of nuisance parameters. In addition, the finite-sample experiment also shows considerable size distortion of using conventional score test.

Taking account of the unfavorable size distortion of conventional score test, it is necessary to develop a test statistic that is robust to local non-exogeneity. This paper proposes an adjusted score-type test statistic which can help improve size over its conventional alternatives by adjusting score test using score functions of nuisance parameters. Taking a close look at the non-central chi-squared distribution of conventional score test, we find that nuisance parameters of local non-exogeneity is the main source of size distortion. Naturally, if we can remove all nuisance parameters from asymptotic distribution of the test statistic under the null, this test statistic can be size-robust to local non-exogeneity conditions. In Section 3, we show this removal of nuisance parameters can be achieved by a simple adjustment of the conventional score test statistic using score functions of nuisance parameters. Therefore, by doing this, we can obtain correct size. In addition, we explicitely show the relationship between conventional score test statistic and its adjusted version with an application of linear structural IV model.

Finally, our Monte Carlo experiments compare the fintie sample performance of conventional GEL score tests and of our adjusted tests. The results (1) confirms the finding in the literature that conventional test staistics suffer from size distortion arising from the invalidity of exogeneity condtions, and (2) shows that our new tests have much improved finite-sample performance in reducing size distortion without losing much of power.

The rest of this paper is organized as follows. In section 2, we introduce our model framework and GEL method. In Section 3 we first develop the asymptotic distribution of conventional score test and explain why the conventional methods fail in the presence of locally non-exogenous variables. Then we provide an adjusted score-type test and show how the newly proposed test method is asymptotically robust regardless of the validity of the exogeneity conditions. A simple application is also discussed in this section. Section 4 reports results of Monte Carlo simulation. Section 5 concludes.

1.2 Model and GEL Estimators

We consider the instrumental variable model as follows:

$$y_i = f(x_i, \beta) + \epsilon_i, \tag{1.1}$$

where x_i is a $k \times 1$ vector of endogenous variables, $\beta \in \mathcal{B} \subset \mathbb{R}^k$ is a $k \times 1$ vector, $f(\cdot,\cdot)$ is a continuous function that maps $\mathbb{R}^k \times \mathcal{B}$ into $\mathbb{R}$. ϵ_i is the error term. We assume there are two types of instruments: z_{vi}, a $(q-\ell) \times 1$ vector of "valid" instruments for which the exogeneity condition is perfectly satisfied, and z_{di}, an $\ell \times 1$ vector of "defective" instruments in which the moment conditions are locally voilated. Therefore, the moment conditions can be defined as:

$$g_i(\beta, \delta) = \mathbb{E} \begin{bmatrix} z_{vi}\epsilon_i \\ z_{di}\epsilon_i \end{bmatrix} = \begin{bmatrix} 0 \\ \delta \end{bmatrix}$$

where δ is a local nonexogeneity parameter with $\delta = C/\sqrt{n}$, and C is a $\ell \times 1$ vector of constants. Under this framework, we characterizes potential local violation of exogeneity condition in instruments z_{di}.

In the following context, we denote parameter $\theta = (\beta', \delta')'$ and $\delta_0 = 0$.

Definition 1.2.1 *(GEL estimation). The GEL estimator $\hat{\theta}$ for θ is defined as (Guggenberger and Smith (2011)):*

$$\widehat{\theta} := arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' g_i(\theta))], \tag{1.2}$$

where Θ is a compact subset of $\mathbb{R}^{k+\ell}$, $\hat{\Lambda}_n(\theta) = \{\lambda \in \mathbb{R}^q : \lambda' g_i(\theta) \in \mathbb{Q}$ for $i = 1, ..., n\}$, , $\mathbb{Q}$ is an open interval of the real line containing 0, and the real-valued function $\rho(.) : \mathbb{Q} \to \mathbb{R}$ is strictly concave on its domain.

Assumption 1.2.1 *(a) The function $\rho(v)$ is twice continuously differentiable in a neighborhood of 0. (b) $\rho_1 = \rho_2 = -1$, where we define $\rho_j(v) = \partial^j \rho(v)/\partial v^j$ with $\rho_j = \rho_j(0)$ for any nonnegative integer j.*

The three mostly used GEL estimators in the literature are the empirical likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994), exponential tilting (ET) estimator of Kitamura and Stutzer (1997), and continuous-updating estimator (CUE) of Hansen, Heaton and Yaron (1996) which correspond to $\rho(\nu) = \ln(1 - \nu)$, $\rho(\nu) = -\exp(\nu)$, and $\rho(\nu) = -(1+\nu)^2/2$, respectively. See Parente and Smith (2014) for a recent survey on GEL methods.

1.3 Test Statistics

In this section we provide a score-type statistic to test for the parameters of interest with $H_0 : \beta = \beta_0$ versus $H_a : \beta \neq \beta_0$. We begin with notation and definitions of the score functions and moment restrictions. Let $\theta_0 = (\beta_0, \delta/\sqrt{n})$, $\overline{g} = 1/n \sum_{i=1}^{n} g_i(\theta)$. The $G_{\beta i} = \partial g_i(\theta)/\partial \beta$ and $G_{\delta i} = \partial g_i(\theta)/\partial \delta$, $G_\beta = \lim_{n \to \infty} \mathbb{E}[n^{-1} \sum_{i=1}^{n} \partial g_i(\theta_0)/\partial \beta]$ and $G_\delta = \lim_{n \to \infty} \mathbb{E}[n^{-1} \sum_{i=1}^{n} \partial g_i(\theta_0)/\partial \delta]$, and $Var(g_i(\theta_0)) = \lim_{n \to \infty} \mathbb{E}[n^{-1} \sum_{i=1}^{n} g_i(\theta_0) g_i(\theta_0)'] \equiv \Omega_{q \times q}$.

Definition 1.3.1 *(Score functions) Let* $\hat{\theta} = (\beta_0, 0)$, $\hat{\lambda} = \arg\sup_{\lambda \in \hat{\Lambda}_n(\hat{\theta})} \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' g_i(\hat{\theta})]$. *By GEL algorithm, the score functions of* β *and* δ *are:*

$$\hat{D}_\beta = \frac{1}{n} \sum_{i=1}^{n} \left[\rho_1(\hat{\lambda}' g_i(\hat{\theta})) G'_{\beta i} \hat{\lambda}\right], \tag{1.3}$$

$$\hat{D}_\delta = \frac{1}{n} \sum_{i=1}^{n} [\rho_1(\hat{\lambda}' g_i(\hat{\theta})) G'_{\delta i} \hat{\lambda}], \tag{1.4}$$

respectively.

Also, let $\Sigma_\beta = G'_\beta \Omega^{-1} G_\beta$, $\Sigma_{\beta\delta} = G'_\beta \Omega^{-1} G_\delta$, *and* $\Sigma_\delta = G'_\delta \Omega^{-1} G_\delta$ *denote the variance-covariance matrices of the score functions.*

We study the asymptotic properties of the GEL-based test statistics under the following assumptions.

Assumption 1.3.1 $\theta_0 \in int(\Theta)$ *is the unique solution to* $E(g(\theta_0)) = 0$.

Assumption 1.3.2 *(a) The moment function* $g(\theta)$ *is continuously differentiable in a neighborhood of* β_0*; (b)* $E[sup_{\theta \in \Theta} \|g(\theta)\|^\alpha] < \infty$ *for some* $\alpha > 2$ *; (c)* $E[sup_{\theta \in \Theta} \|\partial g(\theta)/\partial \theta'\|] < \infty$*; (d)* Ω *is nonsingular.*

Assumption 1.3.3 *The matrices* G_β *and* G_δ *are of full ranks.*

Remark 1.3.1 *Assumptions 2.5.1-1.3.3 follow Newey and Smith(2004) with minor changes(??). In Assumption 2.5.1 the true parameter* θ_0 *contains nuisance parameter* δ *to indicate local nonendogenity. Since*

$$\hat{\lambda} = \arg \sup_{\lambda \in \hat{\Lambda}_n(\hat{\theta})} \frac{1}{n} \sum_{i=1}^{n} [\rho(\hat{\lambda}' g_i(\hat{\theta})],$$

bound conditions for moments evaluated at $\hat{\theta}$ *are needed in the classical expansion theory.*

Lemma 1.3.1 *If exogeneity conditions of instruments* z_{id} *are valid, i.e.,* $\delta = 0$, *under the null hypothesis, the standard score test statistic asymptotically follows a central chi-squared distribution with degrees of freedom* k*:*

$$RS = n\hat{D}_\beta \hat{\Sigma}_\beta^{-1} \hat{D}'_\beta \xrightarrow{d} \chi^2_k(0), \tag{1.5}$$

where $\hat{\Sigma}_\beta$ is a consistent estimator of Σ_β.

1.3.1 Robust Score Test Statistic

This subsection proves why the conventional score-type inference methods fail in the presence of locally non-exogeneity and how the newly proposed test method is asymptotically size-robust using nuisance parameter adjusted score test statistic. We show that the conventioanl score test statistic fails because of its non-central chi-squared distribution where its noncentrality parameter depends on the nuisance parameters of local non-exogeneity. But the newly proposed test statistic is nuisance parameter free and has a central chi-squared distribution. Therefore it is size-robust.

The main source of size distortion in conventional score test is nuisance parameters in the asymptotic distributions. In the following theorem, we prove that when some instruments are locally non-exogenous, conventional score test converges to a non-central chi-squared distribution with non-centrality parameter depending on nuisance parameters of non-exogeneity condition.

Theorem 1.3.1 *Under the null hypothesis, if instruments z_{di} are locally non-exogenous, i.e., $E(z_{di}\epsilon_i) = \delta/\sqrt{n}$, score test statistic is non-pivotal, such that its limiting distribution depends on values of non-exogeneity parameters δ:*

$$RS = n\hat{D}_\beta \hat{\Sigma}_\beta^{-1} \hat{D}'_\beta \xrightarrow{d} \chi^2_K(\mu_2), \tag{1.6}$$

where $\mu_2 = \delta' \Sigma'_{\beta\delta} \Sigma_\beta^{-1} \Sigma_{\beta\delta} \delta$.

Remark 1.3.2 *Theorem 1.3.1 indicates that inference of standard score tests is not size-robust to local non-exogeneity. Its size distortion depends on the value of non-centrality parameter μ_2. If correlation between β and δ are non-zero, that is $\Sigma_{\beta\delta} \neq 0$, the value of μ_2 increases as $|\delta|$ becomes large. But if $\Sigma_{\beta\delta} = 0$, size of the score test may not be affected by the value of $|\delta|$ because $\mu_2 = 0$ anyway. In this case, conventional score test is size-robust despite the validity of exogeneity conditons in z_{di}.*

To remove nuisance parameters δ from the asymptotic distribution of a score test, we adjust its test statistic using score functions of δ so that its asymptotic distribtuion under the null hypothesis is pivotal and centrally distributed.

Theorem 1.3.2 *Define an adjusted score function of β as $\hat{D}^* := \hat{D}_\beta - \hat{\Sigma}_{\beta\delta}\hat{\Sigma}_\delta^{-1}\hat{D}_\delta$. Under the null hypothesis, given $E(z_{di}\epsilon_i) = \delta/\sqrt{n}$, we have:*

1. *adjusted score function of β converges to a normal distribution with mean 0:*

$$\sqrt{n}\hat{D}^* \xrightarrow{d} N(0, \hat{\Sigma}^*), \tag{1.7}$$

where $\hat{\Sigma}^* := \hat{\Sigma}_\beta - \hat{\Sigma}_{\beta\delta}\hat{\Sigma}_\delta^{-1}\hat{\Sigma}'_{\beta\delta}$.

2. *and the adjusted score test statistic:*

$$RS^* := n\hat{D}^{*\prime}(\hat{\Sigma}_\beta - \hat{\Sigma}_{\beta\delta}\hat{\Sigma}_\delta^{-1}\hat{\Sigma}'_{\beta\delta})^{-1}\hat{D}^*. \tag{1.8}$$

converges to a central chi-squared distribution with degrees of freedom K:

$$RS^* \xrightarrow{d} \chi^2_K(0). \tag{1.9}$$

By showing that the adjusted score test has the same asymptotic size as the standard score test under a correctly specified model, the limiting distribution of RS^* indicates the "robustness" of the new test regardless of the presence of nuisance parameters for the local nonexogeneity conditions. For example, in a case where l instruments potentially have direct effect on the outcome variable y, these instruments thus may violate the exogeneity conditions. Our adjusted score test can guarantee to obtain a robust size for any local invalidity of these instruments without knowing the exact level of nonexogeneity. This is convenient especially in empirical research where instrumental variable models are used with multiple instruments. One possible issue associated with this adjusted score test is a trade-off between the type I error and the ability to reject the null hypothesis when β is distinct from β_0. The simulation experiments in section 1.4 provide some evidences for such a trade-off by showing a slightly lower level in power comparing to the standard GEL-based statistics. Note that such a difference in power between the standard and our adjusted test statistics are closely negligible if sample size is large.

Corollary 1.3.3 (No misspecification under the local alternative) *Given* $\delta = 0$, *under the local alternative* $H_a : \beta = \beta_0 + \tau/\sqrt{n}$,

$$RS \xrightarrow{d} \chi^2_k(\mu_1), \tag{1.10}$$

$$RS^* \xrightarrow{d} \chi^2_k(\mu_3), \tag{1.11}$$

where $\mu_1 = \tau'\Sigma_\beta\tau$, $\mu_3 = \tau'(\Sigma_\beta - \Sigma_{\beta\delta}\Sigma_\delta^{-1}\Sigma'_{\beta\delta})\tau$.

This corollary compares asymptotic powers of adjusted and standard score tests under no misspecification. Note that $\mu_3 - \mu_1 = \tau'\Sigma_{\beta\delta}\Sigma_\delta^{-1}\Sigma'_{\beta\delta}\tau \geq 0$ indicates that our adjusted score tests may have loss of asymptotic power relative to standard tests when model is correctly specified. In the language of Bera and Yoon (1993), this is also called a cost of robustification, where correction of asymptotic size causes declines in power.

1.3.2 Example

A simple application considered is a linear structural instrumental variable model with one exogenous variable and two instruments:

$$y_i = x_i\beta + v_i, \tag{1.12}$$

$$x_i = \alpha_1 z_{1i} + \alpha_2 z_{2i} + u_i, \tag{1.13}$$

where x_i is an endogenous variable, u_i and v_i are unobserved disturbances. z_{1i} denotes for a valid instrument, z_{2i} indicates a "nearly" exogenous instrument. Our goal is to test for the structural parameter β in the case of model misspecification, which is specified in the form of local nonexogeneity, i.e., $E(z_{2i}v_i) = \delta/\sqrt{n}$. The moment conditions thus can be obtained as:

$$g_i = \begin{bmatrix} z_{1i}(y_i - x_i\beta) \\ z_{2i}(y_i - x_i\beta) - \delta/\sqrt{n} \end{bmatrix}.$$

With some calculations, we obtain

$$G_\beta = \begin{bmatrix} -\alpha_1\overline{z_1^2} - \alpha_2\overline{z_1 z_2} \\ -\alpha_1\overline{z_1 z_2} - \alpha_2\overline{z_2^2} \end{bmatrix}, \quad G_\delta = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \Omega = \sigma^2 \begin{bmatrix} \overline{z_1^2} & \overline{z_1 z_2} \\ \overline{z_1 z_2} & \overline{z_2^2} \end{bmatrix},$$

where $\overline{A}$ denotes for the sample average of A.

Furthermore,

$$\begin{aligned} \Sigma_\beta &= \frac{1}{\sigma^2(b_1 b_2 - b_{12}^2)}(\alpha_1^2 b_1^2 b_2 + \alpha_2^2 b_2^2 b_1 - \alpha_1^2 b_{12}^2 b_1 - \alpha_2^2 b_{12}^2 b_2 - 2\alpha_1\alpha_2 b_{12}^3 + 2\alpha_1\alpha_2 b_1 b_2 b_{12}), \\ \Sigma_{\beta\delta} &= \frac{\alpha_2}{\sigma^2}, \\ \Sigma_\delta &= \frac{1}{\sigma^2(1-\rho^2)b_2}, \end{aligned}$$

where $b_1 = \overline{z_1^2}$, $b_2 = \overline{z_2^2}$, $b_{12} = \overline{z_1 z_2}$, and $\rho^2 = \frac{b_{12}^2}{b_1 b_2}$.

An interesting observation from this example is if coefficient of the "poor" instrument z_{2i}, α_2, is statistically insignificant our adjusted score test can be equivalent to the standard score test as the adjustment in RS^* is merely zero. When instrument z_{2i} is not or weakly correlated to x_i, this over-identified structural model in (1.12)-(1.13) is closely the same as the model that is just-identified with a single instrument. As a result, RS test and RS^* test asymptotically reach the same limiting distribution, and have the same non-distorted size. Therefore one sufficient condition to avoid size distortion in the standard score test is to have the locally nonexogenous instruments to be non-influential on x.

By Corollary 1.3.3, we can compare the possible power loss when the model is correctly specified, i.e., $\delta/\sqrt{n} = 0$. Unless $\alpha_2 \neq 0$ and $\rho \neq 1$, RS^* has no loss in power relative to RS, and our adjusted score test and the standard score test are generally equivalent.

1.3.3 Generalized Model

In this section we will propose a robust score test under general moment conditions. We define moment conditions as following:

$$E[g(x_i, \theta_{10}, \theta_{20}, \theta_{30})] = 0 \tag{1.14}$$

where θ_1 is a parameter needs to be estimated, θ_2 is a parameter that we want to test i.e $\theta_2 = \theta_{20}$, and θ_3 is potentially locally misspecified i.e $\theta_{30} = C/\sqrt{n}$. Instrument variable model is a special case of this general model. The size-robust score test is formulated as following:

$$RS^* = n\hat{D}^{*\prime}(\hat{\Sigma}_{22} - \hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{32})^{-1}\hat{D}^* \tag{1.15}$$

where $\hat{D}^* = \hat{D}_{\theta_2} - \hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}\hat{D}_{\theta_3}$, $\Sigma_{22} = G_2'\Omega^{-1}G_2 - G_2'\Omega^{-1}G_1(G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1}G_2$, $\Sigma_{23} = G_2'\Omega^{-1}G_3 - G_2'\Omega^{-1}G_1(G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1}G_3$, $\Sigma_{32} = G_3'\Omega^{-1}G_2 - G_3'\Omega^{-1}G_1(G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1}G_2$, and $\Sigma_{33} = G_3'\Omega^{-1}G_3 - G_3'\Omega^{-1}G_1(G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1}G_3$. $\hat{A}$ is a consistent estimator of A.

Theorem 1.3.4 *Given $\theta_3 = C/\sqrt{n}$, under the null hypothesis $H_0 : \theta_2 = \theta_{20}$, the adjusted score test statistic RS^* converges to a central chi-squared distribution with degrees of freedom K:*

$$RS^* \xrightarrow{d} \chi^2_K(0). \tag{1.16}$$

1.4 Monte Carlo Simulation

This simulation experiment is designed for a linear structural model with only one endogenous variable x_i:

$$\begin{aligned} y_i &= x_i\beta + u_i, & (1.17)\\ x_i &= 0.5z_{vi} + 0.5z_{di} + v_i. & (1.18) \end{aligned}$$

For the purpose of identification, two instrumental variables are needed: z_v denotes for a valid instrument and z_d denotes for an invalid (locally nonexogenous) instrument.

We sample z_v from a chi-squared distribution with degrees of freedom 2. The error terms, u_i and v_i, and the invalid instrument z_{di} are generated by a multivariate normal distribution:

$$(u_i, v_i, z_{di}) \sim \mathcal{MN}(0, \Sigma),$$

with

$$\Sigma = \begin{pmatrix} 1 & \varphi & \delta \\ \varphi & 1 & 0 \\ \delta & 0 & 1 \end{pmatrix},$$

where φ controls for the level of endogeneity of x_i, and δ indicates the local nonexogeneity of z_{di}. We set $\varphi = 0.6$. δ is chosen from a set of values $(0, 0.1, 0.15, 0.2)$, where 0 indicates a valid instrument. Each experiment is simulated 1000 times with a sample size chosen from $(200, 500, 1000)$.

For the sake of comparison, we consider six types of tests: adjusted score tests and standard score tests estimated by EL, ET, and CUE algorithm, respectively. Although previous studies on GEL method have pointed out the asymptotic equivalency among EL, ET, and CUE, we should not be surprised to finite-sample differences in sizes and powers obtained by different GEL algorithms. In particular, when sample size is small, for example, 100 or 200, simulation results have shown quite different finite-sample sizes and powers in EL, ET, and CUE-based tests.

By the result in Table 2-4, we compare the relative sizes of our adjusted score tests and the standard score tests. When sample size is small as shown in Table 2, adjusted score test, RS^*, shows little improvement in reducing size distortion caused by the failure of exogeneity conditions.[1] However, in Table 3-4, as sample size increases, size of RS^* significantly shrinks toward the true level while size of RS deviates even far away from it. This pattern of divergence in RS is consistent with our discussion in previous sections that the standard score test fails to control for the size distortion caused by local misspecification in the model.

[1] In an effort to improve size of RS^* in small sample cases, we apply bootstrapping technique for $n = 100$ and 200. After bootstrapping for 5000 times, sizes of RS^* have been greatly reduced towards 0.05 although instruments are locally nonexogenous, while those of RS grow far away from 0.05 as parameter values of δ increases. The bootstrapping results are available upon request.

Based on Theorem 1.3.2, size of the adjusted score test eventually will converge to 0.05, not being affected by any local deviation of δ from 0, as a result of the fact that RS^* test is robust to local nonexogeneity conditions. Among three types of adjusted score tests, CUE-based test has the smallest size regardless of the sample size. Sizes of EL and ET are very close in general. EL-based test is slightly better than ET-based test in large sample cases while ET test performs better in small sample cases. It is worth noting that CUE-based test - either adjusted or unadjusted - pays a price for its excessive under-rejection of the null such that it persistently stays low in power, particularly when sample size is small. Although this is not an unusual trade-off between size and power in statistical inference, we still should be cautious when the CUE-based score test is applied. Kleibergen (2005) provides one possible explanation for such an unexpected small power associated to CUE-based score test that test statistics of these tests are equivalent to the first-order derivative of the GMM objective function thereby spurious results are generated around the value of θ where the objective function reaches its maximal or is at an inflection point. Kitamura (2001) have also mentioned that weighting matrix in the objective function of continuous-updating GMM is likely to be inflated at values of β which are far away from its "true" value, thereby resulting in small value of the objective function and high probability of acceptance of the null hypothesis.

Table 5-16 list the comparison of RS^* and RS test in powers.[2] In the case of no local nonexogeneity, Table 5, 9, 13 show powers of RS^* and RS grow similarly approaching to one as β of the alternative deviates away from 0. The slight decline of RS^* in power, relative to RS, generally reflects the result in Corollary 1.3.3. When the parameter of local nonexogeneity is large, power of RS^* declines if sample size is small, while unsurprisingly power of RS is merely affected by the presence of non-zero δ. In contrast, in Table 3, 13 where sample sizes are large, RS^* test produces good power properties that are similar to RS test.

[2]All powers of RS^* test and RS test reported in this paper are size-adjusted.

1.5 Empirical Example: Democracy and Income

In this section, we re-explore the study on the causal relationship between democracy and income by Ancemoglu, et al. (2008). By examining their instrumental variable strategy, we discuss possible violations of the instruments used to identify the casual effect of countries' economic growth on democracy, and thereby try to provide another view of their story under potential misspecification of the model. Note that our goal is not to reconstruct the economic story in Ancemoglu, et al. (2008), but to investigate the model and data more carefully by examining the exogeneity conditions which are crucial in an instrumental variable model.

For the sake of simplicity, we focus on the comparison of our results to the results in Ancemoglu, et al. (2008) using the basic model studies in their paper:

$$Democracy_{i,t} = \alpha + \beta \cdot Income_{i,t-1} + \epsilon_t. \tag{1.19}$$

The instruments proposed, for example, for per capita income of country i at time $t-1$ are country i's past savings rates, $s_{i,t-2}$ and trade-weighted world income, $\hat{Y}_{i,t-1}$, which reflects trade linkages across countries.

In their paper, Ancemoglu, et al. (2008) find no cross-country correlation between income and democracy after controlling for the country fixed effects. This result is quite different from the conclusion in previous literature in which the result can be summarized as "*democratization came with growth*"[3]. Ancemoglu, et al. claim that such a spurious correlation between income and democracy can be a result when common factors that simultaneously affect both income and democracy, such as country fixed effects, are ignored in the model. Thereby by including country fixed effects in the model, they show the strong correlation between income and democracy disappears.

To identify the causal relationship between income and democracy, they use an instrumental variable model where income is treated as an endogenous variable.[4] To make robust

[3] Ancemoglu, et al. 2008, page 808.

[4] Although Ancemoglue, et al. (2008) have proposed two instruments for variable *Income*, they only use one each time for the regression of their model. In our example, we use both instruments for each of the regressions. Note that the estimation results of β using both instruments in one regression is not much different from the results in their paper.

inference on parameter β, two basic conditions need to be satisfied by the proposed instruments: (1) strong correlation between income and instruments; (2) exogeneity between instruments and ϵ. The first stage regression in their paper shows condition (1) is unlikely to be a problem in practice. For the second condition, the authors admit that they cannot verify it by providing a precise theory but argue the unlikely failure of excludability by making ad-hoc checks on modified models. Thereby it is still possible that one of the exclusion conditions is invalid such that it is correlated to democracy by some other connections. For example, it is difficult to rule out the possibility that past saving rate could be correlated with equilibrium political institutions, which in fact do have impact on democracy.

With this caveat in mind, we test for the significance of β using our adjusted score test, where the test result is shown to be robust to local nonexogeneity of instruments, and the standard score test, which suffers from size distortion as shown in our theory. In this example, we treat past saving rates as an invalid instrument and trade-weighted world income as a valid instrument. The results in Table **??** show a stronger relationship between income and democracy by RS^* test than the results by RS test as test statistic of RS^* is much larger than that of RS. This naturally raises our concerns about the possibility on parameter β that it would remain significant after controlling for the fixed country effects. If this is true, Ancemoglu, et al. (2008) may need to provide stronger evidences to claim that no causal effect of income on democracy.

1.6 Conclusion

This paper proposes an adjusted GEL-based score test statistic for instrumental variable models. The limiting behavior of this test statistic have shown to be robust to local violation of exogeneity conditions, which often appears in empirical studies. Because it converges to a central chi-squared distribution regardless of the presence of nuisance parameters of local nonexogeneity, it produces asymptotically correct size under local misspecification of the model. However we do pay a price of slight reduction in power for such a robustification of size. Furthermore, by studying an example of the linear structural model, we have discussed a sufficient condition where our adjusted score test is asymptotically equivalent to the standard score test in the sense of asymptotic size and power. We also apply our newly proposed test to an empirical example in Ancemoglu, et al. (2008) where causal effect of income on democracy has been extensively explored. By examining the instruments used in their paper, we have shown that when one of their instruments is locally nonexogenous, pooled OLS regression renders a significant β which has much smaller p-value using the adjusted score test than the p-value calculated using the standard score test.

Chapter 2

Mis-specification-Robust Bootstrap for Empirical Likelihood Estimators

2.1 Introduction

Hansen's Generalized method of moment (GMM) (1982) has been widely used for applied economics. Hansen and Singleton (1982) used it to estimate asset pricing models; Christiano and Haan (1996) applied it for business cycle models; Ruge-Murcia (2007) found an application in stochastic dynamic general equilibrium models. Despite the popularity of the GMM method, the GMM estimator has poor finite sample performance. Altonji and Segal (1996) and Hansen, Heaton, and Yaron (1996) have both addressed this subject.

To improve the small sample properties of GMM, multiple alternative estimators have been proposed. The Empirical Likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994), the Exponential Tilting estimator of Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998), the Continuously Updating (CU) estimator of Hansen, Heaton, and Yaron (1996), and the Minimum Distance Estimator (MDE) of Kitamura, Ostu, and Evdokimov (2013). Additionally, other efforts have been made to approximate the small sample distribution of GMM more accurately. This includes bootstrap methods by Hahn (1996), Hall and Horowitz (1996), Andrews (2002), Brown and Newey (2002), Lee (2014), and Allen, Gregory, and Shimotsu (2011).

Although GEL estimators are favorable alternatives to GMM, there is little evidence that the finite sample performance of the GEL test statistics is well enough based on the first-order asymptotics. Guggenberger and Hahn (2005) and Guggenberger(2008) find that the first-order asymptotic approximation to El estimators may be poor. Tt is then natural to consider a bootstrap method to improve the finite sample performance. Brown and Newey(2002) first introduced the bootstrap method to GMM estimators, and Allen, Gregory, and Shimotsu

(2011) extended it to dependent data. And notably, Brown and Newey (2002) invented a method of bootstrapping for GMM that used implied empirical likelihood weights for resampling. They showed that the method achieves significant improvement in the Monte Carlo studies, however, few paper deal with GEL estimators. One important and quite related paper is Lee (2014) which proposes a bootstrap procedure for the GEL estimators. Lee showed that the bootstrap t test statistics achieves sharp asymptotic refinements.

Furthermore, the validity of inferences based on the GEL estimators depends on the model specification. Although model misspecification can be detected asymptotically by an overidentification test, one might not make the correct inference for a finite sample. It is still interesting to explore the properties under possible misspecification. More details can be found in Lee (2014) and Schennach (2007). Since no parameter could satisfy all the moment conditions simultaneously, a pseudo-true value would be defined. Such pseudo-true values are still the object of interest in some cases, i.e., Hellerstein and Imbens (1999), and Bravo (2010).

In this paper I focus on two points: bootstrapping with empirical likelihood weights and robustness of model specification. Lee (2014) also proposes a bootstrap procedure that deals with these two aspects. However, he uses equal weights in the bootstrap procedure because of possible misspecified models that would lose some efficiency if the model were actually correctly specified. In this paper I propose a new bootstrap method based on an overidentification J test to improve the efficiency by using the empirical likelihood weights but not the equal weights.

The plan of the paper is as follows. In Section 2, we briefly introduce the GEL method and model specification.Section 3 presents proof of first-order validity. In Section 4 show some Monte Carlo Studies, and the last section includes conclusion and possible extensions.

2.2 GMM Bootstrap Method

In this section we review the GMM bootstrap method in Brown and Newey (2002) using the same notations. Let z_i (i=1,...,n) be i.i.d observations. They satisfy the moment conditions with true parameter θ_0:

$$E(g(z,\theta_0)) = 0,$$

where $g(z,\theta)$ is an m $\times$ 1 vector of moment conditions and $m \geq p$.

Let $g_i(\theta) = g(z_i,\theta)$, $\hat{g}(\theta) = \sum_{i=1}^n g_i(\theta)/n$, and $\hat{\Omega} = \sum_{i=1}^n g_i(\theta)g_i(\theta)'/n$. A two-step GMM estimator is defined as

$$\hat{\theta} = arg \min_{\theta\in\Theta} \hat{g}(\theta)'\tilde{\Omega}^{-1}\hat{g}(\theta),$$

where $\tilde{\Omega}^{-1} = \hat{\Omega}(\tilde{\theta})$ with some preliminary GMM estimator $\tilde{\beta}$.

Definition 2.2.1 *(GMM-empirical likelihood Bootstrapping)*
We are interested in testing the null hypothesis $H_0 : \theta_k = \theta_{0,k}$.

1. *Calculate*

$$\begin{aligned} \hat{\pi}_i &= \frac{1}{n(1-\hat{\lambda}'\hat{g}_i)}, i = 1,..n, \\ \hat{\lambda} &= \arg\max_{\lambda'\hat{g}_i<1} \sum_{i=1}^n \ln(1-\lambda'\hat{g}_i). \end{aligned}$$

2. *Draw n iid observations* $z_1^b, z_2^b ... z_n^b$ *with replacement from* $z_1, z_2 ... z_n$ *using the distribution with* $Pr(z = z_i) = \hat{\pi}_i$.

3. *Calculate* $t^b = \frac{\hat{\theta}_k^b - \hat{\theta}_k}{\sqrt{\hat{V}_{kk}^b/n}}$.

4. *Repeat steps 2 and 3 B times, to obtain* $t^1, t^2 ... t^B$, *and use the empirical distribution of* $t^1, t^2 ... t^B$ *to compute the critical values.*

This method differs from standard GMM bootstrap in the use of empirical likelihood rather than equal weights in step 2. As it's shown in Brown and Newey (2002) this method is asymptotically efficient, achieving the semi parametric efficiency bound of Brown and Newey (1998) for estimators of the cdf under the moment restrictions.

However, there are two aspects that could be extended to this method. Firstly, as it's shown in Newey and Smith (2004), the EL estimator enjoys better theoretical property than the GMM estimator, so we could use the EL estimator in the first step. 2. Although overidentification test can eventually detect the moment misspecification, it's still useful to utilize the pseudo-true values when there is no parameter that can satisfy all the moment conditions at the same time. In next section we will discuss the method in Lee (2014) and modify it to integrate the empirical likelihood weights.

2.3 GEL estimators And Model Specification

We first review the model structure of Lee (2004). Let z_i (i=1,...,n) be i.i.d observations. They satisfy the moment conditions with true parameter θ_0:

$$E(g(z, \theta_0)) = 0,$$

where $g(z, \theta)$ is an m $\times$ 1 vector of moment conditions where $\theta \in \Theta \subset \mathbb{R}^p$, and $m \geq p$. We follow the same notations as in Lee (2014). Let $G^{(j)}(Z_i, \theta)$ denote the partial derivatives with respect to θ of order j. For instance, $G^{(1)}(Z_i, \theta) \equiv G(Z_i, \theta) \equiv (\partial/\partial\theta')g(Z_i, \theta)$ and $G^{(2)}(Z_i, \theta) \equiv (\partial/\partial\theta')vec\{G(Z_i, \theta)\}$. To simplify the notation, we let $g_i(\theta) = g(Z_i, \theta)$, $G_i^j(\theta) = G^{(1)}(Z_i, \theta)$, $\hat{g}_i = g(Z_i, \hat{\theta})$, and $\hat{G}_i^{(j)} = G^{(j)}(Z_i, \hat{\theta})$, where $\hat{\theta}$ is the EL estimator.

One alternative estimation to GMM is generalized empirical likelihood (GEL). We follow the definition from Guggenberger and Smith (2011):

Definition 2.3.1 *(GEL estimation). The GEL estimator $\hat{\theta}$ for θ is defined as (Guggenberger and Smith (2011)):*

$$\widehat{\theta} := arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' g_i(\theta))], \tag{2.1}$$

where Θ is a compact subset of $\mathbb{R}^{k+\ell}$, $\hat{\Lambda}_n(\theta) = \{\lambda \in \mathbb{R}^q : \lambda' g_i(\theta) \in \mathbb{Q}$ for $i = 1, ..., n\}$, , $\mathbb{Q}$ is an open interval of the real line containing 0, and the real-valued function $\rho(.) : \mathbb{Q} \to \mathbb{R}$ is strictly concave on its domain.

Assumption 2.3.1 *(a) The function $\rho(v)$ is twice continuously differentiable in a neighborhood of 0. (b) $\rho_1 = \rho_2 = -1$, where we define $\rho_j(v) = \partial^j \rho(v)/\partial v^j$ with $\rho_j = \rho_j(0)$ for any* *nonnegative integer j.*

The three most used GEL estimators in the literature are the empirical likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994), the exponential tilting (ET) estimator of Kitamura and Stutzer (1997), and the continuous-updating estimator (CUE) of Hansen, Heaton and Yaron (1996) which corresponds to $\rho(\nu) = \ln(1-\nu)$, $\rho(\nu) = -\exp(\nu)$, and $\rho(\nu) = -(1+\nu)^2/2$, respectively. See Parente and Smith (2014) for a recent survey on GEL methods.

The first-order conditions of equation (1) are

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_i)\hat{G}_i{}'\hat{\lambda} = 0,$$

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i = 0.$$

The model is correctly specified if there is a unique θ_0 satisfy $Eg(z_i, \theta_0) = 0$. If the model is misspecified then there is no θ_0 that can satisfy $Eg(z_i, \mu_0) = 0$ simultaneously under the overidentified case. Following Schennach(2007) we define the pseudo-true values $\beta_0 = (\theta_0', \lambda_0')$ solving the population version of the FOCs

$$E\rho_1(\lambda_0' g_{i0})G_{i0}'\lambda_0 = 0,$$

$$E\rho_1(\lambda_0' g_{i0})g_{i0} = 0.$$

These FOC conditions hold regardless of model specification so these pseudo-true values also hold even if the model is not correctly specified. For the EL estimator, Chen, Hong, and Shum (2007) provide regularity conditions for $\sqrt{n}$ consistency and asymptotic normality under misspecification. Particularly, they assume that the moment function is uniformly bounded:

$$\sup_{\theta\in\Theta, z\in Z} ||g(z,\theta)|| < \infty, \inf_{\theta\in\Theta,\lambda\in\Lambda(\theta),z\in Z}(1-\lambda' g(z,\theta)) > 0,$$

under some regularity conditions we could have the asymptotic property of the pseudo-true estimators as follows:

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, \Gamma^{-1}\Psi(\Gamma')^{-1}) \tag{2.2}$$

where $\Gamma = E(\partial/\partial\theta')\psi(x_i, \theta_0)$ and $\Psi = E(\psi(x_i, \theta_0)\psi(x_i, \theta_0)')$. and,

$$\psi(x_i, \theta) = \begin{pmatrix} -\frac{1}{(1-\lambda' g_i(\beta))} G_i(\beta)'\lambda \\ -\frac{1}{(1-\lambda' g_i(\beta))} g_i(\beta) \end{pmatrix}.$$

Γ and Ψ can be estimated by $\hat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n} \partial\psi(x_i, \hat{\theta})/\partial\theta'$ and$\hat{\Psi} = \frac{1}{n}\sum_{i=1}^{n} \psi(x_i, \hat{\theta})\psi(x_i, \hat{\theta})'$.

The upper left submatrix of $\Gamma^{-1}\Psi(\Gamma')^{-1}$ is the asymptotic variance matrix of $\sqrt{n}(\hat{\theta}-\theta_0)$, which could be used for a testing problem.

2.4 Bootstrapping on testing

Let $\hat{\theta}$ be the EL estimator and $\hat{\Sigma}$ be the corresponding variance matrix estimator. Let θ_k denote the kth element of θ, and $\hat{\Sigma}_k$ is the kth diagonal element of $\hat{\Sigma}$. The t statistic for testing the null hypothesis $H_0 : \theta_k = \theta_{0,k}$ is

$$T = \frac{\hat{\theta}_k - \theta_{0,k}}{\sqrt{\hat{\Sigma}_k/n}}, \tag{2.3}$$

which has an asymptotic N(0,1) distribution under H_0 without assuming the correct models. However, applying this asymptotic normal distribution as a reference distribution does not work well in the finite sample. Please see the Monte Carlo Studies in Lee (2014). He proposed a nonparametric bootstrap method by resampling $Z_1^*, ..., Z_n^*$ randomly with replacement from the sample $Z_1, ..., Z_n$. Let $\hat{\Sigma}^*$ be the upper left $p \times p$ submatrix of the bootstrap version of the covariance matrix. Then the bootstrap t test statistics is defined as

$$T^* = \frac{\hat{\theta}_k^* - \hat{\theta}_k}{\sqrt{\hat{\Sigma}_r^*/n}}.$$

It is natural to implement the bootstrap method with EL weights when we resample the observations; however, as shown in Lee(2014), the cdf estimators based on such weights would be inconsistent for the true cdf if the model is misspecified because $\sum_i p_i g(z_i, \hat{\theta}) = 0$ holds even for large sample, while $Eg(z_i, \theta_0) \neq 0$. Hence Lee (2014) proposed to using equal weights in his paper since the empirical distribution function is always consistent. However, we will lose some efficiency if the moment condition is correctly specified as shown in Brown and Newey (2002). In this paper we extend Lee's method to use the empirical likelihood weights.

Definition 2.4.1 *(Adapted Bootstrap Weights)*

$$w_i = \hat{p}_i \mathbb{1}(\widehat{OI} < \chi^2_{(m-1,1-\alpha)}) + \frac{1}{n}\mathbb{1}(\widehat{OI} \geq \chi^2_{(m-1,1-\alpha)}) \tag{2.4}$$

where $\hat{p}_i$ is the EL implied weights, $\widehat{OI} = 2[nln(1/n) - \sum_{i=1}^{n} ln\hat{p}_i]$ is the overidentification test statistics, α is the size of the overidentification test, and l is the number of moment conditions.

The intuition is to utilize the EL weights when we don't have strong evidence to reject the overidentification test and only equal weights when we have strong evidence to reject the overidentification test.

2.5 Main Results

We rely extensively on the results of Lee (2014). P^* is the probability distribution of the bootstrap sample.

Assumption 2.5.1

$$Z_i, i = 1...n \quad are \quad i.i.d.$$

Assumption 2.5.2

1. *Γ is nonsingular and Ψ is positive definite.*

2. $g(z, \theta)$ *is d+1 times differentiable with respect to* θ *in the neighborhood of* θ_0*, for all* z *in the domain.*

3. *There is a function C(Z) such that* $||G^{(j)}(z, \theta) - G^{(j)}|| < C(z)||\theta - \theta_0||$ *for all* z *in the domain and all* θ *in the domain.*

4. *There is a function* $C_2(Z)$ *such that* $|\rho_j(\lambda' g(z, \theta)) - \rho_j(\lambda_0' g(z, \theta_0))| \leq C_2(z)||(\theta', \lambda') - (\theta_0', \lambda_0')||$ *for all* z *in the domain and* θ *in the domain;* $EC_2(Z) < \infty$*.*

Assumption 2.5.3

1. Θ *is compact and* θ_0 *is an interior point of* Θ*;* $\Lambda(\theta)$ *is a compact and* $\lambda(\theta)$ *contains a zero vector.*

2. $(\hat{\theta}, \hat{\lambda})$ *is the EL estimator;* (θ_0, λ_0) *is the pseudo-true value that uniquely solves the FOC of population version.*

3. *For some Function* $C_3(Z)$ *such that* $||g(z, \theta_1 - g(z, \theta_2)|| < C_3(Z)||\theta_1 - \theta_2||$ *for all* z *in the domain and* θ *in the domain;* $EC_3(Z) < \infty$*.*

4. *For some Function* $C_4(Z)$ *such that* $||\rho(\lambda_1' g(z, \theta_1)) - \rho(\lambda_2' g(z, \theta_{12}))|| < C_4(Z)||(\theta_1', \lambda_1') - (\theta_2', \lambda_2')||$ *for all* x *in the domain and* θ *in the domain;* $EC_4(Z) < \infty$*.*

Lemma 2.5.1 *Suppose Assumptions 5.1, 5.2, 5.3, and the uniformly bounded condition hold. Then* $\hat{\lambda} \xrightarrow{p} 0$*, and* $\hat{\theta} - \theta_0 \xrightarrow{p} 0$*.*

Lemma 2.5.2 *Suppose Assumptions 5.1, 5.2, 5.3, and the uniformly bounded condition hold. Then For any* $\epsilon > 0$ *and* $\delta > 0$*,*

$$\lim_{n \to \infty} P[P^*[|\theta^* - \hat{\theta}| > \epsilon] > \delta] = 0$$

.

Theorem 2.5.1 *Under Assumptions 5.1, 5.2, 5.3, and the uniformly bounded condition holds.,*

$$\sup_{z \in \mathbb{R}} |P^*(T^* \leq z) - P(T \leq z)| \xrightarrow{p} 0. \tag{2.5}$$

2.6 Simulation Study

In this section, multiple Monte Carlo Studies compare the finite sample performance of using different reference distributions under correct specification and misspecification. We use the same warp-speed Monte Carlo method of Lee (2014). The Warp-speed method only draws one bootstrap sample for each Monte Carlo repetition rather than B times, allowing for a significant computation advantage. The number of Monte Carlo repetition is 5,000. We compare 4 tests:

1. "Asymp"-Use normal distribution without bootstrapping
2. "Boot-equal"-Use equal weights in the bootstrap process
3. "Boot-EL"-Use EL weights in the bootstrap process
4. "Boot-Adapted"-Use adapted EL weights in the bootstrap process

2.6.1 Correct Model Specification

Consider the same AR(1) dynamic panel model of Lee (2014). Suppose $i = 1, 2, ..n$ and $t = 1, 2.., 4$, and the DGP is:

$$y_{it} = \rho_0 y_{i,t-1} + \eta_i + \nu_{it}, \tag{2.6}$$

where $\eta_i \sim N(0,1)$; $\nu_{it} \sim \frac{\chi_1^2 - 1}{\sqrt{2}}$; $y_{i1} = \eta_i/(1-\rho_0) + \mu_{i1}$; $\mu_{i1} \sim N(0, 1/(1-\rho_0^2))$, and the moments that are used:

$$\begin{aligned}
Ey_{i2}(\triangle y_{i4} - \rho_0 \triangle y_{i3}) &= 0 && (2.7)\\
Ey_{i1}(\triangle y_{i4} - \rho_0 \triangle y_{i3}) &= 0 && (2.8)\\
Ey_{i1}(\triangle y_{i3} - \rho_0 \triangle y_{i2}) &= 0 && (2.9)\\
E\triangle y_{i3}(y_{i4} - \rho_0 y_{i3}) &= 0 && (2.10)\\
E\triangle y_{i2}(y_{i3} - \rho_0 y_{i2}) &= 0 && (2.11)
\end{aligned}$$

Moment conditions (6), (7), and (8) are derived from taking the differences of (5), and the lagged values of y_{it} are used as instruments. Moment conditions (9) and (10) are using lagged values as the instruments. n=100, 200 are considered. The size of t test is 0.1. ρ_0 is chosen from (0.4, 0.9), where the latter represents the near unit root process. α is the size of overidentification test, and we choose from (0.05, 0.1, 0.2).

Table 2-7 show the rejection rates under different scenarios. We observe that all three bootstrap methods outperform the method using normal distribution , and if the moment conditions are correctly specified, Boot-EL has the best performance and Boot-Adapted performs better than Boot-equal. If the size of the J-test is smaller, Boot-adapted is closer to Boot-EL since it has less chance to reject the null hypothesis when moment conditions are actually correctly specified.

2.6.2 Mis-specified Moment Condition

We consider the same misspecified model of Lee (2014) in which the DGP follows an AR(2) while the model is based on the AR(1) specification. Suppose $i = 1, 2, ..n$ and $t = 1, 2.., 4$, and the true DGP is:

$$y_{it} = \rho_1 y_{i,t-1} + \rho_2 y_{i,t-2} + \eta_i + \nu_{it}, \tag{2.12}$$

where $\eta_i \sim trN(0,1)$; $\nu_{it} \sim \frac{tr\chi_1^2 - 1}{\sqrt{2}}$; $y_{i1} = \eta_i/(1-\rho_1-\rho_2)+\mu_{i1}$; $\mu_{i1} \sim trN(0,1)$; $y_{i2} \sim trN(0,1)$, and trN(0,1) and $tr\chi_1^2$ are truncated standard normal between -4 and 4, and truncated chi-square between 0 and 16. The moments are still from previous AR(1) model as:

$$\begin{aligned}
Ey_{i2}(\triangle y_{i4} - \rho_0 \triangle y_{i3}) &= 0 && (2.13)\\
Ey_{i1}(\triangle y_{i4} - \rho_0 \triangle y_{i3}) &= 0 && (2.14)\\
Ey_{i1}(\triangle y_{i3} - \rho_0 \triangle y_{i2}) &= 0 && (2.15)\\
E\triangle y_{i3}(y_{i4} - \rho_0 y_{i3}) &= 0 && (2.16)\\
E\triangle y_{i2}(y_{i3} - \rho_0 y_{i2}) &= 0 && (2.17)
\end{aligned}$$

In this model there is no ρ_0 that could satisfy all the 5 moments simultaneously. Interestingly, four of the moment conditions identify $\rho_a = \rho_1 - \rho_2$ and the other identifies $\rho_b = \rho_1 + \frac{\rho_2}{\rho_1 - \rho_2}$. The pseudo-true value ρ_0 is defined as $\rho_0 = w\rho_a + (1-w)\rho_b$ where w is between 0 and 1.n=100, 200, and 500 are considered. The size of t test is 0.1. Two sets of rho_1 and rho_2 are chosen. $\rho_1 = 0.6, \rho_2 = 0.2$, and $\rho_1 = 0.3, \rho_2 = 0.4$. The pseudo-true value is simulated using sample size n=30,000. The size of overidentification is chosen from (0.05, 0.1, 0.2).

Tables 7-12 show the rejection rates under misspecified moment conditions. We find that three Bootstrap methods still work better than using the normal distribution. In addition, the rejection rates of Boot-EL start to diverge as n goes large, which is consistent with the theoretical findings.

2.7 Conclusion

We propose an adapted bootstrap procedure for the empirical likelihood estimators. This method extends the bootstrap method in Lee (2014) by using the empirical likelihood weights, which improves efficiency if the moment condition model is correctly specified. Simulation Studies support the findings.

Chapter 3

Higher Order MSE Comparisons of Generalized Empirical Likelihood Estimators

3.1 Introduction

This paper calculates asymptotic higher-order mean squared error (MSE) for a simple linear model with only two moment conditions with a univariate parameter. This is done without assuming that generalized third moments of moment conditions are 0. We find that the Empirical Likelihood (EL) estimator no longer has the least higher-order MSE in the GEL family, and we also propose a data-driven GEL estimator that could minimize the higher-order MSE. This paper is motivated by two questions: (i) Does the EL estimator still has the least asymptotic higher-order MSE? and (ii) If not, can we find a favorable GEL estimator to minimize the higher order MSE?

Traditionally, generalized method of moments (GMM) estimators of Hansen (1982) have been invented to estimate method of moments. It is known that two-step GMM estimators might not yield good small sample performances. Hansen, Heaton, and Yaron (1996) provided a good explanation on this matter. To improve the small sample properties of GMM, multiple alternative estimators have been proposed. Empirical likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994), the exponential tilting estimator of Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998), the continuously updating (CU) estimator of Hansen, Heaton, and Yaron (1996), and the minimum distance estimator (MDE) of Kitamura, Ostu, and Evdokimov (2013). Additionally, other efforts have been made to approximate the small sample distribution of GMM more accurately. This includes bootstrap methods by Hahn (1996), Hall and Horowitz (1996), Andrews (2002), Brown and Newey (2002), Lee (2014), and Allen, Gregory, and Shimotsu (2011).

Newey and Smith (2004) developed a family of Generalized empirical (GEL) estimators that includes EL, ET, and CUE estimators. They showed that GEL and GMM estimators have the same first order asymptotic distribution but different higher order asymptotic distributions. Anatolyev (2005) generalized this to a dependent data structure and found that EL still has the nice small bias property. They found that EL's asymptotic bias does not grow with the number of moment restrictions. Hence the finite sample bias would be smaller with many moment conditions than GMM and other GEL estimators.

Although the small bias property of EL is attractive, there are few papers talking about higher order MSE comparisons on the GEL estimator. Newey and Smith (2004) showed that after correcting bias and applying EL probabilities, the higher order variance is relatively smaller compared to other bias corrected estimators. However, it is also interesting to compare higher order MSE when bias is not corrected, which is more common for practitioners. Imbens and Spady (2005) calculated higher-order asymptotic bias and MSE of GEL estimators and assumed third moments of moment conditions to be zero. They found that all GEL estimators have equivalent higher order properties. Meanwhile, it is not straightforward to observe an EL estimator's advantage from finite sample simulations. Guggenberger (2008) did comprehensive Monte Carlo studies to compare the finite sample properties of GEL estimators and other instrumental variable (IV) estimators. He found no significant advantages of using an EL estimator and suggested using two-stage least square estimators, which are simpler to compute. Lee (2016) also argued that there is little evidence that GEL estimators have a better approximation of the finite sample distribution.

Hence there is a need to explore the higher order MSE properties of GEL estimators. In this paper we use a very simple moment condition structure to compute the higher order MSE of GEL estimators. Due to the simple structure, we do not need to assume that the estimator is bias corrected nor that third moment conditions are zero. I find that higher-order MSE depends on the third derivatives of the ρ function in GEL estimator, and hence the EL estimator might not yield the least higher order MSE. In the end I also derive a data-driven GEL estimator that minimizes the higher order MSE.

The plan of the paper is as follows. In section 2, we briefly introduce our GEL method and review the equivalence between GEL estimators and minimum discrepancy (MD) estimators . In section 3 we introduce our moment condition model and the stochastic expansions.

Section 4 presents the higher order MSE of GEL estimators and proposes a data-driven estimator minimizing the higher order MSE. Section 5 concludes and discusses possible extensions and future work.

3.2 The model and Estimators

We first review the model structure from Newey and Smith (2004). We use the same notations. Let z_i (i=1,...,n) be i.i.d observations. They satisfy the moment conditions with true parameter θ_0:

$$E(g(z,\theta_0)) = 0,$$

where $g(z,\theta)$ is an m $\times$ 1 vector of moment conditions and $m \geq p$.

One alternative estimation to GMM is the generalized empirical likelihood (GEL). We follow the definition from Guggenberger and Smith (2011):

Definition 3.2.1 *(GEL estimation). The GEL estimator* $\hat{\theta}$ *for* θ *is defined as (Guggenberger and Smith (2011)):*

$$\widehat{\theta} := arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' g_i(\theta))], \tag{3.1}$$

where Θ *is a compact subset of* $\mathbb{R}^{k+\ell}$, $\hat{\Lambda}_n(\theta) = \{\lambda \in \mathbb{R}^q : \lambda' g_i(\theta) \in \mathbb{Q}$ *for* $i = 1, ..., n\}$, , $\mathbb{Q}$ *is an open interval of the real line containing* 0*, and the real-valued function* $\rho(.) : \mathbb{Q} \to \mathbb{R}$ *is strictly concave on its domain.*

Assumption 3.2.1 *(a) The function* $\rho(v)$ *is twice continuously differentiable in a neighborhood of 0. (b)* $\rho_1 = \rho_2 = -1$*, where we define* $\rho_j(v) = \partial^j \rho(v)/\partial v^j$ *with* $\rho_j = \rho_j(0)$ *for any nonnegative integer* j*.*

The three most used GEL estimators in the literature are the empirical likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994), the exponential tilting (ET) estimator of Kitamura and Stutzer (1997), and the continuous-updating estimator (CUE) of Hansen, Heaton and Yaron (1996) which correspond to $\rho(\nu) = \ln(1-\nu)$, $\rho(\nu) = -\exp(\nu)$, and

$\rho(\nu) = -(1+\nu)^2/2$, respectively. See Parente and Smith (2014) for a recent survey on GEL methods.

Newey and Smith (2004) presented the dual relationship between GEL and the minimum discrepancy (MD) estimator. The MD estimation is formulated as follows:

$$\bar{\theta} := arg \min_{\theta \in \Theta} \sum_{i=1}^{n} [h(\pi_i)],$$

subject to:

$$\sum_{i=1}^{n} \pi_i g_i(\theta) = 0,$$

$$\sum_{i=1}^{n} \pi_i = 1.$$

For each GEL estimator there is a dual MD estimator of a member in Cressie and Read (1984) family, where $h(\pi) = [\gamma(\gamma+1)]^{-1}[(n\pi)^{\gamma+1} - 1]/n$. They prove that the GEL and MD estimators are equivalent under the following relationship:

$$\rho(v) = -(1+\gamma v)^{(\gamma+1)/\gamma}/(\gamma+1).$$

The CR family estimator has been widely used in statistics and econometrics. Hence it's also interesting to explore what is the " best " CR family estimator.

3.2.1 Stochastic Expansion

In this section we will repeat the stochastic expansion of Newey and Smith (2004), which is used as the basis for our higher-order MSE calculations. They find the stochastic expansions for GEL estimator as below:

$$\sqrt{n}(\hat{\theta} - \theta_0) = \tilde{\psi} + Q_1(\tilde{\psi}, \tilde{a}, F_0)/\sqrt{n} + Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}, F_0)/n + R_n,$$

The details of this expansion can be found in the Appendix. Applying this expansion

Newey and Smith (2004) showed that the asymptotic higher order bias is given by

$$Bias(\hat{\theta}) = E[Q_1(\psi_i, a_i, F_0)/n],$$

and they got the higher order bias expressions for GMM and GEL estimators

$$\begin{aligned} Bias(\hat{\theta}_{GMM}) &= B_I + B_G + B_\Omega + B_W, \\ Bias(\hat{\theta}_{GEL}) &= B_I + (1 + \rho_3/2)B_\Omega, \end{aligned}$$

where $B_I = H(-a + E[G_i H g_i])/n$, $B_\Omega = -\Sigma E[G_i' P g_i]/n$, $B_\Omega = HE[g_i g_i' P g_i]/n$, and $B_W = -H \sum_{i=1}^{p} \bar{\Omega}_{\beta_j}(H_W - H)' e_j/n$. Especially, $Bias(\hat{\theta}_{EL}) = B_I$. This shows their conclusion that EL has the preferred higher order bias property.

3.2.2 Higher Order MSE Calculation

The model we consider here is to estimate the mean of a bivariate model. To describe it, suppose

$$X = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

, and $\{X_i\}_{i=1}^n$ are i.i.d samples from some unknown distribution F. From previous knowledge we know that the model has a true parameter μ_0 satisfying the moment conditions:

$$E(z_1) = \mu_0 \tag{3.2}$$

$$E(z_2) = \mu_0 \tag{3.3}$$

An important estimator of μ is the two-step GMM estimator of Hansen(1982). The alternatives to GMM we consider here are generalized empirical likelihood(GEL) estimators, as in Smith(1997). To describe GEL let $\rho(v)$ be a function of a scalar v that is concave on its domain. The estimator is the solution to a saddle point problem of (1), where $g_i = (x_i - \mu)$. The EL estimator is a special case with $\rho(v) = \ln(1 - v)$. The exponential tilting estimator is a special case with $\rho(v) = -e^v$, and the continuous updating estimator(CUE) is a special case with $\rho(v) = -(1+v)^2/2$.

We follow the work of Newey and Smith(2004) to derive the expansion equation. Here are some notations that are needed in the later expressions. Without loss of generality, we assume $E(z_{1i}-\mu)^2 = E(z_{2i}-\mu)^2 = 1$. Let $a_{1n} = \frac{\sum_{i=1}^{n}(z_{1i}-\mu)}{n}$, $a_{2n} = \frac{\sum_{i=1}^{n}(z_{2i}-\mu)}{n}$, $c_{1n} = \sum_{i=1}^{n}(z_{1i}-\mu)^2/n - 1$, $c_{2n} = \sum_{i=1}^{n}(z_{2i}-\mu)^2/n - 1$, $b_n = \sum_{i=1}^{n}(z_{1i}-\mu)(z_{2i}-\mu)/n$, $d_{1n} = \sum_{i=1}^{n}(z_{1i}-\mu)^2(z_{2i}-\mu)/n$ $d_{2n} = \sum_{i=1}^{n}(z_{2i}-\mu)^2(z_{1i}-\mu)/n$, $e_{1n} = \sum_{i=1}^{n}(z_{1i}-\mu)^3/n - \mu_3$, and $e_{2n} = \sum_{i=1}^{n}(z_{2i}-\mu)^3/n - \mu_3$.

Theorem 3.2.1 *Following Lemma A4 of Newey and Smith(2004) with Assumption 1 and 2 in the Appendix, then,*

$$\begin{aligned}
\hat{\mu} - \mu_0 &= 1/2(a_{1n}+a_{2n}) + 1/4(c_{1n}-c_{2n})(a_{2n}-a_{1n}) - 1/8\rho_3\mu_3(a_{2n}-a_{1n})^2 \\
&\quad + 1/8(c_{1n}-c_{2n})(c_{1n}+c_{2n}-2b_n)(a_{1n}-a_{2n}) + 1/8\rho_3\mu_3(a_{2n}-a_{1n})^2(c_{1n}+c_{2n}-2b_n)) \\
&\quad + 1/8(a_{2n}-a_{1n})^2(a_{1n}+a_{2n}+1/16\rho_3(d_{1n}+d_{2n}-e_{1n}-e_{2n})(a_{1n}-a_{2n})^2 \\
&\quad + 1/8\rho_3(a_{1n}-a_{2n})^2(a_{1n}+a_{2n}) + O_p(n^{-2}).
\end{aligned}$$

Corollary 3.2.2 *Following Theorem 1 then,*

$$E(\hat{\mu}-\mu_0)^2 = C + \frac{3}{16}\frac{\rho_3^2\mu_3^2}{n^2} + \frac{\rho_3\mu_3^2}{n^2} + \frac{3}{4}\frac{\rho_3}{n^2} - \frac{1}{4}\frac{\rho_3\mu_4}{n^2} + o(n^{-2}) \tag{3.4}$$

Remark 3.2.1 *With some simple calculations we get the higher order bias of the GEL estimator as follows:*

$$E(\hat{\mu}-\mu_0) = -\frac{\mu_3}{4n}(2+\rho_3) + o(n^{-1}) \tag{3.5}$$

When the EL estimator, i.e, $\rho_3 = -2$, the first term of higher-order bias vanishes. This result is consistent with the result in Newey and Smith(2004).

Remark 3.2.2 *From the higher-order MSE expansion we observe that if $\mu_3 = 0$, the leading term vanishes, and the minimum point of MSE would not exist. All the GEL estimators would have the same higher order MSE, and this result is consistent with the findings in Imbens and Spady (2005).*

Remark 3.2.3 *If* $\mu_3 \neq 0$ *when* $\hat{\rho}_3 = -3/8 + \frac{2\mu_4 - 6}{3\mu_3^2}$ *the higher-order MSE is minimized. Equivalently we get the MD estimator* $\hat{\gamma}$ *by* $\rho_3 = -(1-\gamma)$.

3.3 Conclusion and Future Work

In this paper we calculate the higher-order MSE of GEL estimators with a bivariate mean model. We find that the EL estimator no longer enjoys the smallest higher-order bias property in the criteria of MSE. From the expansion equation we observe that all GEL estimator would have the same higher order MSE if the third moment is 0. In addition we calculate a data-driven GEL estimators that minimizes the higher order MSE. In this paper we only explore the higher-order MSE property of the simple bivariate mean model, which significantly reduces the calculations. In the future more work could be conducted in more generalized settings.

Appendix A

Robust Inference for Instrumental Variable Models with Locally Non-exogenous Instruments

Proof of Lemma 1: Lemma 1 is a classical result, so we just give a sketchy proof. Under the null hypothesis, Lagrange multiplier λ in the GEL objective function can be consistently estimated as:

$$\hat{\lambda} = \sup_{\lambda \in \hat{\Lambda}_n(\hat{\theta})} \sum_{i=1}^{n} \frac{1}{n} [\rho(\lambda' g_i(\hat{\theta})]. \tag{A.1}$$

By first order conditions, we obtain:

$$\frac{1}{n} \sum_{i=1}^{n} \left[\rho_1(\hat{\lambda}' g_i(\hat{\theta})) g_i(\hat{\theta})\right] = 0. \tag{A.2}$$

Following Newey and Smith (2004) (proof of Theorem 3.2), we expand our first order conditions around $\lambda = 0$. This gives:

$$\sqrt{n}\overline{g} + \sqrt{n}\Omega\hat{\lambda} = o_p(1) \quad \Rightarrow \quad \sqrt{n}\hat{\lambda} \xrightarrow{d} N(0, \Omega^{-1}). \tag{A.3}$$

Since $\hat{D}_\beta = -G_1' \hat{\lambda} + o_p(1)$, we obtain:

$$n\hat{D}_\beta \hat{\Sigma}_1^{-1} \hat{D}_\beta' \xrightarrow{d} \chi_K^2(0).$$

Q.E.D.

To proof theorem 1 we need some lemmas. These lemmas extensively rely on Newey and Smith(2004).In particular, I use Lemma A1, A2, Theorem 3.1, and Theorem 3.2 of Newey and Smith(2004). We let $g_i(\theta) = z_i(y_i - f(x_i, \beta)) - \Delta = g_i(\beta) - \Delta$, $\theta_0 = (\beta_0, C_\delta/\sqrt{n})$ denotes the true parameter, and $\hat{\theta} = (\beta_0, 0)$.

Lemma 1 *If Assumption 1 is satisfied, then for any* $1/\alpha < \zeta < 1/2$ *and* $\Lambda_n = \{\lambda : ||\lambda|| \leq$

$n^{-\zeta}\}$, $\sup_{\beta\in B,\lambda\in\Lambda_n} |\lambda' g_i(\beta)| \xrightarrow{p} 0$.

This lemma is just the Lemma A1 of Newey and Smith(2004).

Lemma 2 *If assumption 1 is satisfied, and* $\overline{g}(\beta_0) = O_p(n^{-1/2})$,

then $\hat{\lambda} = \sup_{\lambda\in\hat{\Lambda}_n(\hat{\theta})} \sum_{i=1}^{n} \frac{1}{n}[\rho(\lambda' g_i(\beta_0)]$ *exists w.p.a.1, and* $\hat{\lambda} = O_p(n^{-1/2})$.

Proof. Let $g(\beta) = E[g(x,\beta)]$, and by uniform weak law of large numbers(UWL), $\sup_{\beta in B} ||\overline{g}(\beta) - g(\beta)|| \xrightarrow{p} 0$, and according to Assumption 1 $E(g(x,\beta_0)) = C_\delta/\sqrt{n}$. By triangle inequalities $\overline{g}(\beta_0) = O_p(n^{-1/2})$. Since $\rho(v)$ is twice continuously differentiable in a neighborhood of zero, then $\hat{\lambda} = \sup_{\lambda\in\hat{\Lambda}_n(\hat{\theta})} \sum_{i=1}^{n} \frac{1}{n}[\rho(\lambda' g_i(\beta_0)]$ exists w.p.a.1. Follow the same argument of Lemma A2 of Newey and Smith(2004) $||\hat{\lambda}|| = O_p(n^{-1/2})$.

Proof of Theorem 1: Let $\eta = (\lambda, \beta, \delta)$, and β_0 denote for the true value of β. Considering different values of η, we specify $\eta_0 = (0, \beta_0, C_\delta/\sqrt{n})$, $\eta* = (0, \beta_0, 0)$, and $\hat{\eta} = (\hat{\lambda}, \beta_0, 0)$, where $\hat{\lambda} = \arg\sup_{\lambda\in\hat{\Lambda}_n(\hat{\theta})} \frac{1}{n}\sum_{i=1}^{n}[\rho(\lambda' g_i(\hat{\theta})]$. Let$P(\eta) = \frac{1}{n}\sum_{i=1}^{n}[\rho(\lambda' g_i(\theta)]$.

Taylor expansion of the score function w.r.t. λ around η_0 gives

$$\sqrt{n}\frac{\partial P(\eta*)}{\partial\lambda} = \sqrt{n}\frac{\partial P(\eta_0)}{\partial\lambda} + \sqrt{n}\frac{\partial^2 P(\eta_0)}{\partial\lambda\partial\delta'}(0 - \frac{C_\delta}{\sqrt{n}}) = o_p(1), \tag{A.4}$$

$$\Rightarrow \sqrt{n}\frac{\partial P(\eta*)}{\partial\lambda} = -\sqrt{n}\overline{g} + G_2 C_\delta + o_p(1). \tag{A.5}$$

By definition, $\frac{\partial P(\hat{\eta})}{\partial\lambda} = 0$. Taylor expansion of the same score function around $\hat{\eta}$ gives:

$$\sqrt{n}\frac{\partial P(\eta*)}{\partial\lambda} = \sqrt{n}\frac{\partial P(\hat{\eta})}{\partial\lambda} + \sqrt{n}\frac{\partial^2 P(\hat{\eta})}{\partial\lambda\partial\lambda'}(0 - \hat{\lambda}) + o_p(1), \tag{A.6}$$

$$\Rightarrow \sqrt{n}\frac{\partial P(\eta*)}{\partial\lambda} = \sqrt{n}\Omega\hat{\lambda} + o_p(1). \tag{A.7}$$

From equation (A.20) and (A.22) we obtain

$$\sqrt{n}\hat{\lambda} = -\sqrt{n}\Omega^{-1}\overline{g} + \Omega^{-1}G_2 C_\delta + o_p(1). \tag{A.8}$$

Taylor expansion on the score function of β at $\hat{\eta}$ around value of η_0 gives:

$$
\begin{aligned}
\sqrt{n}\hat{D}_{\beta} &= \sqrt{n}\frac{\partial P(\hat{\eta})}{\partial \beta} && \text{(A.9)}\\
&= \sqrt{n}\frac{\partial P(\eta_0)}{\partial \beta} + \sqrt{n}\frac{\partial^2 P(\eta_0)}{\partial \beta \partial \lambda'}(\hat{\lambda} - 0) + \sqrt{n}\frac{\partial^2 P(\eta_0)}{\partial \beta \partial \delta'}(0 - C_{\delta}/\sqrt{n}) + o_p(1) && \text{(A.10)}\\
&= -\sqrt{n}G_1'\hat{\lambda} + o_p(1). && \text{(A.11)}
\end{aligned}
$$

By (A.8) and (A.11), it is easy to show

$$\sqrt{n}\hat{D}_{\beta} = \sqrt{n}G_1'\Omega^{-1}\overline{g} - G_1'\Omega^{-1}G_2C_{\delta} + o_p(1). \tag{A.12}$$

According to Lindberg-Levy CLT, therefore,

$$n\hat{D}_{\beta}\hat{\Sigma}_1^{-1}\hat{D}_{\beta}' \xrightarrow{d} \chi_K^2(\mu_2), \tag{A.13}$$

where $\mu_2 = C_{\delta}'\Sigma_{12}'\Sigma_1^{-1}\Sigma_{12}C_{\delta}$.

Q.E.D.

Proof of Lemma 2: Use the same argument of Theorem 1.3.1, we obtain

$$\sqrt{n}\hat{D}_{\delta} = -\sqrt{n}G_2'\hat{\lambda} + o_p(1) = \sqrt{n}G_2'\Omega^{-1}\overline{g} - G_2'\Omega^{-1}G_2C_{\delta} + o_p(1). \tag{A.14}$$

Q.E.D.

Proof of Theorem 2: By Theorem 1 and Lemma 2, we have

$$
\begin{aligned}
\sqrt{n}\hat{D}_{\beta_0} &= \sqrt{n}G_1'\Omega^{-1}\overline{g} - G_1'\Omega^{-1}G_2C_{\delta} + o_p(1), && \text{(A.15)}\\
\sqrt{n}\hat{D}_{\delta} &= \sqrt{n}G_2'\Omega^{-1}\overline{g} - G_2'\Omega^{-1}G_2C_{\delta} + o_p(1). && \text{(A.16)}
\end{aligned}
$$

$$\Rightarrow \hat{D}_{\beta} - \hat{\Sigma}_{12}\hat{\Sigma}_2^{-1}\hat{D}_{\delta} = (G_1'\Omega^{-1} - \Sigma_{12}\Sigma_2^{-1}G_2'\Omega^{-1})\overline{g} + o_p(1). \tag{A.17}$$

Therefore, we can show

$$\sqrt{n}(\hat{D}_{\beta} - \hat{\Sigma}_{12}\hat{\Sigma}_2^{-1}\hat{D}_{\delta}) \xrightarrow{d} N(0, \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'). \tag{A.18}$$

Q.E.D.

Proof of Theorem 3: The proof of Theorem 3 is quite similar to the proof of Theorem 1. The only difference is θ_1 needs to be estimated. Let $\eta = (\lambda, \theta_1, \theta_2, \theta_3)$. Considering different values of η, we specify $\eta_0 = (0, \theta_{10}, \theta_{20}, C/\sqrt{n})$, $\eta^* = (0, \theta_{10}, \theta_{20}, 0)$, and $\hat{\eta} = (\hat{\lambda}, \hat{\theta}_1, \theta_{20}, 0)$, where $\hat{\theta}_1 = \arg\min_{\theta_1 \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\hat{\theta})} \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' g_i(\theta_1, \theta_{20}, 0)]$. Let$P(\eta) = \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' g_i(\theta)]$.

Taylor expansion of the score function w.r.t. λ around η_0 gives

$$\sqrt{n}\frac{\partial P(\eta*)}{\partial \lambda} = \sqrt{n}\frac{\partial P(\eta_0)}{\partial \lambda} + \sqrt{n}\frac{\partial^2 P(\eta_0)}{\partial \lambda \partial \theta_{30}'}(0 - \frac{C}{\sqrt{n}}) + o_p(1), \tag{A.19}$$

$$\Rightarrow \sqrt{n}\frac{\partial P(\eta*)}{\partial \lambda} = -\sqrt{n}\overline{g} + G_3 C + o_p(1). \tag{A.20}$$

By definition, $\frac{\partial P(\hat{\eta})}{\partial \lambda} = 0$. Taylor expansion of the same score function around $\hat{\eta}$ gives:

$$\sqrt{n}\frac{\partial P(\eta*)}{\partial \lambda} = \sqrt{n}\frac{\partial P(\hat{\eta})}{\partial \lambda} + \sqrt{n}\frac{\partial^2 P(\hat{\eta})}{\partial \lambda \partial \lambda'}(0 - \hat{\lambda}) + \sqrt{n}\frac{\partial^2 P(\hat{\eta})}{\partial \lambda \partial \theta_1'}(\theta_{10} - \hat{\theta}_1) + o_p(1), \tag{A.21}$$

$$\Rightarrow \sqrt{n}\frac{\partial P(\eta*)}{\partial \lambda} = \sqrt{n}\Omega\hat{\lambda} - \sqrt{n}G_1(\theta_{10} - \hat{\theta}_1) + o_p(1). \tag{A.22}$$

From equation (A.20) and (A.22) we obtain

$$\sqrt{n}\Omega\hat{\lambda} = -\sqrt{n}\overline{g} + G_3 C_\delta - \sqrt{n}G_1(\hat{\theta}_1 - \theta_{10}) + o_p(1). \tag{A.23}$$

Taylor expansion of the score function w.r.t θ_1 around η_0 gives:

$$\sqrt{n}\frac{\partial P(\hat{\eta})}{\partial \theta_1} = \sqrt{n}\frac{\partial P(\eta_0)}{\partial \theta_1} + \sqrt{n}\frac{\partial^2 P(\eta_0)}{\partial \theta_1 \partial \lambda'}(\hat{\lambda} - 0) + \sqrt{n}\frac{\partial^2 P(\eta_0)}{\partial \theta_1 \partial \theta_1'}(\hat{\theta}_1 - \theta_1) \tag{A.24}$$

$$+\sqrt{n}\frac{\partial^2 P(\eta_0)}{\partial \theta_1 \partial \theta_3'}(0 - C/\sqrt{n}) + o_p(1), \tag{A.25}$$

$$\Rightarrow \sqrt{n}G_1'\hat{\lambda} = o_p(1) \tag{A.26}$$

By substituting equation(A.23) into equation(A.26) we have,

$$\sqrt{n}(\hat{\theta}_1 - \theta_{10}) = -(G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1}\sqrt{n}\overline{g} + (G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1}G_3 C + o_p(1) \tag{A.27}$$

Plugging equation(A.27) into equation(A.23) we will get,

$$\sqrt{n}\hat{\lambda} = -(\Omega^{-1} - \Omega^{-1}G_1(G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1})\sqrt{n}\overline{g}$$
$$+\Omega^{-1}(G_3C - G_1(G_1'\Omega^{-1}G_1)^{-1}G_1'\Omega^{-1}G_3C) + o_p(1)$$

Meanwhile,

$$\hat{D}^* = \hat{D}_{\theta_2} - \hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}\hat{D}_{\theta_3} = (\hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}G_3' - G_2')\sqrt{n}\hat{\lambda} + o_p(1) \tag{A.28}$$

Then according to Lindeberg-Levy CLT,

$$\sqrt{n}\hat{D}^* \xrightarrow{d} N(0, \Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32}) \tag{A.29}$$

Q.E.D.

Table A.1: Size: N=200

endogeneity	ELadj_size	ETadj_size	CUEadj_size	EL_size	ET_size	CUE_size	t_size
0	0.065	0.062	0.025	0.065	0.066	0.023	0.052
0.01	0.078	0.068	0.03	0.071	0.068	0.024	0.051
0.02	0.046	0.048	0.024	0.046	0.037	0.013	0.042
0.03	0.056	0.055	0.021	0.056	0.052	0.018	0.065
0.04	0.073	0.071	0.026	0.067	0.062	0.02	0.064
0.05	0.062	0.062	0.029	0.072	0.067	0.022	0.076
0.06	0.087	0.079	0.034	0.076	0.071	0.027	0.086
0.07	0.086	0.08	0.039	0.076	0.07	0.024	0.095
0.08	0.065	0.062	0.027	0.071	0.062	0.021	0.102
0.09	0.087	0.086	0.03	0.07	0.067	0.022	0.09
0.1	0.085	0.078	0.045	0.067	0.064	0.015	0.098
0.11	0.067	0.063	0.03	0.054	0.051	0.014	0.101
0.12	0.076	0.075	0.032	0.066	0.065	0.019	0.128
0.13	0.088	0.077	0.042	0.072	0.066	0.018	0.134
0.14	0.098	0.095	0.037	0.085	0.082	0.014	0.132
0.15	0.086	0.08	0.035	0.075	0.069	0.016	0.148
0.16	0.114	0.109	0.043	0.072	0.072	0.011	0.146
0.17	0.122	0.126	0.065	0.098	0.089	0.016	0.179
0.18	0.102	0.107	0.049	0.095	0.095	0.012	0.199
0.19	0.112	0.115	0.057	0.085	0.074	0.008	0.195
0.2	0.128	0.128	0.055	0.091	0.082	0.01	0.222

Table A.2: Size: N=500

endogeneity	adjusted EL	adjusted ET	adjusted CUE	EL	ET	CUE	t
0	0.044	0.043	0.033	0.04	0.041	0.03	0.043
0.01	0.052	0.054	0.039	0.051	0.053	0.039	0.054
0.02	0.076	0.075	0.056	0.073	0.072	0.052	0.071
0.03	0.047	0.045	0.027	0.051	0.047	0.034	0.06
0.04	0.044	0.046	0.038	0.055	0.059	0.039	0.07
0.05	0.055	0.059	0.044	0.071	0.073	0.054	0.08
0.06	0.057	0.062	0.041	0.063	0.068	0.049	0.083
0.07	0.051	0.051	0.034	0.076	0.078	0.056	0.101
0.08	0.062	0.062	0.041	0.085	0.083	0.051	0.111
0.09	0.051	0.055	0.037	0.08	0.084	0.056	0.115
0.1	0.063	0.073	0.046	0.093	0.1	0.065	0.129
0.11	0.066	0.06	0.037	0.111	0.108	0.074	0.163
0.12	0.057	0.061	0.035	0.101	0.106	0.064	0.168
0.13	0.061	0.066	0.034	0.104	0.108	0.051	0.182
0.14	0.065	0.07	0.033	0.127	0.138	0.087	0.216
0.15	0.068	0.079	0.039	0.132	0.136	0.082	0.227
0.16	0.071	0.081	0.041	0.15	0.153	0.091	0.243
0.17	0.069	0.084	0.04	0.165	0.178	0.095	0.289
0.18	0.081	0.093	0.053	0.161	0.17	0.088	0.294
0.19	0.095	0.111	0.047	0.156	0.169	0.077	0.338
0.2	0.091	0.109	0.056	0.171	0.184	0.083	0.344

Table A.3: Size: N=1000

endogeneity	adjusted EL	adjusted ET	adjusted CUE	EL	ET	CUE	t
0	0.044	0.047	0.043	0.056	0.058	0.052	0.053
0.01	0.044	0.045	0.039	0.038	0.04	0.032	0.044
0.02	0.05	0.052	0.04	0.058	0.058	0.052	0.063
0.03	0.043	0.041	0.038	0.056	0.057	0.055	0.067
0.04	0.05	0.053	0.045	0.069	0.072	0.061	0.088
0.05	0.048	0.049	0.039	0.073	0.075	0.062	0.087
0.06	0.058	0.061	0.051	0.088	0.092	0.076	0.113
0.07	0.062	0.066	0.05	0.086	0.092	0.066	0.113
0.08	0.063	0.066	0.048	0.126	0.128	0.107	0.167
0.09	0.066	0.07	0.048	0.136	0.144	0.12	0.2
0.1	0.058	0.058	0.043	0.148	0.154	0.125	0.215
0.11	0.065	0.069	0.055	0.164	0.179	0.147	0.245
0.12	0.056	0.061	0.042	0.179	0.196	0.146	0.273
0.13	0.057	0.07	0.041	0.208	0.217	0.167	0.321
0.14	0.056	0.067	0.042	0.214	0.23	0.175	0.347
0.15	0.076	0.085	0.045	0.221	0.24	0.188	0.379
0.16	0.075	0.093	0.058	0.262	0.283	0.194	0.418
0.17	0.08	0.096	0.056	0.236	0.269	0.172	0.457
0.18	0.081	0.093	0.066	0.259	0.291	0.194	0.492
0.19	0.093	0.121	0.071	0.293	0.328	0.208	0.542
0.2	0.115	0.153	0.094	0.302	0.334	0.21	0.576

Table A.4: Power: N=200, δ=0

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.652	1	1	0.832	1
-0.275	1	1	0.764	1	1	0.889	1
-0.25	1	1	0.887	1	1	0.949	1
-0.225	1	1	0.91	1	1	0.963	1
-0.2	0.996	0.999	0.94	0.998	1	0.967	0.999
-0.175	0.985	0.986	0.945	0.99	0.994	0.962	0.99
-0.15	0.904	0.914	0.867	0.943	0.946	0.902	0.946
-0.125	0.805	0.813	0.752	0.837	0.838	0.789	0.809
-0.1	0.651	0.657	0.639	0.657	0.655	0.64	0.65
-0.075	0.4	0.412	0.403	0.419	0.429	0.432	0.355
-0.05	0.199	0.199	0.195	0.217	0.212	0.21	0.182
-0.025	0.121	0.124	0.119	0.122	0.109	0.115	0.093
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.059	0.052	0.05	0.062	0.07	0.053	0.101
0.05	0.098	0.099	0.066	0.125	0.116	0.093	0.137
0.075	0.206	0.195	0.136	0.267	0.245	0.18	0.345
0.1	0.3	0.293	0.162	0.358	0.376	0.294	0.513
0.125	0.398	0.368	0.176	0.497	0.491	0.332	0.678
0.15	0.603	0.616	0.316	0.679	0.71	0.535	0.773
0.175	0.659	0.671	0.27	0.73	0.759	0.558	0.883
0.2	0.745	0.734	0.319	0.869	0.871	0.682	0.938
0.225	0.859	0.876	0.397	0.934	0.943	0.651	0.975
0.25	0.841	0.857	0.247	0.935	0.946	0.594	0.98
0.275	0.93	0.919	0.261	0.973	0.981	0.621	0.994
0.3	0.932	0.939	0.235	0.984	0.987	0.59	0.998

Table A.5: Power: N=200, δ=0.01

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.633	1	1	0.826	1
-0.275	1	1	0.77	1	1	0.901	1
-0.25	1	1	0.905	1	1	0.95	1
-0.225	1	1	0.905	1	1	0.945	1
-0.2	0.998	0.998	0.946	0.999	0.999	0.97	1
-0.175	0.976	0.979	0.936	0.981	0.987	0.963	0.991
-0.15	0.921	0.926	0.879	0.929	0.945	0.899	0.947
-0.125	0.823	0.833	0.784	0.839	0.841	0.815	0.829
-0.1	0.626	0.64	0.601	0.618	0.623	0.582	0.689
-0.075	0.387	0.418	0.403	0.403	0.432	0.441	0.421
-0.05	0.202	0.207	0.224	0.212	0.215	0.237	0.199
-0.025	0.095	0.103	0.095	0.095	0.09	0.096	0.085
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.055	0.051	0.038	0.065	0.072	0.068	0.075
0.05	0.094	0.094	0.067	0.136	0.125	0.105	0.155
0.075	0.145	0.139	0.107	0.206	0.215	0.176	0.298
0.1	0.335	0.31	0.23	0.381	0.408	0.337	0.462
0.125	0.425	0.427	0.255	0.535	0.586	0.478	0.649
0.15	0.569	0.562	0.269	0.677	0.706	0.542	0.771
0.175	0.643	0.647	0.327	0.799	0.824	0.633	0.899
0.2	0.803	0.8	0.415	0.881	0.903	0.691	0.943
0.225	0.861	0.857	0.36	0.944	0.955	0.763	0.98
0.25	0.922	0.938	0.407	0.985	0.984	0.789	0.991
0.275	0.927	0.919	0.238	0.978	0.984	0.672	0.99
0.3	0.929	0.908	0.194	0.978	0.985	0.64	0.995

Table A.6: Power: N=200, δ=0.1

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.728	1	1	0.76	1
-0.275	1	1	0.806	1	1	0.803	1
-0.25	1	1	0.854	1	1	0.84	1
-0.225	0.998	0.999	0.93	0.998	0.999	0.896	1
-0.2	0.994	0.996	0.929	0.989	0.99	0.858	1
-0.175	0.981	0.983	0.954	0.966	0.97	0.845	0.986
-0.15	0.923	0.933	0.885	0.869	0.877	0.754	0.944
-0.125	0.774	0.78	0.765	0.665	0.667	0.564	0.811
-0.1	0.589	0.596	0.597	0.467	0.439	0.368	0.657
-0.075	0.36	0.369	0.379	0.225	0.226	0.206	0.364
-0.05	0.193	0.207	0.222	0.101	0.102	0.098	0.162
-0.025	0.099	0.105	0.114	0.056	0.051	0.045	0.073
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.033	0.036	0.029	0.088	0.088	0.085	0.075
0.05	0.069	0.057	0.035	0.176	0.181	0.16	0.162
0.075	0.165	0.149	0.085	0.371	0.374	0.342	0.348
0.1	0.24	0.214	0.086	0.501	0.539	0.425	0.516
0.125	0.364	0.347	0.151	0.684	0.7	0.572	0.688
0.15	0.492	0.478	0.142	0.744	0.779	0.643	0.78
0.175	0.605	0.568	0.205	0.83	0.854	0.712	0.852
0.2	0.68	0.639	0.177	0.926	0.94	0.702	0.932
0.225	0.78	0.738	0.182	0.95	0.967	0.709	0.979
0.25	0.786	0.741	0.12	0.965	0.976	0.687	0.98
0.275	0.822	0.76	0.087	0.984	0.989	0.663	0.994
0.3	0.863	0.822	0.093	0.988	0.994	0.623	0.998

Table A.7: Power: N=200, δ=0.2

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.725	1	1	0.576	1
-0.275	1	1	0.774	1	1	0.678	1
-0.25	0.999	0.999	0.859	0.998	0.999	0.665	1
-0.225	0.999	0.999	0.912	0.996	0.998	0.689	1
-0.2	0.987	0.993	0.91	0.965	0.973	0.682	0.996
-0.175	0.978	0.985	0.935	0.923	0.929	0.664	0.98
-0.15	0.881	0.902	0.902	0.769	0.762	0.542	0.929
-0.125	0.794	0.809	0.823	0.546	0.545	0.391	0.856
-0.1	0.521	0.557	0.545	0.351	0.345	0.268	0.544
-0.075	0.363	0.375	0.364	0.172	0.156	0.112	0.382
-0.05	0.237	0.226	0.214	0.086	0.081	0.063	0.202
-0.025	0.091	0.095	0.112	0.044	0.037	0.035	0.074
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.032	0.029	0.019	0.102	0.111	0.088	0.059
0.05	0.05	0.036	0.023	0.261	0.232	0.186	0.162
0.075	0.057	0.029	0.011	0.336	0.367	0.304	0.304
0.1	0.087	0.057	0.007	0.416	0.445	0.358	0.439
0.125	0.151	0.107	0.015	0.602	0.633	0.482	0.571
0.15	0.255	0.153	0.02	0.691	0.735	0.489	0.745
0.175	0.361	0.262	0.028	0.849	0.881	0.62	0.891
0.2	0.438	0.34	0.026	0.875	0.886	0.543	0.921
0.225	0.529	0.438	0.035	0.928	0.955	0.545	0.965
0.25	0.652	0.468	0.028	0.942	0.961	0.516	0.981
0.275	0.65	0.52	0.02	0.956	0.983	0.515	0.988
0.3	0.733	0.581	0.009	0.971	0.991	0.47	0.997

Table A.8: Power: N=500, δ=0

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.962	1	1	0.996	1
-0.275	1	1	0.981	1	1	0.997	1
-0.25	1	1	0.99	1	1	0.998	1
-0.225	1	1	0.994	1	1	0.999	1
-0.2	1	1	0.999	1	1	0.999	1
-0.175	1	1	0.997	1	1	0.998	1
-0.15	1	1	1	1	1	1	1
-0.125	0.989	0.992	0.991	0.994	0.995	0.995	0.996
-0.1	0.942	0.949	0.951	0.961	0.966	0.964	0.963
-0.075	0.766	0.764	0.753	0.784	0.799	0.799	0.776
-0.05	0.4	0.396	0.419	0.436	0.428	0.433	0.458
-0.025	0.174	0.17	0.164	0.162	0.161	0.157	0.156
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.125	0.127	0.112	0.124	0.118	0.117	0.145
0.05	0.239	0.233	0.218	0.273	0.283	0.267	0.319
0.075	0.517	0.542	0.495	0.617	0.623	0.607	0.609
0.1	0.772	0.778	0.743	0.866	0.874	0.858	0.896
0.125	0.891	0.901	0.854	0.929	0.94	0.933	0.967
0.15	0.95	0.958	0.92	0.968	0.976	0.966	0.986
0.175	0.988	0.991	0.947	0.997	0.999	0.992	1
0.2	0.989	0.993	0.95	1	1	0.998	1
0.225	0.999	1	0.959	1	1	0.997	1
0.25	1	1	0.943	1	1	0.994	1
0.275	1	1	0.919	1	1	0.995	1
0.3	0.999	1	0.867	1	1	0.99	1

Table A.9: Power: N=500, δ=0.01

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.95	1	1	0.993	1
-0.275	1	1	0.978	1	1	0.995	1
-0.25	1	1	0.985	1	1	0.996	1
-0.225	1	1	0.991	1	1	0.998	1
-0.2	1	1	0.997	1	1	0.999	1
-0.175	1	1	0.999	1	1	0.999	1
-0.15	1	1	0.999	0.999	1	1	1
-0.125	0.993	0.996	0.994	0.998	0.998	0.996	0.998
-0.1	0.95	0.945	0.946	0.955	0.958	0.954	0.957
-0.075	0.727	0.729	0.728	0.747	0.752	0.748	0.768
-0.05	0.413	0.417	0.422	0.409	0.402	0.402	0.443
-0.025	0.147	0.139	0.158	0.15	0.151	0.151	0.103
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.051	0.065	0.071	0.094	0.095	0.097	0.111
0.05	0.238	0.241	0.251	0.322	0.352	0.34	0.411
0.075	0.561	0.545	0.503	0.635	0.671	0.669	0.67
0.1	0.723	0.755	0.704	0.812	0.828	0.82	0.855
0.125	0.859	0.862	0.818	0.939	0.942	0.937	0.952
0.15	0.959	0.963	0.923	0.984	0.987	0.983	0.992
0.175	0.987	0.992	0.955	0.996	0.997	0.995	0.998
0.2	0.992	0.994	0.961	1	1	0.998	1
0.225	0.999	1	0.937	1	1	0.995	1
0.25	0.999	1	0.914	1	1	0.998	1
0.275	0.999	0.999	0.88	1	1	0.99	1
0.3	0.999	1	0.808	1	1	0.989	1

Table A.10: Power: N=500, δ=0.1

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.966	1	1	0.969	1
-0.275	1	1	0.978	1	1	0.974	1
-0.25	1	1	0.992	1	1	0.983	1
-0.225	1	1	0.997	1	1	0.995	1
-0.2	1	1	0.998	1	1	0.99	1
-0.175	1	1	1	1	1	0.99	1
-0.15	0.999	1	1	0.997	0.998	0.986	1
-0.125	0.989	0.991	0.992	0.95	0.951	0.91	0.993
-0.1	0.951	0.95	0.954	0.786	0.8	0.751	0.962
-0.075	0.71	0.731	0.747	0.472	0.474	0.434	0.717
-0.05	0.425	0.426	0.439	0.155	0.161	0.146	0.382
-0.025	0.173	0.162	0.175	0.049	0.041	0.035	0.156
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.053	0.061	0.049	0.187	0.194	0.179	0.119
0.05	0.179	0.188	0.142	0.43	0.435	0.437	0.309
0.075	0.423	0.409	0.281	0.679	0.711	0.703	0.654
0.1	0.6	0.607	0.429	0.849	0.883	0.871	0.82
0.125	0.767	0.754	0.552	0.944	0.962	0.95	0.948
0.15	0.907	0.91	0.683	0.991	0.996	0.993	0.993
0.175	0.951	0.963	0.725	0.997	0.998	0.994	0.999
0.2	0.99	0.994	0.818	0.999	1	0.992	0.999
0.225	0.986	0.991	0.665	1	1	0.995	1
0.25	0.994	0.997	0.695	0.999	1	0.988	1
0.275	0.994	1	0.589	1	1	0.967	1
0.3	0.996	0.999	0.53	1	1	0.973	1

Table A.11: Power: N=500, δ=0.2

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.95	1	1	0.758	1
-0.275	1	1	0.966	1	1	0.804	1
-0.25	1	1	0.987	1	1	0.881	1
-0.225	1	1	0.995	1	1	0.909	1
-0.2	1	1	0.997	0.998	1	0.902	1
-0.175	1	1	0.997	0.991	0.998	0.884	1
-0.15	0.999	0.998	0.999	0.965	0.968	0.831	0.998
-0.125	0.984	0.988	0.995	0.812	0.838	0.662	0.991
-0.1	0.922	0.929	0.937	0.565	0.566	0.411	0.927
-0.075	0.732	0.743	0.763	0.21	0.212	0.169	0.758
-0.05	0.44	0.453	0.476	0.06	0.056	0.039	0.419
-0.025	0.172	0.175	0.18	0.022	0.017	0.017	0.141
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.022	0.015	0.014	0.184	0.201	0.205	0.132
0.05	0.03	0.023	0.011	0.434	0.441	0.402	0.341
0.075	0.147	0.096	0.03	0.7	0.741	0.672	0.686
0.1	0.309	0.236	0.067	0.799	0.834	0.773	0.847
0.125	0.494	0.414	0.127	0.927	0.944	0.917	0.93
0.15	0.645	0.598	0.174	0.957	0.986	0.927	0.984
0.175	0.732	0.655	0.113	0.996	0.999	0.981	1
0.2	0.864	0.861	0.177	0.994	1	0.959	1
0.225	0.863	0.846	0.102	0.994	1	0.947	1
0.25	0.931	0.92	0.107	0.998	1	0.948	1
0.275	0.939	0.926	0.069	1	1	0.923	1
0.3	0.988	0.977	0.114	1	1	0.852	1

Table A.12: Power: N=1000, δ=0

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.997	1	1	1	1
-0.275	1	1	0.999	1	1	1	1
-0.25	1	1	1	1	1	1	1
-0.225	1	1	1	1	1	1	1
-0.2	1	1	1	1	1	1	1
-0.175	1	1	1	1	1	1	1
-0.15	1	1	1	1	1	1	1
-0.125	1	1	1	1	1	1	1
-0.1	0.996	0.995	0.995	0.998	0.998	0.998	0.998
-0.075	0.951	0.952	0.951	0.959	0.959	0.955	0.955
-0.05	0.626	0.628	0.635	0.683	0.679	0.682	0.67
-0.025	0.208	0.209	0.215	0.236	0.228	0.226	0.197
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.178	0.186	0.185	0.219	0.229	0.231	0.23
0.05	0.529	0.52	0.517	0.573	0.572	0.576	0.632
0.075	0.859	0.858	0.835	0.89	0.899	0.899	0.935
0.1	0.986	0.989	0.986	0.995	0.996	0.996	0.997
0.125	0.997	0.997	0.996	0.998	0.998	0.998	0.998
0.15	1	1	1	1	1	1	1
0.175	1	1	1	1	1	1	1
0.2	1	1	1	1	1	1	1
0.225	1	1	1	1	1	1	1
0.25	1	1	0.994	1	1	1	1
0.275	1	1	0.993	1	1	1	1
0.3	1	1	0.988	1	1	1	1

Table A.13: Power: N=1000, δ=0.01

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.995	1	1	1	1
-0.275	1	1	0.998	1	1	1	1
-0.25	1	1	1	1	1	1	1
-0.225	1	1	1	1	1	1	1
-0.2	1	1	1	1	1	1	1
-0.175	1	1	1	1	1	1	1
-0.15	1	1	1	1	1	1	1
-0.125	1	1	1	1	1	1	1
-0.1	0.998	0.998	0.998	0.999	0.999	0.999	0.999
-0.075	0.945	0.946	0.945	0.964	0.968	0.967	0.972
-0.05	0.665	0.663	0.65	0.699	0.699	0.69	0.712
-0.025	0.205	0.2	0.21	0.212	0.205	0.197	0.213
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.18	0.18	0.18	0.21	0.216	0.21	0.23
0.05	0.493	0.507	0.493	0.576	0.576	0.582	0.569
0.075	0.814	0.816	0.807	0.887	0.896	0.883	0.903
0.1	0.961	0.964	0.956	0.983	0.983	0.98	0.985
0.125	0.989	0.992	0.99	0.997	0.996	0.997	0.996
0.15	1	1	0.999	1	1	1	1
0.175	1	1	1	1	1	1	1
0.2	1	1	1	1	1	1	1
0.225	1	1	0.999	1	1	1	1
0.25	1	1	0.999	1	1	1	1
0.275	1	1	0.995	1	1	1	1
0.3	1	1	0.98	1	1	1	1

Table A.14: Power: N=1000, δ=0.1

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.998	1	1	0.997	1
-0.275	1	1	0.999	1	1	1	1
-0.25	1	1	0.999	1	1	0.999	1
-0.225	1	1	1	1	1	0.999	1
-0.2	1	1	1	1	1	0.999	1
-0.175	1	1	1	1	1	0.998	1
-0.15	1	1	1	1	1	0.998	1
-0.125	1	1	1	1	1	0.997	1
-0.1	0.999	0.999	0.999	0.954	0.967	0.952	0.999
-0.075	0.934	0.934	0.949	0.636	0.642	0.614	0.961
-0.05	0.703	0.715	0.722	0.239	0.227	0.213	0.677
-0.025	0.246	0.25	0.266	0.028	0.025	0.025	0.243
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.085	0.083	0.06	0.323	0.321	0.322	0.222
0.05	0.399	0.394	0.331	0.732	0.741	0.723	0.627
0.075	0.671	0.65	0.557	0.92	0.935	0.927	0.895
0.1	0.931	0.92	0.851	0.987	0.991	0.991	0.989
0.125	0.976	0.986	0.956	0.999	1	1	1
0.15	0.993	0.999	0.987	1	1	1	1
0.175	0.998	1	0.989	1	1	1	1
0.2	1	1	0.983	1	1	1	1
0.225	1	1	0.98	1	1	0.999	1
0.25	1	1	0.959	1	1	1	1
0.275	1	1	0.964	1	1	1	1
0.3	1	1	0.898	1	1	1	1

Table A.15: Power: N=1000, δ=0.2

Beta	adjEL_scp	adjET_scp	adjCUE_scp	EL_scp	ET_scp	CUE_scp	t_scp
-0.3	1	1	0.996	1	1	0.933	1
-0.275	1	1	0.999	1	1	0.952	1
-0.25	1	1	0.997	1	1	0.953	1
-0.225	1	1	0.999	1	1	0.972	1
-0.2	1	1	0.999	1	1	0.964	1
-0.175	1	1	1	0.999	1	0.981	1
-0.15	1	1	1	0.998	0.999	0.969	1
-0.125	1	1	1	0.959	0.966	0.884	1
-0.1	0.992	0.997	0.997	0.757	0.784	0.643	0.996
-0.075	0.934	0.936	0.958	0.331	0.351	0.27	0.948
-0.05	0.697	0.676	0.726	0.069	0.059	0.029	0.722
-0.025	0.239	0.228	0.248	0.008	0.006	0.005	0.236
5.55E-17	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.016	0.006	0.004	0.266	0.277	0.284	0.17
0.05	0.061	0.026	0.009	0.607	0.651	0.615	0.544
0.075	0.303	0.213	0.067	0.885	0.924	0.905	0.906
0.1	0.529	0.47	0.152	0.964	0.984	0.973	0.982
0.125	0.749	0.714	0.213	0.992	1	0.994	0.998
0.15	0.886	0.87	0.378	0.998	1	0.999	1
0.175	0.936	0.963	0.476	0.997	1	0.999	1
0.2	0.961	0.981	0.375	0.999	1	0.998	1
0.225	0.986	0.993	0.369	1	1	0.999	1
0.25	0.986	0.997	0.302	0.999	1	0.997	1
0.275	0.994	0.999	0.328	1	1	0.991	1
0.3	0.994	1	0.242	0.999	1	0.988	1

Appendix B

Mis-specification-Robust Bootstrap for Empirical Likelihood Estimators

The lemmas that establish the consistency of the bootstrap reference are based on the results in Lee (2014), Schennach (2007), and Allen, Gregory, and Shimotsu (2008). We hereafter refer them as L2014, S2007, and AGS. The intuition of the proof is that when moment conditions are correctly specified either equal weights or EL weights are consistent, and when moment conditions are misspecified equal weights will only be used since the overidentification test would ultimately reject the null hypothesis with probability equals to 1.

Proof of Lemma 1

First, we note that the EL probabilities $\hat{p}_i = 1/n(1 - \hat{\lambda}' g(z_i, \theta^*))^{-1}$ need to be positive. If $\hat{\lambda}$ does not converge to 0 in probability, $\max_{i \leq n} \hat{\lambda}' g(z_i, \theta^*)$ would be unbounded, which would make some $\hat{p}_i$ negative. To show the consistency we use NS2014 Lemma A3 that $||\hat{g}(\hat{\theta})|| = O_p(n^{-1/2})$. Then by UWL, $\sup_{\theta \in \Theta} ||\hat{g}(\theta) - g(\theta)|| \xrightarrow{p} 0$. As $g(\theta) = 0$ has a unique zero at θ_0, $||g(\theta)||$ must be bounded away from zero outside any neighborhood of θ_0 . Hence, $\hat{\theta}$ must be inside any neighborhood of θ_0. This proves the consistency of $\hat{\theta}$.

Proof of Lemma 2

To prove Lemma 2 we first modify Lemma 3 of L2014.

$$\lim_{n \to \infty} n^{\alpha} P(P^*(\sup_{\theta \in \Theta} ||\hat{\theta}^* - \hat{\theta}|| > \epsilon) > n^{-a}) = 0, \tag{B.1}$$

For a given $\epsilon > 0$, there exists $\eta > 0$ independent of n such that $||\theta - \hat{\theta}|| > \epsilon$ implies that $0 < \eta \leq n^{-1} \sum_i (\rho(\hat{\lambda}' g_i(\theta) - \rho(\hat{\lambda}' \hat{g}_i(\theta))$ with probability equals to 1. This can be shown by using a similar logic in Lemma 3 of L2014. Then, we have

$$P(P^*(\sup_{\theta\in\Theta}||\hat{\theta}^* - \hat{\theta}|| > \epsilon) > n^{-a})$$
$$\leq P(P^*(\sup_{\theta\in\Theta}\sup_{\lambda\in\Lambda(\theta)}|n^{-1}\sum_i(\rho(\lambda' g_i^*(\theta) - \rho(\lambda' g_i(\theta))| > \eta/2) > n^{-a}) = o(n^{-a}).$$

Proof of Theorem 1

We follow the proof of Theorem 1 in Lee (2014). Firstly, we use the same result of Lemma 8 in L2014 as follows:

$$\lim_{n\to\infty} n^{\alpha}\sup_{z\in\mathbb{R}}|P(T\leq z) - [1+\sum_{i=1}^{2a} n^{-i/2\pi_i}]\Phi(z)| = 0$$

,

Then by the triangle inequality,

$$P(\sup_{z\in\mathbb{R}}|P(T\leq z) - P^*(T^*\leq z)| > n^{-1/2+\eta}\epsilon) \leq$$
$$P(\sup_{z\in\mathbb{R}}|P(T\leq z) - (1+\sum_{i=1}^{2} n^{-i/2}\pi_i(\delta, v_1))\Phi(z)| > n^{-1/2+\eta}\epsilon/4) +$$
$$P(\sup_{z\in\mathbb{R}}|P(T\leq z) - (1+\sum_{i=1}^{2} n^{-i/2}\pi_i(\delta, v_{n,1}^*))\Phi(z)| > n^{-1/2+\eta}\epsilon/4) +$$
$$P(\sup_{z\in\mathbb{R}} n^{-1/2}|\pi_1(\delta.v_1) - \pi_1(\delta, v_{n,1}^*)|\Phi(z) > n^{-1/2+\eta}\epsilon/4) +$$
$$P(\sup_{z\in\mathbb{R}} n^{-1}|\pi_2(\delta.v_1) - \pi_2(\delta, v_{n,1}^*)|\Phi(z) > n^{-1/2+\eta}\epsilon/4) = o(n^{-1}).$$

The last equality holds by lemma 8(a)-(b) in L2014.

Q.E.D.

Table B.1: Correct Model: $\rho_0 = 0.4$, α=0.05

n	100	200
Asymp	0.361	0.267
Boot-equal	0.0622	0.0512
Boot-EL	0.093	0.0954
Boot-Adapted	0.09	0.0904
J-test	0.0624	0.0484

Table B.2: Correct Model: $\rho_0 = 0.4$, α=0.1

n	100	200
Asymp	0.361	0.267
Boot-equal	0.0622	0.0512
Boot-EL	0.093	0.0954
Boot-Adapted	0.089	0.0864
J-test	0.125	0.0967

Table B.3: Correct Model: $\rho_0 = 0.4$, α=0.2

n	100	200
Asymp	0.361	0.267
Boot-equal	0.0622	0.0512
Boot-EL	0.093	0.0954
Boot-Adapted	0.07	0.0616
J-test	0.224	0.186

Table B.4: Correct Model: $\rho_0 = 0.9$, α=0.05

n	100	200
Asymp	0.724	0.6602
Boot-equal	0.0478	0.05
Boot-EL	0.0712	0.0908
Boot-Adapted	0.0662	0.0806
J-test	0.0786	0.0684

Table B.5: Correct Model: $\rho_0 = 0.9$, α=0.1

n	100	200
Asymp	0.724	0.6602
Boot-equal	0.0478	0.05
Boot-EL	0.0712	0.0908
Boot-Adapted	0.0652	0.0786
J-test	0.1168	0.0854

Table B.6: Correct Model: $\rho_0 = 0.9$, α=0.2

n	100	200
Asymp	0.724	0.6602
Boot-equal	0.0478	0.05
Boot-EL	0.0712	0.0908
Boot-Adapted	0.0502	0.0686
J-test	0.246	0.228

Table B.7: Misspecified Model: $\rho_1 = 0.6, \rho_2 = 0.2, \rho_0 = 0.3961$, α=0.05

n	100	200	500
Asymp	0.5682	0.4726	0.3798
Boot-equal	0.0588	0.0644	0.0766
Boot-EL	0.106	0.159	0.202
Boot-Adapted	0.0606	0.0822	0.0807
J-test	0.202	0.2212	0.508

Table B.8: Misspecified Model: $\rho_1 = 0.6, \rho_2 = 0.2, \rho_0 = 0.3961$, α=0.1

n	100	200	500
Asymp	0.5682	0.4726	0.3798
Boot-equal	0.0588	0.0644	0.0766
Boot-EL	0.106	0.159	0.202
Boot-Adapted	0.0806	0.0992	0.087
J-test	0.2224	0.2942	0.5608

Table B.9: Misspecified Model: $\rho_1 = 0.6, \rho_2 = 0.2, \rho_0 = 0.3961$, α=0.2

n	100	200	500
Asymp	0.5682	0.4726	0.3798
Boot-equal	0.0588	0.0644	0.0766
Boot-EL	0.106	0.159	0.202
Boot-Adapted	0.0846	0.108	0.0908
J-test	0.348	0.402	0.6608

Table B.10: Misspecified Model: $\rho_1 = 0.3, \rho_2 = 0.4, \rho_0 = -0.09765$, α=0.05

n	100	200	500
Asymp	0.5218	0.4842	0.4496
Boot-equal	0.0686	0.0946	0.1094
Boot-EL	0.1624	0.2578	0.33
Boot-Adapted	0.0844	0.1288	0.1106
J-test	0.368	0.524	0.845

Table B.11: Misspecified Model: $\rho_1 = 0.3, \rho_2 = 0.4, \rho_0 = -0.09765$, α=0.1

n	100	200	500
Asymp	0.5218	0.4842	0.4496
Boot-equal	0.0686	0.0946	0.1094
Boot-EL	0.1624	0.2578	0.33
Boot-Adapted	0.0944	0.1078	0.1086
J-test	0.4586	0.6624	0.955

Table B.12: Misspecified Model: $\rho_1 = 0.3, \rho_2 = 0.4, \rho_0 = -0.09765$, α=0.2

n	100	200	500
Asymp	0.5218	0.4842	0.4496
Boot-equal	0.0686	0.0946	0.1094
Boot-EL	0.1624	0.2578	0.33
Boot-Adapted	0.0954	0.1088	0.1026
J-test	0.6808	0.7624	0.988

Appendix C

Higher Order MSE Comparisons of Generalized Empirical Likelihood Estimators

We follow Newey and Smith (2004) Lemma A.4 as below to get the higher order expansion equation for our model.

Assumption C.0.1 *(a) $\mu_0 \in U$ is the unique solution to $E(g(z, \mu_0] = 0$;(b) U is compact;*

(c) $E[sup_{\mu \in U} \|g(z, \mu)\|^{\alpha}] < \infty$;(d)$\rho(v)$ is twice continuously differentiable in a neighborhood of zero.

Assumption 1 is adapted from Assumption 1 of Newey and Smith(2004) with our bivariate mean model.

Assumption C.0.2 *There is b(z) with $E[b(z_i)^6] < \infty$ such that for $0 \leq j \geq 4$ and all z,$\nabla^j g(z, \mu)$ exists on a neighborhood N ofμ_0,$\sup_{\mu \in N} \|\nabla^j g(z, \mu)\| \leq b(z)\|\mu - \mu_0\|$, and for each $\mu \in N$, $\|\nabla^4 g(z, \mu) - \nabla^4 g(z, \mu_0)\| \leq b(z)\|\mu - \mu_0\|$,$\rho(v)$ os four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.*

This assumption is the same as the Assumption (3) in Newey and Smith(2004) which is needed for the stochastic expansions.

Theorem C.0.1 *Suppose that the estimator $\hat{\theta}$ and vector of functions$m(z, \theta)$satisfies a)$\hat{\theta} = \theta_0 + O_p(n^{-1/2})$;b)$\sum_{i=1}^{n} m(z_i, \hat{\theta} = 0$,w.p.a.1;c)For some $\zeta > 2, d(z)$with$E(d(z) < \infty$,and $T_n = \theta : \| \theta - \theta_0 \| \leq n^{-1/\theta}$, w.p.a.a for $i = 1, 2..., n$, $m(z_i, \theta)$ is three times continuously differentiable on T_n and for $\theta \in T_n$,*

$$\| \partial^3(m(z_i, \theta))/\partial\theta_j\partial\theta_k\partial\theta_l - \partial^3(m(z_0), \theta)/\partial\theta_j\partial\theta_k\partial\theta_l \| \leq d(z_i) \| \theta - \theta_0 \| \tag{C.1}$$

d)$E(m(z, \theta_0) = 0$ and $M = E(\partial m(z, \theta_0)/\partial\theta)$exists and is nonsingular. Let $M_j =$

$E[\partial^2 m(z,\theta_0)/\partial\theta j\partial\theta], M_{jk} = E[\partial^3 m(z,\theta_0)/\partial\theta j\partial\theta_k\partial\theta],$

$A(z) = \partial m(z,\theta_0)/\partial\theta - M,\ B_j(z) = \partial^2 m(z,\theta_0)/\partial\theta j\partial\theta - M_j,$

$\psi(z) = -M^{-1}m(z,\theta_0),\ a(z) = vecA(z),\ b(z) = vec[B_1(z), ..., B_q(z)].$

Then,

$$\hat{\theta} - \theta_0 = \frac{\tilde{\psi}}{n^{1/2}} + \frac{\tilde{Q}_1}{n} + \frac{\tilde{Q}_2}{n^{3/2}} + O_p(n^{-2}) \tag{C.2}$$

Where, $\tilde{Q}_1 = -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}^{q}\tilde{\psi}_j M_j\tilde{\psi}/2],$

$\tilde{Q}_2 = -M^{-1}[\tilde{A}\tilde{Q}_1 + \sum_{j=1}^{q}\tilde{\psi}_j M_j\tilde{Q}_1 + \tilde{Q_{1j}}M_j\tilde{\psi} + \tilde{\psi}_j\tilde{B}_j\tilde{\psi}/2 + \sum_{j,k=1}^{q}\tilde{\psi}_j\tilde{\psi}_k M_{jk}\tilde{\psi}/6]$

Now we apply this theorem to this simple two dimension model. To simplify the calculation we assume z_1 and z_2 are independent, $,var(x_1) = var(x_2) = 1$, and $E(z_1-\mu)^3 = E(z_2-\mu)^3 = \mu_3$. These are some mild assumptions,and could be easily generalized. According to the first order conditions we could easily get $m(z_i,\theta)$ as a 3×1 matrix

$$\begin{pmatrix} \rho_1(\lambda_1(z_{1i}-\mu) + \lambda_2(z_{2i}-\mu))(-\lambda_1-\lambda_2) \\ \rho_1(\lambda_1(z_{1i}-\mu) + \lambda_2(z_{2i}-\mu))(z_{1i}-\mu) \\ \rho_1(\lambda_1(z_{1i}-\mu) + \lambda_2(z_{2i}-\mu))(z_{2i}-\mu) \end{pmatrix}$$

We could then easily get

$$M = E[\partial m(z,\theta_0)/\partial\theta] = \begin{pmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

We could then get the first component in our expansion equation,

$$\tilde{\psi} = \sqrt{n}\begin{pmatrix} 1/2(a_{1n}+a_{2n}) \\ 1/2(a_{2n}-a_{1n}) \\ 1/2(a_{1n}-a_{2n}) \end{pmatrix}.$$

Where $a_{1n} = \frac{\sum_{i=1}^{n}(z_{1i}-\mu)}{n}$, $a_{2n} = \frac{\sum_{i=1}^{n}(z_{2i}-\mu)}{n}$

Similarly the second component in the expansion equation,

$\tilde{Q}_1 = -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}^{3} \tilde{\psi}_1 M_1 \tilde{\psi}/2]$ as

$$n/4 \begin{pmatrix} (c_{1n} - c_{2n})(a_{2n} - a_{1n}) - 1/2\rho_3\mu_3(a_{2n} - a_{1n})^2 \\ (c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n}) \\ (c_{1n} + c_{2n} - 2b_n)(a_{2n} - a_{1n}) \end{pmatrix}$$

Where $c_{1n} = \sum_{i=1}^{n}(z_{1i} - \mu)^2/n - 1$, $c_{2n} = \sum_{i=1}^{n}(z_{2i} - \mu)^2/n - 1$, and $b_n = \sum_{i=1}^{n}(z_{1i} - \mu)(z_{2i} - \mu)/n$.

Here we start to derive the last part in the expansion equation, which is the most complicated one.

$$\tilde{Q}_2 = -M^{-1}[\tilde{A}\tilde{Q}_1 + \sum_{j=1}^{3}(\tilde{\psi}_j M_j \tilde{Q}_1 + \tilde{Q_{1j}} M_j \tilde{\psi} + \tilde{\psi}_j \tilde{B}_j \tilde{\psi})/2 + \sum_{j,k=1}^{3} \tilde{\psi}_j \tilde{\psi}_k M_{jk} \tilde{\psi}/6] \tag{C.3}$$

With some further calculations we could get

$$-M^{-1}\tilde{A}\tilde{Q}_1 = 1/4n^{2/3} \begin{pmatrix} 1/2(c_{1n} - c_{2n})(c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n}) \\ 1/2(c_{2n} - c_{1n})(c_{1n} + c_{2n} - 2b_n)(a_{2n} - a_{1n}) \\ 1/2(c_{1n} - c_{2n})(c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n}) \end{pmatrix}.$$

and,

$$-M^{-1}\sum_{j=1}^{3} \tilde{\psi}_j M_j \tilde{Q}_1 = 1/16n^{2/3} \begin{pmatrix} \rho_3\mu_3(c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n})^2 \\ 0 \\ 0 \end{pmatrix}.$$

$$-M^{-1}\sum_{j=1}^{3} \tilde{Q_{1j}} M_j \tilde{\psi} = 1/16n^{2/3} \begin{pmatrix} \rho_3\mu_3(c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n})^2 \\ 0 \\ 0 \end{pmatrix}.$$

$$-M^{-1}\sum_{j=1}^{3}\tilde{\psi}_j\tilde{B}_j\tilde{\psi} = 1/8n^{2/3}\begin{pmatrix} (a_{1n}-a_{2n})^2(a_{1n}+a_{2n})+1/2\rho_3 f_n(a_{1n}-a_{2n})^2 \\ \rho_3(d_{1n}+d_{2n}-e_{1n}-e_{2n})(a_{1n}-a_{2n})^2 \\ -\rho_3(d_{1n}+d_{2n}-e_{1n}-e_{2n})(a_{1n}-a_{2n})^2 \end{pmatrix}.$$

where, $d_{1n} = 1/n\sum_{i=1}^{n}(z_{1i}-\mu)^2(z_{2i}-\mu)$ $d_{2n} = 1/n\sum_{i=1}^{n}(z_{2i}-\mu)^2(z_{1i}-\mu)$, $e_{1n} = 1/n\sum_{i=1}^{n}(z_{1i}-\mu)^3-\mu_3$, $f_n = (d_{1n}+d_{2n}-c_{1n}-c_{2n})$. and $e_{2n} = 1/n\sum_{i=1}^{n}(z_{2i}-\mu)^3-\mu_3$ Finally, the last part in $\tilde{Q}_2$.

So lastly,

$$-M^{-1}\sum_{j,k=1}^{3}\tilde{\psi}_j\tilde{\psi}_k M_{jk}\tilde{\psi}/6 = 3/4n^{3/2}\begin{pmatrix} \rho_3(a_{1n}+a_{2n})(a_{1n}-a_{2n})^2 \\ \rho_4\mu_4(a_{2n}-a_{1n}) \\ \rho_4\mu_4(a_{2n}-a_{1n}) \end{pmatrix}.$$

So, we derive the expansion equation:

$$\hat{\mu}-\mu_0 = 1/2(a_{1n}+a_{2n})+1/4(c_{1n}-c_{2n})(a_{2n}-a_{1n})-1/8\rho_3\mu_3(a_{2n}-a_{1n})^2 \tag{C.4}$$
$$+1/8(c_{1n}-c_{2n})(c_{1n}+c_{2n}-2b_n)(a_{1n}-a_{2n})+1/8\rho_3\mu_3(a_{2n}-a_{1n})^2(c_{1n}+c_{2n}-2b_n) \tag{C.5}$$
$$+1/8(a_{2n}-a_{1n})^2(a_{1n}+a_{2n})+1/16\rho_3(d_{1n}+d_{2n}-e_{1n}-e_{2n})(a_{1n}-a_{2n})^2 \tag{C.6}$$
$$+1/8\rho_3(a_{1n}-a_{2n})^2(a_{1n}+a_{2n})+O_p(n^{-2}) \tag{C.7}$$

Meanwhile,

$$1/2(a_{1n}+a_{2n}) = O_p(n^{-1/2}) \tag{C.8}$$
$$1/4(c_{1n}-c_{2n})(a_{2n}-a_{1n})-1/8\rho_3\mu_3(a_{2n}-a_{1n})^2 = O_p(n^{-1}) \tag{C.9}$$

$$1/8c_{1n} - c_{2n}(c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n}) + 1/8\rho_3\mu_3(a_{2n} - a_{1n})^2(c_{1n} + c_{2n} - 2b_n)$$
$$+ 1/8(a_{2n} - a_{1n})^2(a_{1n} + a_{2n}) + 1/16\rho_3(d_{1n} + d_{2n} - e_{1n} - e_{2n}))(a_{1n} - a_{2n})^2$$
$$+ 1/8\rho_3(a_{1n} - a_{2n})^2(a_{1n} + a_{2n}) = O_p(n^{-3/2}) \quad \text{(C.10)}$$

So collecting those terms with ρ we would get

$$E(\hat{\mu} - \mu_0)^2 = C + 1/64\rho_3^2\mu_3^2 E(a_{2n} - a_{1n})^4 - 1/8\rho_3\mu_3 E((a_{2n} - a_{1n})^2(a_{2n} + a_{1n}))$$
$$- 1/16\rho_3\mu_3 E((a_{2n} - a_{1n})^3(c_{1n} - c_{2n})) + 1/8\rho_3\mu_3 E((c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n})^2(a_{1n} + a_{2n}))$$
$$+1/16\rho_3 E((d_{1n}+d_{2n}-e_{1n}-e_{2n})(a_{1n}-a_{2n})^2(a_{1n}+a_{2n}))+1/8\rho_3 E((a_{1n}-a_{2n})^2(a_{1n}+a_{2n})^2+o_p(n^{-2}) \quad \text{(C.11)}$$

where C is a constant term(not relevant with ρ),and we could easily compute these expectations:

$$\begin{aligned}
E(a_{2n} - a_{1n})^4 &= 12/n^2 + o_p(n^{-2}) \\
E((a_{2n} - a_{1n})^2(a_{2n} + a_{1n})) &= 2\mu_3/n^2 + o_p(n^{-2}) \\
E((a_{2n} - a_{1n})^3(c_{1n} - c_{2n})) &= -12\mu_3/n^2 + o_p(n^{-2}) \\
E((c_{1n} + c_{2n} - 2b_n)(a_{1n} - a_{2n})^2(a_{1n} + a_{2n})) &= 4\mu_3/n^2 + o_p(n^{-2}) \\
E((d_{1n} + d_{2n} - e_{1n} - e_{2n})(a_{1n} - a_{2n})^2(a_{1n} + a_{2n})) &= 4/n^2 - 4\mu_4/n^2 + o_p(n^{-2}) \\
E((a_{1n} - a_{2n})^2(a_{1n} + a_{2n})^2 &= 4/n^2 + o_p(n^{-2})
\end{aligned}$$

Finally,

$$E(\hat{\mu} - \mu_0)^2 = C + 3/16\frac{\rho_3^2\mu_3^2}{n^2} + \frac{\rho_3\mu_3^2}{n^2} + 3/4\frac{\rho_3}{n^2} - 1/4\frac{\rho_3\mu_4}{n^2} + o_p(n^{-2}) \quad \text{(C.12)}$$

Minimizing this quadratic function to get the optimal $\hat{\rho}$ as (suppose $\mu_3 \neq 0$)

$$\hat{\rho_3} = -3/8 + \frac{2\mu_4 - 6}{3\mu_3^2} \quad \text{(C.13)}$$

and, we could get the $\hat{\gamma}$ by $\rho_3 = -(1-\gamma)$.

Q.E.D.

References

[1] Allen, J., A. W., and Shimotsu, K. *Empirical likelihood block bootstrapping*. Journal of Econometrics, 161(2), 110-121, 2011.

[2] Altonji, J. G., Segal, L. M. *Small-sample bias in GMM estimation of covariance structures*. Journal of Business and Economic Statistics, 14(3), 353-366, 1996.

[3] Anatolyev, S. *GMM, GEL, Serial Correlation, and Asymptotic Bias*. Econometrica, 73(3), 983-1002, 2005.

[4] Berkowitz, D.,Caner, M. and Fang, Y. *Are "Nearly Exogenous Instruments" reliable?* Economic Letters,101(1), 20-23, 2008.

[5] Berkowitz, D., Caner, M. and Fang, Y. *The validity of instruments revisited*. Journal of Econometrics, 166(2), 255-256, 2012.

[6] Brown,B. W.,and Newey, W.K., *Generalized method of moments, efficient bootstrapping, and improved inference*. Journal of Business and Economic Statistics, 20(4), 507-517, 2002.

[7] Guggenberger, P. *On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption*. Econometric Theory, 28(02), 387-421, 2012.

[8] Guggenberger, P., and Hahn, J. *Finite Sample Properties of the Two-Step Empirical Likelihood Estimator.*. Econometric Reviews, 27(4-6), 526-541, 2005.

[9] Guggenberger, P., and Richard, S. *Generalized empirical likelihood estimators and tests under partial, weak, and strong identification*. Econometric Theory, 21(04), 3667-709, 2005.

[10] Hahn, J. *A note on bootstrapping generalized method of moments estimators*. Econometric Theory, 12, 187-197, 1996.

[11] Hall, P., and Horowitz, J. L. *Bootstrap critical values for tests based on generalized method of moments estimators*. Econometrica, 64(4), 891-916, 1996.

[12] Hall, P. *The bootstrap and Edgeworth expansion*. Springer. 1992.

[13] Hansen, L. P. *Large sample properties of generalized method of moments estimators.* Econometrica 50(4), 1029-1054, 1982.

[14] Hansen, L. P., J.Heaton and A.Yaron *Finite-Sample Properties of Some Alternative GMM estimators*. Journal of Business and Economic Statistics 14, 262-280, 1996.

[15] Imbens, G. W. *Generalized method of moments and empirical likelihood*. The review of Economic Studies, 64(3), 359-383, 1997.

[16] Kitamura, Y., and Stutzer, M. *An information-theoretic alternative to generalized method of moments estimation*. Econometrica, 65(4), 861-874, 1997.

[17] Lee, S. *Asymptotic Refinements of a Misspecification-Robust Bootstrap for GEL Estimators*. Working Paper, 2014.

[18] Newey, W.K, and Smith ,R.J. *Higher order properties of GMM and generalized empirical likelihood estimators*. Econometrica,72(1), 219-255, 2004.

[19] Owen,A. *Empirical Likelihood Ratio Confidence Intervals for a Single Functional*. Biometrika 75, 237-249, 1998.

[20] Qin, J., and Lawless, J. *Empirical likelihood and general estimating equations*. The Annals of Statistics, 300-325, 1994.