OPTIMAL COMPARISON OF MISSPECIFIED MOMENT RESTRICTION MODELS UNDER A CHOSEN MEASURE OF FIT

By

Vadim Marmer and Taisuke Otsu

August 2009
Revised July 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1724

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Optimal Comparison of Misspecified Moment Restriction Models under a Chosen Measure of Fit[1]

Vadim Marmer[2]
University of British Columbia

Taisuke Otsu[3]
Yale University

July 25, 2011

[1]We thank Donald Andrews, Nikolay Gospodinov, Bruce Hansen, Hiroyuki Kasahara, Yuichi Kitamura, Katsumi Shimotsu, and participants at the 2009 Applied Econometrics and Forecasting in Macroeconomics and Finance Workshop at the Federal Reserve Bank of St. Louis, 2009 SETA Conference in Kyoto, and 2010 Info-Metrics Conference hosted by American University for helpful comments. We would like to also thank the anonymous referees and the guest co-editor whose comments greatly helped to improve the paper. Our research is supported by the Social Science and Research Council of Canada under grants 410-2007-1998 and 410-2010-1394 (Marmer), and the National Science Foundation under SES-0720961 (Otsu).

[2]Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC, V6T 1Z1, Canada. Tel.: 1-604-822-8217, email: vadim.marmer@ubc.ca.

[3]Cowles Foundation and Department of Economics, Yale University, P.O. Box 208281, New Haven, CT, 06520-8281, USA. Tel.: 1-203-432-9778, email: taisuke.otsu@yale.edu.

Abstract

Suppose that the econometrician is interested in comparing two misspecified moment restriction models, where the comparison is performed in terms of some chosen measure of fit. This paper is concerned with describing an optimal test of the Vuong (1989) and Rivers and Vuong (2002) type null hypothesis that the two models are equivalent under the given measure of fit (the ranking may vary for different measures). We adopt the generalized Neyman-Pearson optimality criterion, which focuses on the decay rates of the type I and II error probabilities under fixed non-local alternatives, and derive an optimal but practically infeasible test. Then, as an illustration, by considering the model comparison hypothesis defined by the weighted Euclidean norm of moment restrictions, we propose a feasible approximate test statistic to the optimal one and study its asymptotic properties. Local power properties, one-sided test, and comparison under the generalized empirical likelihood-based measure of fit are also investigated. A simulation study illustrates that our approximate test is more powerful than the Rivers-Vuong test.

JEL classification: C12; C14; C52
Keywords: Moment restriction; Model comparison; Misspecification; Generalized Neyman-Pearson optimality; Generalized method of moments

1 Introduction

Econometric models are often defined in the form of moment restrictions and estimated by the generalized method of moments (GMM) (Hansen, 1982), empirical likelihood (EL) (Owen, 1988; Qin and Lawless, 1994), or their variants (see, e.g., Newey and Smith (2004) and Kitamura (2007)).[1] Moment restriction models are semiparametric and allow flexible distribution forms of data. However, in many applications it is often reasonable to suspect that those moment restrictions are misspecified. While misspecified models are typically rejected with probability approaching one by some overidentifying restriction test, they nevertheless can be of interest as approximations to the true unknown data generating process.[2] In this context, choosing one model having the best measure of fit among several competing misspecified models is of great importance for practitioners.

Misspecified models and inference procedures for such models have been discussed extensively in econometrics. White (1982) studied the properties of maximum likelihood under misspecification. Hendry (1979) and Maasoumi and Phillips (1982) discussed estimation and inference with invalid instruments in a linear regression model. In more general frameworks allowing nonlinear models, Gallant and White (1988) and Hall and Inoue (2003) discussed properties of the GMM estimator for misspecified moment restriction models.

In a seminal paper, Vuong (1989) proposed a test of the null hypothesis that two misspecified parametric models provide an equivalent approximation to the data generating distribution in terms of their Kullback-Leibler information criteria (KLIC). This approach was extended in Rivers and Vuong (2002) (RV, hereafter) to a more general framework which includes misspecified moment restriction models. In a recent paper, Hall and Pelletier (2011) (see also Hall and Pelletier (2007) for a detailed theoretical argument and simulation study) analyzed in depth the properties of the RV test for moment restriction models, and showed that the asymptotic null distribution of the RV test statistic depends on the degree of misspecification. Furthermore, they pointed out that the ranking obtained by the RV test crucially depends on the choice of weighting matrices used to define the model comparison hypotheses. Kitamura (2000) and Kitamura (2003) developed information theoretic approaches to compare misspecified unconditional and conditional moment restriction models, respectively, in terms of their KLIC to the data generating distribution. This information theoretic approach was employed, for example, in Christoffersen, Hahn, and Inoue (2001) for a comparison of Value-at-Risk measures and Kitamura and Stutzer (2002) for a comparison of stochastic discount factor models. Corradi and Swanson (2007) proposed a Kolmogorov-type test to compare misspecified dynamic stochastic general equilibrium models. Dridi, Guay, and Renault (2007) address the issue of misspecification from a perspective of indirect inference (Gouriéroux, Monfort, and Renault, 1993). They show how (some components of) structural parameters can be consistently estimated despite

[1] See Hall (2005) for a comprehensive review on the GMM.

[2] For example, Prescott (1991) argued that a model is only an approximation and should not be regarded as a null hypothesis to be statistically tested.

misspecification when certain requirements from the encompassing principle are satisfied.

This paper considers optimal testing of model comparison hypotheses for misspecified unconditional moment restriction models under some chosen measure of fit. Our focus is not on the choice of a measure of fit used to set up the model comparison hypotheses, but on the choice of a test given the measure of fit. To set up (a list of) model comparison hypotheses, the researcher must acknowledge that the rankings crucially depend on the choice of the hypotheses as emphasized by Hall and Pelletier (2011) and assess the validity of the chosen hypotheses based on economic or statistical considerations. This paper starts from the situation where the researcher has already chosen a model comparison hypothesis of interest. This paper does not make any specific recommendation about how to choose the measure of fit for the hypothesis.

To evaluate different tests for a given model comparison hypothesis, we employ the large deviation approach.[3] In particular, we adopt the generalized Neyman-Pearson (GNP) optimality criterion, which compares the decay rate of the type II error probability under fixed non-local alternatives subject to a constraint on the decay rate of the type I error probability. Based on Hoeffding (1965), Zeitouni and Gutman (1991) developed the notion of the GNP optimality and applied it to hypothesis testing problems in parametric models. Kitamura (2001) and Kitamura, Santos, and Shaikh (2009) studied the GNP optimality for testing the validity of overidentified moment restrictions. This paper extends these GNP optimality analyses to model comparison tests. Based on a modified version of the GNP optimality criterion, we derive an optimal test that is defined by the KLIC between a neighborhood of the empirical measure and a set of measures satisfying the model comparison null hypothesis.

Since the derived optimal test is generally infeasible, we then discuss its feasible approximation. As an illustration, we first consider a measure of fit defined by the weighted Euclidean (WE) norm of moment restrictions, propose an approximate test statistic, and study its asymptotic properties. The weight matrix used to define the WE norm may be unknown and may contain unknown parameters. We also demonstrate how to extend our approach to the generalized EL-based measures of fit.

The asymptotic analysis for the WE norm example reveals that our approximate test shares several common features with the conventional RV test: (i) for both tests, preliminary estimation of unknown parameters in the models does not affect the asymptotic null distributions of the test statistics, and the null distributions are the same as if the pseudo-true parameter values were known; and (ii) under some local alternatives, both tests have the same local power functions. On the other hand, advantages of our approach over the RV test are: (i) our test is motivated by the GNP optimality and shows better power properties in the simulation study; and (ii) although the asymptotic null distribution of the RV test statistic depends on whether the models are nested, non-nested, or overlapping in a certain sense (see Rivers and Vuong (2002) and Hall and Pelletier (2011)), our test statistic has

[3]See, e.g., Dembo and Zeitouni (1998) for a review on large deviation theory.

the same null distribution regardless of the nested, non-nested, or overlapping structure of the competing models.

While we focus on the global power properties of model comparison tests under fixed non-local alternatives, another conventional way to evaluate power properties of a test is to evaluate the local power function under Pitman-type local alternatives, where the data generating distribution drifts to the null hypothesis as the sample size increases. In Section 5.1, we illustrate that our approximate test and the RV test have the same local power function against local alternatives in a simple setup. Therefore, the local power analysis may not be informative enough to explain the different finite sample performances of these tests as presented in Section 6. On the other hand, since different tests typically show different global power properties (see, e.g., Hoeffding (1965)), the GNP optimality analysis for global power can be useful to explain superior finite sample power properties of our approximate test observed in the simulation study.

Lastly, we would like to emphasize that the problem of comparison of misspecified models should be discerned from non-nested hypothesis testing problems (Davidson and MacKinnon, 1981; MacKinnon, 1983; Smith, 1992). The non-nested hypothesis testing literature is concerned with testing whether one of the competing models is correctly specified. On the other hand, we compare two misspecified models in terms of their measures of fit. Thus, the two approaches, non-nested testing and model comparison testing, are not competing but rather complementary. The former can be used in a search for correct specification, while the latter can be used when the econometrician suspects that all competing models are misspecified or when those models have been rejected by some specification tests. Examples of non-nested testing for moment restriction models are Singleton (1985) and Ghysels and Hall (1990). They consider specification tests of Euler equations when some information about specific non-nested alternatives is available.[4]

The rest of the paper is organized as follows. Section 2 describes the testing framework with some examples and briefly discusses the choice of a measure of fit. Section 3 conducts the GNP optimality analysis for model comparison testing under a given measure of fit. Section 4 illustrates implementation of the approximate test when measures of fit are defined by the WE norms of moment restrictions, and studies the asymptotic properties of the proposed test statistic. Section 5 contains three extensions: local power analysis for the RV test and ours, one-sided hypothesis testing, and test for the hypotheses defined using the generalized EL-based measures of fit. In Section 6, we conduct a simulation study. Section 7 concludes. All proofs are given in the Appendix.

We use the following notation. Let $cl\,(A)$ and $int\,(A)$ be the closure and interior of a set A, respectively. Let $\Pr\{A : \mu\}$ and E_μ be the probability of an event A and the expectation under a probability measure μ, respectively. The data generating measure is denoted by μ_0, and the expectation under μ_0 is denoted by E_{μ_0} or simply by E. Let $\|z\| = \sqrt{\text{trace}\,(zz')}$

[4] The EL-based non-nested tests for moment restriction models are considered by Smith (1997), Ramalho and Smith (2002), and Otsu and Whang (2008).

be the Euclidean norm for a vector or matrix z and $\|z\|_W = \sqrt{z'Wz}$ be the WE norm for a vector z and a symmetric positive definite matrix W.

2 Model comparison

2.1 Setup

Suppose that we observe an i.i.d. sample $\{w_i\}_{i=1}^n$ of a random vector $w \in \mathbb{R}^q$ drawn from the unknown probability measure μ_0. Consider the unconditional moment restriction model:

$$E_{\mu_0} g(w, \theta_0) = 0, \tag{1}$$

where $g : \mathbb{R}^q \times \Theta \to \mathbb{R}^{l_g}$ is a known function up to unknown parameters $\theta_0 \in \Theta \subset \mathbb{R}^{p_g}$ with $l_g > p_g$. In this paper, we denote moment restriction models by their corresponding moment functions. For example, the model in equation (1) is called model g. If model g is correctly specified (i.e., (1) is satisfied at some $\theta_0 \in \Theta$), then we can apply the standard GMM theory for estimation and inference of θ_0. The condition $l_g > p_g$ implies that model g is overidentified.

The focus of this paper is to compare two misspecified moment restriction models. To formalize our idea, we introduce some notation. Let $\mathcal{M}$ be the space of all probability measures on $\mathbb{R}^q$ and define

$$\mathcal{P}_\theta^g = \{\mu \in \mathcal{M} : E_\mu g(w, \theta) = 0\}, \quad \mathcal{P}^g = \cup_{\theta \in \Theta} \mathcal{P}_\theta^g,$$

i.e., $\mathcal{P}_\theta^g$ is a set of measures satisfying the moment restrictions of model g at a given $\theta \in \Theta$, and $\mathcal{P}^g$ is a set of measures satisfying the moment restrictions at some $\theta \in \Theta$. Then misspecification of model g is defined as follows.[5]

Definition 1 (Misspecification)**.** *Model g is said to be misspecified if $\mu_0 \notin \mathcal{P}^g$.*

An alternative moment restriction model is similarly defined as $E_{\mu_0} h(w, \beta_0) = 0$, where $h : \mathbb{R}^q \times B \to \mathbb{R}^{l_h}$ is a known function up to unknown parameters $\beta_0 \in B \subset \mathbb{R}^{p_h}$ with $l_h > p_h$. For model h, we also define the sets $\mathcal{P}_\beta^h = \{\mu \in \mathcal{M} : E_\mu h(w, \beta) = 0\}$ and $\mathcal{P}^h = \cup_{\beta \in B} \mathcal{P}_\beta^h$.

We consider the situation where models g and h are both misspecified and we wish to compare these models in terms of their goodness of fit. Let $D(g, \mu_0)$ and $D(h, \mu_0)$ be measures of fit of models g and h to the data generating measure μ_0, respectively. For example, Vuong (1989) and Kitamura (2000) adopted the KLIC:

$$D_{KL}(g, \mu_0) = \inf_{\mu \in \mathcal{P}^g} I(\mu_0 \| \mu),$$

[5] The overidentification condition $l_g > p_g$ is needed for a model to be misspecified in the sense of Definition 1 as discussed in Hall and Inoue (2003, Proposition 1).

where

$$I(\mu_0 \| \mu) = \begin{cases} \int \log\left(\frac{d\mu_0}{d\mu}\right) d\mu_0 & \text{if } \mu_0 \text{ is absolutely continuous with respect to } \mu, \\ \infty & \text{otherwise.} \end{cases}$$

Another example, studied in depth by Hall and Pelletier (2011), is the WE norm of (the violation of) the moment restrictions:

$$D_{WE}(g, \mu_0) = \min_{\theta \in \Theta} \|E_{\mu_0} g(w, \theta)\|^2_{W_g}, \tag{2}$$

where the weight matrix W_g may be unknown and needs to be estimated. The measures of fit $D_{KL}(h, \mu_0)$ and $D_{WE}(h, \mu_0)$ for model h are similarly defined. Once the researcher has chosen a measure of fit of interest, model comparison testing problems can be defined as follows.

Definition 2 (Model comparison testing problem)**.**

(i) *The two-sided model comparison testing problem between models g and h under the measure of fit D is the one to test*

$$H_0 : D(g, \mu_0) = D(h, \mu_0) \quad \textit{against} \quad H_1 : D(g, \mu_0) \neq D(h, \mu_0). \tag{3}$$

(ii) *The one-sided model comparison testing problem between models g and h under the measure of fit D is the one to test*

$$H_0 \textit{ or } H_0^g : D(g, \mu_0) \leq D(h, \mu_0) \quad \textit{against} \quad H_1^h : D(g, \mu_0) > D(h, \mu_0),$$

where the roles of models g and h can be interchanged.

We first present our main results for the two-sided test and then discuss the one-sided test in Section 5.2. We close this subsection by providing two economic examples of model comparison testing. The first example is concerned with misspecified linear instrumental variable regression models (e.g., Hendry, 1979; Maasoumi and Phillips, 1982), and the second example is borrowed from the asset pricing literature (e.g., Hansen and Jagannathan, 1997).

Example 1 (Instumental variable regression models)**.** Let $w = \left(y, x', z^{g\prime}, z^{h\prime}\right)'$, where y is a dependent variable, x is a vector of endogenous regressors, and z^g and z^h are different vectors of instruments for models g and h, respectively (in general, the regressors may vary with the models as well). Moment functions for models g and h are defined as $g(w, \theta) = z^g(y - x'\theta)$ and $h(w, \beta) = z^h(y - x'\beta)$, respectively. Both models are assumed to be overidentified, i.e., $\dim(z^g) > \dim(x)$ and $\dim\left(z^h\right) > \dim(x)$. If the researcher is interested in the WE measure of fit D_{WE} in (2), the null hypothesis for the model comparison is

$$H_0 : \min_{\theta \in \Theta} \left\|E_{\mu_0} z^g \left(y - x'\theta\right)\right\|^2_{W_g} = \min_{\beta \in B} \left\|E_{\mu_0} z^h \left(y - x'\beta\right)\right\|^2_{W_h}.$$

In Section 6, we use this example for our simulation study.

Example 2 (Linear factor asset pricing models)**.** Let f^g and f^h be two different vectors of factors (including a constant) used to define stochastic discount factors $f^{g\prime}\theta$ and $f^{h\prime}\beta$ based on linear factor asset pricing models g and h, respectively. Let R be an l-vector of asset returns and $w = \left(f^{g\prime}, f^{h\prime}, R'\right)'$. Asset pricing models can be evaluated by comparing the WE norms of pricing errors $E_{\mu_0} R\left(f^{g\prime}\theta\right) - 1_l$ and $E_{\mu_0} R\left(f^{h\prime}\beta\right) - 1_l$, where 1_l is the l-vector of ones (see, e.g., Hansen and Jagannathan (1997), Kan and Robotti (2009), and Kan, Robotti, and Shanken (2009)). In this case, the moment functions for models g and h are defined as $g(w,\theta) = R\left(f^{g\prime}\theta\right) - 1_l$ and $h(w,\beta) = R\left(f^{h\prime}\beta\right) - 1_l$, respectively, and the null hypothesis for the model comparison can be written as

$$H_0 : \min_{\theta\in\Theta} \left\| E_{\mu_0} R\left(f^{g\prime}\theta\right) - 1_l \right\|_W^2 = \min_{\beta\in B} \left\| E_{\mu_0} R\left(f^{h\prime}\beta\right) - 1_l \right\|_W^2 .$$

Alternatively, Kitamura and Stutzer (2002) suggested to compare the asset pricing models by the KLIC, where the null hypothesis for the model comparison can be written as

$$H_0 : \min_{\theta\in\Theta} \max_{\gamma_g\in\mathbb{R}^{l_g}} -E_{\mu_0} e^{\gamma_g'\{R(f^{g\prime}\theta)-1_l\}} = \min_{\beta\in B} \max_{\gamma_h\in\mathbb{R}^{l_h}} -E_{\mu_0} e^{\gamma_h'\left\{R\left(f^{h\prime}\beta\right)-1_l\right\}}.$$

2.2 Remarks on the choice of measure of fit

A characteristic feature of the model comparison testing is that the researcher needs to specify a measure of fit D to set up the testing problem, and thus the conclusion drawn from the test crucially depends on the choice of D, as was minutely studied by Hall and Pelletier (2011). This subsection clarifies the focus and contribution of the paper and then discusses several issues concerning the choice of D.

First of all, we would like to emphasize that the main question of this paper is not how to choose D, but how to test a given hypothesis *after* the researcher has chosen some D to set up a model comparison testing problem. Therefore, this paper does not develop any specific recommendation on the choice of D. The GNP optimality analysis developed in Section 3 applies to any choice of D as far as the assumptions therein are satisfied. For the illustration purposes, Section 4 considers the case of D_{WE}, and Section 5.3 describes how to extend our approach to the generalized EL-based measures of fit.

However, we would like to point out that to avoid misleading interpretations of the outcome of a model comparison test, it is necessary to report the employed D together with results of the test. In some cases, it can be desirable to consider several different candidates for D to ensure robustness of the conclusion.

To find reasonable candidates for D, in each application the researcher needs to explore the implications of those measures of fit from the economic theory perspective or using some other frameworks such as information theory. For instance, in the asset pricing framework considered in Example 2, it is a common practice to use the Hansen-Jagannathan measure of fit, which is based on the minimized WE norm $\min_{\theta\in\Theta} \left\| E_{\mu_0} R\left(f^{g\prime}\theta\right) - 1_l \right\|_W^2$ with the weight matrix $W = \left(E_{\mu_0} RR'\right)^{-1}$. As discussed in Hansen and Jagannathan (1997), this choice corresponds to the case where the difference between the true and proxy stochastic discount factors is evaluated using a quadratic loss function. Thus, the Hansen-Jagannathan measure of fit may not be appropriate if the econometrician's preferences are represented by some other loss function. For example, Kan, Robotti, and Shanken (2009) considered several choices of W including the unit matrix.[6] Also, an alternative KLIC-based measure of fit for stochastic discount factors was proposed in Kitamura and Stutzer (2002).

If economic theory is not informative enough to specify an appropriate choice of D, the researcher may adopt one of the popular measures of fit considered in the information theory and statistics literature. The KLIC is one such example. One can also consider the class of generalized EL functions (Newey and Smith, 2004):

$$D_\rho (g, \mu_0) = \min_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}^{l_g}} E_{\mu_0} \rho \left(\gamma' g (w, \theta) \right), \tag{4}$$

where $\rho(\cdot)$ is a criterion function, such as $\rho(v) = \log(1-v)$ (EL), $\rho(v) = -(1+v)^2/2$ (Euclidean likelihood or continuous updating GMM), and $\rho(v) = -e^v$ (exponential tilting). Note that the KLIC-based measure of fit of Kitamura and Stutzer (2002) is included here as a special case. Although these measures of fit do not involve weight matrices as in the case of D_{WE}, the researcher still needs to choose the criterion function ρ. Therefore, the caveat pointed out in Hall and Pelletier (2011) still applies: conclusions of the model comparison test, in general, depend on the choice of ρ.

Hereafter, we assume that the researcher has already deliberately chosen a measure of fit D and wishes to find a test having desirable statistical properties. Section 3 below derives a GNP δ-optimal test. Sections 4.1 and 4.2 propose a feasible approximation to the GNP δ-optimal test and study its asymptotic properties when the WE measure of fit with a known weight matrix W is used. Section 4.3 considers the case where W has to be estimated as in the example of the Hansen-Jagannathan measure of fit. Section 5.3 discusses the case where the measure of fit is defined by a generalized EL function.

3 GNP optimal test

We now address the issue of optimal model comparison testing under a chosen measure of fit. Among several optimality criteria of statistical tests (see, for example, Serfling (1980, Chapter 12)), we adopt the GNP optimality criterion developed by Zeitouni and Gutman (1991) and Kitamura (2001), among others. The GNP optimality criterion focuses on global properties of a test, in particular asymptotic behaviors of error probabilities under fixed non-local data generating distributions (in contrast to Pitman-type drifting distributions).

[6] We would like to emphasize here that in the context of model comparison, the choice of a weight matrix in the WE norm represents the choice of a loss function used to evaluate violations of moment restrictions and should be discerned from optimal weighting used to achieve asymptotic efficiency in the usual inference framework.

Under fixed data generating distributions, the type I and II error probabilities of a test with an adequate critical value typically decrease to zero at exponential rates. The GNP optimality criterion compares the decay rate of the type II error probability under some restriction on the decay rate of the type I error probability.

To formalize the notion of the GNP optimality, we need some notation. Let μ_n be the empirical measure based on the sample $\{w_i\}_{i=1}^n$ and

$$\mathcal{P}_0 = \{\mu \in \mathcal{M} : D(g, \mu) = D(h, \mu)\}$$

be the set of measures satisfying the null hypothesis H_0 in (3). Consider a test $\Omega = (\Omega_0, \Omega_1)$ based on μ_n defined by the partition (Ω_0, Ω_1) for $\mathcal{M}$, i.e., accept H_0 if $\mu_n \in \Omega_0$ and reject H_0 if $\mu_n \in \Omega_1 = \mathcal{M} \setminus \Omega_0$.[7] Then the type I and II error probabilities are defined as

$$\begin{aligned} &\Pr\{\mu_n \in \Omega_1 : \mu_0\} \quad \text{for} \quad \mu_0 \in \mathcal{P}_0, \\ &\Pr\{\mu_n \in \Omega_0 : \mu_0\} \quad \text{for} \quad \mu_0 \notin \mathcal{P}_0, \end{aligned}$$

respectively. By adapting the original idea of the Neyman-Pearson optimality to the decay rate analogs, the GNP optimality criterion is described as

$$\begin{aligned} &\text{minimize } \lim_{n\to\infty} \tfrac{1}{n} \log \Pr\{\mu_n \in \Omega_0 : P_1\} \quad \text{for each } P_1 \in \mathcal{M} \setminus \mathcal{P}_0, \\ &\text{subject to } \sup_{P_0 \in \mathcal{P}_0} \lim_{n\to\infty} \tfrac{1}{n} \log \Pr\{\mu_n \in \Omega_1 : P_0\} \le -\alpha. \end{aligned} \tag{5}$$

To analyze these decay rates of the error probabilities, we can apply the large deviation theory for the empirical measure. In particular, Sanov's theorem is useful for our purpose. Let $D_L(\mu, \nu)$ be the Lévy metric between $\mu \in \mathcal{M}$ and $\nu \in \mathcal{M}$, that is

$$D_L(\mu, \nu) = \inf\{\epsilon > 0 : F_\mu(w - \epsilon 1_q) - \epsilon \le F_\nu(w) \le F_\mu(w - \epsilon 1_q) + \epsilon \quad \text{for all } w \in \mathbb{R}^q\},$$

where F_μ and F_ν are the distribution functions of μ and ν, respectively.

Theorem 1 (Sanov)**.** *Suppose that $\{w_i\}_{i=1}^n$ is an i.i.d. sample from $\mu_0 \in \mathcal{M}$. Then its empirical measure μ_n satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{\mu_n \in G : \mu_0\} \le -\inf_{\nu \in G} I(\nu \| \mu_0),$$

for any closed set $G \subset \mathcal{M}$ with respect to the Lévy metric, and

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{\mu_n \in H : \mu_0\} \ge -\inf_{\nu \in H} I(\nu \| \mu_0),$$

for any open set $H \subset \mathcal{M}$ with respect to the Lévy metric.

[7] We focus on the class of tests defined by a partition for the empirical measure. For example, the conventional RV test statistic, which may be written as $D_{WE}(g, \mu_n) - D_{WE}(h, \mu_n)$, belongs to this class. An analogous argument to Zeitouni and Gutman (1991, Lemma 1) may yield a sufficiency result to restrict on this class of tests.

The proof of Sanov's theorem can be found in Deuschel and Stroock (1989), for example. Sanov's theorem says that the error probabilities written in terms of the empirical measure are determined by the KLIC between the data generating measure μ_0 and the sets of interest G and H. This result is particularly useful for establishing the bounds on the decay rates of the type I and II errors probabilities. On the other hand, Sanov's theorem has some rough nature: we can only obtain the upper (or lower) bound for closed (or open) sets with respect to the Lévy metric. In general, however, the rejection regions defined in terms of the KLIC is not necessarily closed, and this fact makes derivation of the GNP optimality in the sense of (5) very difficult (see Zeitouni and Gutman (1991) and Kitamura (2001) for more discussions). Therefore, we consider a modified version of the GNP optimality, called the GNP δ-optimality.

To define the GNP δ-optimality, we need more notation. Let $B(\mu,\delta) = \{\nu \in \mathcal{M} : D_L(\mu,\nu) < \delta\}$ be an open ball around $\mu \in \mathcal{M}$ with radius $\delta > 0$. For a test $\Omega = (\Omega_0, \Omega_1)$, define the partition $\Omega^\delta = \left(\Omega_0^\delta, \Omega_1^\delta\right)$ with $\Omega_1^\delta = \cup_{\mu\in\Omega_1} B(\mu,\delta)$ and $\Omega_0^\delta = \mathcal{M} \setminus \Omega_1^\delta$. The set Ω_1^δ is often called the δ-blowup (or δ-smoothing) of the critical region Ω_1 by the Lévy ball. The GNP δ-optimality is defined as follows.

Definition 3 (GNP δ-optimality)**.** *A test defined by a partition $\Lambda = (\Lambda_0, \Lambda_1)$, which may depend on δ, is called GNP δ-optimal if for each $\delta > 0$,*

(a) $\sup_{P_0\in\mathcal{P}_0} \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \Lambda_1^\delta : P_0\right\} \le -\alpha$ *for some $\alpha > 0$,*

(b) *for any test $\Omega = (\Omega_0, \Omega_1)$ satisfying*

$$\sup_{P_0\in\mathcal{P}_0} \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \Omega_1^{\bar{\delta}} : P_0\right\} \le -\alpha \quad \textit{for some } \bar{\delta} > \delta,$$

it holds that for all $P_1 \in \mathcal{M} \setminus \mathcal{P}_0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \mathcal{M} \setminus \Lambda_1^\delta : P_1\right\} \le \liminf_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \mathcal{M} \setminus \Omega_1^\delta : P_1\right\}.$$

From Theorem 1, we can expect that a test based on the KLIC between the set of measures $\mathcal{P}_0$ satisfying H_0 and the empirical measure μ_n would enjoy the GNP optimal property. Based on Zeitouni and Gutman (1991), we consider the (δ-dependent) KLIC-based test $\Lambda_\delta = (\Lambda_{0,\delta}, \Lambda_{1,\delta})$:

$$\begin{aligned} &\text{accept } H_0 \text{ if } \mu_n \in \Lambda_{0,\delta} = \left\{\nu \in \mathcal{M} : \inf_{\mu\in\mathcal{P}_0} \inf_{\nu'\in cl(B(\nu,c\delta))} I\left(\nu' \| \mu\right) \le \alpha\right\}, \\ &\text{reject } H_0 \text{ if } \mu_n \in \Lambda_{1,\delta} = \mathcal{M} \setminus \Lambda_{0,\delta}, \end{aligned}$$

for some $c > 1$. In other words, we test H_0 by the test statistic

$$T_{n,\delta} = \inf_{\mu\in\mathcal{P}_0} \inf_{\nu\in cl(B(\mu_n,c\delta))} I(\nu\|\mu), \tag{6}$$

with the critical value α. The following theorem establishes the GNP δ-optimality of this KLIC-based test Λ_δ.

Theorem 2 (GNP δ-optimal test)**.** *Suppose that $\{w_i\}_{i=1}^n$ is i.i.d. and the set $\{\nu \in \mathcal{M} : \inf_{\mu \in \mathcal{P}_0} I(\nu \| \mu) \leq \alpha\}$ is compact with respect to the Lévy metric. Then the KLIC-based test Λ_δ is GNP δ-optimal to test the model comparison hypothesis H_0 against H_1.*

Remarks. (a) The i.i.d. assumption on the sample $\{w_i\}_{i=1}^n$ is a major limitation. Although this assumption is often reasonable for the models based on cross-section data such as instrumental variable regression models in Example 1, many applications such as asset pricing models discussed in Example 2 involves time series data. This assumption is required to apply Sanov's theorem to control large deviation behaviors of the empirical measure. Under weakly dependent data, large deviation properties of the empirical measure can be analyzed by Gärtner-Ellis' theorem (Dembo and Zeitouni, 1998, Theorem 2.3.6), where the exponential convergence rate is characterized by the long-run limit of the moment generating function instead of the KLIC. Nevertheless, it is not clear how to apply our technical argument to dependent data. For example, it is not clear what kind of topology should be employed for stochastic processes. Even for simpler setups such as parametric models, we are not aware of any GNP optimality analysis. We note that even though we lose a rationale from the GNP optimality, the approximate test statistics in the next Section can be generalized to dependent data.

(b) To prove the GNP optimality of a test in the sense of (5) by Sanov's theorem, one needs closedness of the rejection region $\{\nu \in \mathcal{M} : \inf_{\mu \in \mathcal{P}_0} I(\nu \| \mu) \geq \alpha\}$ which is generally not true (see, Zeitouni and Gutman, 1991, p. 287). For the GNP δ-optimality, we impose a weaker condition that the set $\{\nu \in \mathcal{M} : \inf_{\mu \in \mathcal{P}_0} I(\nu \| \mu) \leq \alpha\}$ is compact. This compactness condition is easier to verify and holds if $\inf_{\mu \in \mathcal{P}_0} I(\nu \| \mu)$ is lower semicontinuous in ν under the Lévy metric, for example.

(c) The compactness condition on the set $\{\nu \in \mathcal{M} : \inf_{\mu \in \mathcal{P}_0} I(\nu \| \mu) \leq \alpha\}$ restricts the form of the null hypothesis $\mathcal{P}_0$ (not only the forms of the moment functions g and h, but also the form of D). For example, suppose that $D(g, \mu_0)$ and $D(h, \mu_0)$ are continuous in μ_0 under the Lévy metric, which is satisfied if g and h are bounded and the WE measure of fit D_{WE} in (2) with a known weight matrix is adopted. In this case, an application of the maximum theorem (Leininger, 1984) combined with the lower semicontinuity of the KLIC (Chaganty and Karandikar, 1996) implies the lower semicontinuity of $\inf_{\mu \in \mathcal{P}_0} I(\nu \| \mu)$ in ν under the Lévy metric, which in turn implies the compactness of $\{\nu \in \mathcal{M} : \inf_{\mu \in \mathcal{P}_0} I(\nu \| \mu) \leq \alpha\}$ under the Lévy metric. Note that the KLIC-based measure of fit $D_{KLIC}(g, \mu_0)$ or $D_{KLIC}(h, \mu_0)$ does not necessarily satisfy the continuity of μ_0 even if g and h are bounded.

(d) Although the GNP δ-optimality is a weaker notion of optimality than the original Neyman-Pearson or the GNP optimality in the sense of (5), this theorem is insightful: the test statistic $T_{n,\delta}$ should be constructed by taking the minimum KLIC between the space $\mathcal{P}_0$ and the closed Lévy ball $cl(B(\mu_n, 2\delta))$ around the empirical measure μ_n.

(e) Note that the second inequality in Definition 3 (b) is a weak one. Thus, similar to other optimality or admissibility statements, the GNP δ-optimality is silent about the uniqueness of the optimal test. Along with our KLIC-based test Λ^{δ}, there may exist other GNP δ-optimal tests.

(f) An obvious limitation of this theorem is the fact that both the optimal test Λ^{δ} and alternative test Ω^{δ} depend on the blowup constant δ. For the optimal test Λ^{δ}, we can apply a similar argument to Zeitouni and Gutman (1991, Corollary 3) and construct a positive and monotone decreasing sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with $\delta_n \to 0$ such that the n-dependent test $\left\{\Lambda^{\delta_n}\right\}_{n\in\mathbb{N}}$ satisfies the GNP δ-optimality. On the other hand, for the alternative test Ω^{δ}, suppose that the test Ω^{δ} is "regular" in the sense of Zeitouni and Gutman (1991), i.e.,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\mu_n\in\Omega_1^{\delta}:P_0\right\}=\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\mu_n\in\Omega_1:P_0\right\},$$

for each $P_0\in\mathcal{P}_0$, which is satisfied when $\inf_{\mu\in int(\Omega_1)} I\left(\mu\|P_0\right)=\inf_{\mu\in cl(\Omega_1)} I\left(\mu\|P_0\right)$ (see, Zeitouni and Gutman, 1991, Lemma 4). Then we can replace the blowup critical regions $\Omega_1^{\bar{\delta}}$ and Ω_1^{δ} in Definition 3 with the original one Ω_1.

(g) We note that the same argument applies to the case of one-sided testing, i.e., H_0 or $H_{0g}:\mu_0\in\mathcal{P}_{0,g}=\{\mu\in\mathcal{M}:D(g,\mu)\leq D(h,\mu)\}$ against $H_h:\mu_0\in\{\mu\in\mathcal{M}:D(g,\mu)> D(h,\mu)\}$ (the roles of models g and h can be interchanged). In this case, as far as the set $\left\{\nu\in\mathcal{M}:\inf_{\mu\in\mathcal{P}_{0g}} I\left(\nu\|\mu\right)\leq\alpha\right\}$ is compact with respect to the Lévy metric, the same technical argument goes through and the test statistic $T^{g}_{n,\delta}=\inf_{\mu\in\mathcal{P}_{0g}}\inf_{\nu\in cl(B(\mu_n,c\delta))} I\left(\nu\|\mu\right)$ yields the GNP δ-optimal test for H_{0g} against H_h. Moreover, since $\mathcal{P}_0\subset\mathcal{P}_{0,g}$, this optimality result in turn implies that $T^{g}_{n,\delta}$ yields the GNP δ-optimal test for H_0 against H_h as well. Section 5.2 proposes a feasible approximation to $T^{g}_{n,\delta}$.

4 Approximation of the optimal test

As discussed in the previous section, a test based on the statistic $T_{n,\delta}$ enjoys the GNP δ-optimality property. However, in practice it is difficult to compute $T_{n,\delta}$ due to the δ-blowup in its definition. In this section, we propose a feasible approximation to $T_{n,\delta}$ and study its statistical properties.

To simplify the notation, hereafter let E be the expectation under the data generating measure μ_0, $g_i(\theta)=g(w_i,\theta)$, $h_i(\beta)=h(w_i,\beta)$, $\bar{g}(\theta)=\frac{1}{n}\sum_{i=1}^{n} g_i(\theta)$, and $\bar{h}(\beta)=\frac{1}{n}\sum_{i=1}^{n} h_i(\beta)$.

4.1 Construction of the approximate test statistic

As an illustration of our approach, we first consider the case where the (two-sided) model comparison hypothesis is written by the WE measure of fit, i.e.,

$$\mathcal{P}_0=\{\mu\in\mathcal{M}:\min_{\theta\in\Theta}\|Eg_i(\theta)\|^2_{W_g}=\min_{\beta\in B}\|Eh_i(\beta)\|^2_{W_h}\}$$

for some known matrices W_g and W_h. The first step for approximation of the optimal test statistic $T_{n,\delta}$ in (6) is to remove the δ-blowup and focus on the statistic

$$T_n = \inf_{\mu \in \mathcal{P}_0} I\left(\mu_n \| \mu\right), \tag{7}$$

i.e., take the infimum of the KLIC from the empirical measure μ_n to the set $\mathcal{P}_0$.[8] Similarly to the construction of EL statistics, we consider the following discretized analog of (7):

$$\min_{\{p_i\}_{i=1}^n} -\frac{1}{n}\sum_{i=1}^n \log\left(np_i\right), \tag{8}$$

$$\text{s.t.} \quad p_i > 0, \quad \sum_{i=1}^n p_i = 1, \quad \min_{\theta\in\Theta}\left\|\sum_{i=1}^n p_i g_i\left(\theta\right)\right\|_{W_g}^2 = \min_{\beta\in B}\left\|\sum_{i=1}^n p_i h_i\left(\beta\right)\right\|_{W_h}^2.$$

Although this minimization problem looks similar to that of EL, we cannot directly apply the standard implementation and asymptotic theory of EL because of the following two reasons: (i) the last constraint in the above minimization problem is nonlinear in the weights p_i's; and (ii) in the last constraint, we need to evaluate infimum with respect to θ and β for each possible choice of p_i's.

The next step is to find a more practical and technically tractable approximation to the minimization problem in (8). Let us set θ and β in the last constraint of (8) to their population pseudo-true values:

$$\theta^* = \arg\min_{\theta\in\Theta}\left\|Eg_i\left(\theta\right)\right\|_{W_g}^2, \quad \beta^* = \arg\min_{\beta\in B}\left\|Eh_i\left(\beta\right)\right\|_{W_h}^2. \tag{9}$$

Then the minimization problem in (8) reduces to that of the conventional EL for a smooth function of means (Hall and La Scala, 1990):

$$T_n^{*A} = \min_{\left\{\eta_g\in\mathbb{R}^{l_g}, \eta_h\in\mathbb{R}^{l_h} : \|\eta_g\|_{W_g}^2 = \|\eta_h\|_{W_h}^2\right\}} \ell^*\left(\eta_g, \eta_h\right), \tag{10}$$

where

$$\ell^*\left(\eta_g, \eta_h\right) = \min_{\{p_i\}_{i=1}^n} -\frac{1}{n}\sum_{i=1}^n \log\left(np_i\right), \tag{11}$$

$$\text{s.t.} \quad p_i > 0, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i g_i\left(\theta^*\right) = \eta_g, \quad \sum_{i=1}^n p_i h_i\left(\beta^*\right) = \eta_h.$$

[8] When μ_0 has finite support, this approximate statistic T_n is GNP optimal and no smoothing is required (see, Zeitouni and Gutman, 1991, Section II). The GNP optimality of the likelihood ratio test in multinomial models is established by Hoeffding (1965).

Note that the above formulation of EL involves two optimization problems with (11) nested into (10). Note also that η_g and η_h are the running arguments for the minimization in (10). The null of model equivalence is imposed through the constraint $\|\eta_g\|^2_{W_g} = \|\eta_h\|^2_{W_h}$ in (10). The constraints in (11) are now linear in p_i's, and $\ell^*(\eta_g, \eta_h)$ has a convenient dual representation which involves optimization only with respect to $l_g + l_h$ variables:

$$\ell^*(\eta_g, \eta_h) = -\max_{\lambda_g \in \mathbb{R}^{l_g}, \lambda_h \in \mathbb{R}^{l_h}} \frac{1}{n} \sum_{i=1}^{n} \log\left(1 + \lambda_g' (g_i(\theta^*) - \eta_g) + \lambda_h' (h_i(\beta^*) - \eta_h)\right).$$

The last step is to derive a feasible test statistic. By replacing θ^* and β^* in $\ell^*(\eta_g, \eta_h)$ with their empirical analogs

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \|\bar{g}(\theta)\|^2_{W_g}, \hat{\beta} = \arg\min_{\beta \in B} \left\|\bar{h}(\beta)\right\|^2_{W_h},$$

our approximate test statistic is now defined as

$$T_n^A = \min_{\left\{\eta_g \in \mathbb{R}^{l_g}, \eta_h \in \mathbb{R}^{l_h} : \|\eta_g\|^2_{W_g} = \|\eta_h\|^2_{W_h}\right\}} -2n\ell(\eta_g, \eta_h), \tag{12}$$

where

$$\ell(\eta_g, \eta_h) = -\max_{\lambda_g \in \mathbb{R}^{l_g}, \lambda_h \in \mathbb{R}^{l_h}} \frac{1}{n} \sum_{i=1}^{n} \log\left(1 + \lambda_g' \left(g_i\left(\hat{\theta}\right) - \eta_g\right) + \lambda_h' \left(h_i\left(\hat{\beta}\right) - \eta_h\right)\right). \tag{13}$$

The optimization problems in (12) and (13) must be solved numerically. For example, this can be implemented in MATLAB using a nested structure: employ unconstrained minimization procedure *fminunc* for (13) and employ constrained minimization procedure *fmincon* for (12).[9] The following natural starting values can be used for minimization: $\bar{g}\left(\hat{\theta}\right)$ for η_g and $\bar{h}\left(\hat{\beta}\right)$ for η_h to implement (12), and zero vectors for λ_g and λ_h to implement (13). (Note that under H_0, the solutions for λ_g and λ_h in (13) converge in probability to zero.)

4.2 Asymptotic properties

In this section, we derive the asymptotic properties of the approximate test statistic T_n^A introduced in (12). Observe that the conventional theory for EL (e.g., Hall and La Scala, 1990) implies

$$-2n \min_{\left\{\eta_g \in \mathbb{R}^{l_g}, \eta_h \in \mathbb{R}^{l_h} : \|\eta_g\|^2_{W_g} = \|\eta_h\|^2_{W_h}\right\}} \ell^*(\eta_g, \eta_h) \xrightarrow{d} \chi^2_1 \quad \text{if } \|Eg_i(\theta^*)\|^2_{W_g} = \|Eh_i(\beta^*)\|^2_{W_h}.$$

[9] In MATLAB, *fmincon* allows nonlinear constraints.

Here we show that the asymptotic distribution of the approximate test statistic T_n^A, defined by $\ell(\eta_g, \eta_h)$ instead of $\ell^*(\eta_g, \eta_h)$, is unaffected by preliminary estimation of θ^* and β^* under the model comparison null hypothesis H_0. To this end, we impose the following assumptions.

Assumption 1.

(a) $\{w_i\}_{i=1}^n$ *is i.i.d.*

(b) *Model g is misspecified and $\|Eg_i(\theta)\|_{W_g}^2$ has a unique minimum at $\theta^* \in int(\Theta)$; model h is misspecified and $\|Eh_i(\beta)\|_{W_h}^2$ has a unique minimum at $\beta^* \in int(B)$. Θ and B are compact.*

(c) *$g_i(\theta)$ is twice continuously differentiable on Θ almost surely; $h_i(\beta)$ is twice continuously differentiable on B almost surely.*

(d) $\left(E\frac{\partial g_i(\theta)}{\partial \theta'}\right)' W_g \left(E\frac{\partial g_i(\theta)}{\partial \theta'}\right) + \left(I_{p_g} \otimes (W_g E g_i(\theta))\right)' \left(E\frac{\partial}{\partial \theta'} vec\left(\frac{\partial g_i(\theta)}{\partial \theta'}\right)\right)$ *is nonsingular in a neighborhood N_{θ^*} of θ^*;*
$\left(E\frac{\partial h_i(\beta)}{\partial \beta'}\right)' W_h \left(E\frac{\partial h_i(\beta)}{\partial \beta'}\right) + \left(I_{p_h} \otimes (W_h E h_i(\beta))\right)' \left(E\frac{\partial}{\partial \beta'} vec\left(\frac{\partial h_i(\beta)}{\partial \beta'}\right)\right)$ *is nonsingular in a neighborhood N_{β^*} of β^*.*

(e) *For some $\varepsilon > 0$, $\sup_{\theta \in N_{\theta^*}} \|g_i(\theta)\|^{2+\varepsilon}$, $\sup_{\theta \in N_{\theta^*}} \left\|\frac{\partial g_i(\theta)}{\partial \theta'}\right\|^2$, and $\sup_{\theta \in N_{\theta^*}} \left\|\frac{\partial}{\partial \theta'} vec\left(\frac{\partial g_i(\theta)}{\partial \theta'}\right)\right\|$ as well as $\sup_{\beta \in N_{\beta^*}} \|h_i(\beta)\|^{2+\varepsilon}$, $\sup_{\beta \in N_{\beta^*}} \left\|\frac{\partial h_i(\beta)}{\partial \beta'}\right\|^2$, and $\sup_{\beta \in N_{\beta^*}} \left\|\frac{\partial}{\partial \beta'} vec\left(\frac{\partial h_i(\beta)}{\partial \beta'}\right)\right\|$ are integrable.*

(f) $E\left[(g_i(\theta^*) - Eg_i(\theta^*))(g_i(\theta^*) - Eg_i(\theta^*))'\right]$ *is positive definite;*
$E\left[(h_i(\beta) - Eh_i(\beta^*))(h_i(\beta) - Eh_i(\beta^*))'\right]$ *is positive definite.*

Assumption 1 (a) excludes dependent data. Although it loses a rationale based on the GNP δ-optimality in Theorem 2, the construction of the test statistic T_n^A itself can be adapted to weakly dependent data. In particular, we can replace the moment functions $g_i(\cdot)$ and $h_i(\cdot)$ used to define T_n^A with their blocked analogs as in Kitamura (1997). Then a modified argument of Kitamura (1997) will yield the asymptotic properties of the test statistic using blocked moments.

Assumption 1 (b) requires uniqueness of the pseudo-true values, which is often assumed in the literature of misspecification analysis (e.g., Vuong, 1989; Kitamura, 2000; Rivers and Vuong, 2002; Hall and Pelletier, 2011). However, it should be acknowledged that this assumption commonly fails in practice. Although it is beyond the scope of this paper, we suggest two directions to relax this assumption. First, Assumption 1 (b) may be generalized by using the notion of "identifiably unique parameters" in Domowitz and White (1982, Definition 2.1), which are not required to converge to a limit (see also Bates and White (1985) for a detailed discussion on consistency). Second, a recent paper by Shi (2009) proposed the notion of "pseudo-true sets" and developed an RV-type test to compare misspecified moment inequality models, where measures of fit are minimized on some sets of parameters. It is interesting to assess how our approach can be adapted to such scenarios.

Assumption 1 (c) is standard for nonlinear models. Assumption 1 (d) requires that the Hessians of $\|Eg_i(\theta)\|^2_{W_g}$ and $\|Eh_i(\beta)\|^2_{W_h}$ are nonsingular in neighborhoods of θ^* and β^*, respectively. This assumption appears, for example, in Hall and Inoue (2003).[10] Assumption 1 (e) assumes that the moment functions g and h are sufficiently smooth in some neighborhoods of θ^* and β^* respectively, and the distribution of the data has sufficiently thin tails; they are similar to Kitamura (2000, Assumption 2 (f)). Assumption 1 (f) is similar to Hall and Inoue (2003, Assumption 4), and together with other conditions it ensures that $\sqrt{n}\left(\hat{\theta}-\theta^*\right) = O_p(1)$ and $\sqrt{n}\left(\hat{\beta}-\beta^*\right) = O_p(1)$ (see, e.g., Domowitz and White, 1982).

The asymptotic properties of the approximate test statistic T_n^A to the GNP δ-optimal one $T_{n,\delta}$ (in the case of the WE measure of fit) are described in the following theorem.

Theorem 3. *Suppose that Assumption 1 holds. Then under H_0, $T_n^A \xrightarrow{d} \chi_1^2$. Also, under H_1, $\Pr\left\{T_n^A > c\right\} \to 1$ for any $c > 0$.*

Remarks: (a) Let $\chi^2_{1,\alpha}$ be the $(1-\alpha)$-th quantile of the χ_1^2 distribution. According to Theorem 3, an asymptotic size α model comparison test is defined by the following rule:

accept H_0 if $T_n^A \le \chi^2_{1,\alpha}$,

reject H_0 and select the model g if $T_n^A > \chi^2_{1,\alpha}$ and $\left\|\bar{g}\left(\hat{\theta}\right)\right\|^2_{W_g} \le \left\|\bar{h}\left(\hat{\beta}\right)\right\|^2_{W_h}$,

reject H_0 and select the model h if $T_n^A > \chi^2_{1,\alpha}$ and $\left\|\bar{g}\left(\hat{\theta}\right)\right\|^2_{W_g} > \left\|\bar{h}\left(\hat{\beta}\right)\right\|^2_{W_h}$.

Note that except for the test statistic and its critical value, our test procedure is same as that of RV whose test statistic is based on the contrast $\|\bar{g}\left(\hat{\theta}\right)\|^2_{W_g} - \|\bar{h}\left(\hat{\beta}\right)\|^2_{W_h}$ after normalization with a normal critical value.

(b) An interesting difference with the RV test is that for Theorem 3, the positive definiteness of the variance matrix of $(g_i(\theta^*)', h_i(\beta^*)')$ is not required (even though the variance matrices of $g_i(\theta^*)$ and $h_i(\beta^*)$ are assumed to be positive definite respectively in Assumption 1 (f)). Thus, even in the so-called nested or overlapping cases where $g_i(\theta^*)$ and $h_i(\beta^*)$ share common elements, the null asymptotic distribution of our statistic remains χ_1^2. The reason for this is that, when the rank of the variance of $(g_i(\theta^*)', h_i(\beta^*)')$ is less than $l_g + l_h$, one can express some elements of $(g_i(\theta^*)', h_i(\beta^*)')$ as a linear combination of the remaining elements and reformulate the problem so that the new random vector in the reformulated problem has a positive definite variance matrix. This is one of the important

[10] In contrast to the correctly specified case, the Hessian involves an extra term when the model is misspecified.

advantages of our testing approach over that of RV. In the RV approach, the difference of estimated criterion functions can have a non-standard non-normal asymptotic distribution when the models are nested or overlapping.[11] Therefore, to implement the RV test, one typically has to employ pre-tests or two-step testing approach which introduces additional practical and theoretical complications. Unlike RV, our approach does not require such pre-tests since the asymptotic null distribution does not depend on whether the models are non-nested, nested, or overlapping.

(c) Theorem 3 establishes that estimation of θ^* and β^* does not affect the asymptotic null distribution of the test statistic. The reason for this, as can be seen in the proof of this theorem, is that the stochastic terms in $\ell(\eta_g, \eta_h)$ created by estimation of θ^* and β^* are asymptotically orthogonal to a linear space defined by the null hypothesis. This orthogonality is guaranteed by the first-order conditions for $\hat{\theta}$ and $\hat{\beta}$. A similar phenomenon occurs in the case of the RV test statistic.

(d) When g and h are both correctly specified, i.e., $Eg_i(\theta*) = 0$ and $Eh_i(\beta^*) = 0$, one can show that T_n^A converges to zero in probability (see footnote 13 in the proof of Theorem 3). Hence, H_0 of models equivalence will be accepted with probability approaching one, which is a correct decision, since in this case the models are equivalent.

4.3 Estimated weight matrices

We now consider the situation where the weight matrices W_g and W_h are unknown and estimated by the estimators $\hat{W}_g$ and $\hat{W}_h$, respectively. In this case, our test statistic (denoted by $\hat{T}_n^A$) is defined by replacing W_g and W_h in (12) with $\hat{W}_g$ and $\hat{W}_h$. Based on Hall and Pelletier (2011), we impose the following assumptions.

Assumption 2.

(a) *$\hat{W}_g$ and $\hat{W}_h$ satisfy* $\sqrt{n}\begin{pmatrix} vec\left(\hat{W}_g - W_g\right) \\ vec\left(\hat{W}_h - W_h\right)\end{pmatrix} = A\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i + o_p(1)$, *where* $\{f_i\}_{i=1}^n$ *is an i.i.d. mean-zero k-vector and A is an* $\left(l_g^2 + l_h^2\right) \times k$ *constant matrix.*

(b) $E\begin{pmatrix} g_i(\theta^*) - Eg_i(\theta^*) \\ h_i(\beta^*) - Eh_i(\beta^*) \\ f_i \end{pmatrix}\begin{pmatrix} g_i(\theta^*) - Eg_i(\theta^*) \\ h_i(\beta^*) - Eh_i(\beta^*) \\ f_i \end{pmatrix}' = \begin{pmatrix} \Omega & \Sigma_1 \\ \Sigma_1' & \Sigma_2 \end{pmatrix}$ *is positive definite.*

Assumption 2 (a) says that the estimators $\hat{W}_g$ and $\hat{W}_h$ have asymptotic linear forms. This assumption is typically satisfied if $\hat{W}_g$ and $\hat{W}_h$ are functions of sample means and/or contain $\sqrt{n}$-consistent estimators for nuisance parameters. See Domowitz and White (1982) for primitive conditions needed to derive the $\sqrt{n}$-consistency and asymptotic linear forms for the nuisance parameter estimators. Assumption 2 (b) is a rank condition. This assumption

[11] See also Vuong (1989).

can be relaxed by allowing overlapping elements between $g_i(\theta^*)$ and $h_i(\beta^*)$. In that case, Ω would denote the variance of the elements of $\left(g_i(\theta^*)', h_i(\beta^*)'\right)$ without the duplicates.

Let $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega & \Sigma_1 \\ \Sigma_1' & \Sigma_2 \end{pmatrix}\right)$. The asymptotic properties of $\hat{T}_n^A$ are obtained as follows.

Theorem 4. *Suppose that Assumptions 1 and 2 hold. Under H_0,*

$$\hat{T}_n^A \xrightarrow{d} \|\pi\|_\Omega^{-2}\left(\pi' X_1 + \kappa' A X_2/2\right)^2,$$

where

$$\pi = \begin{pmatrix} W_g E g_i(\theta^*) \\ -W_h E h_i(\beta^*) \end{pmatrix}, \quad \kappa = \begin{pmatrix} E g_i(\theta^*) \otimes E g_i(\theta^*) \\ -E h_i(\beta^*) \otimes E h_i(\beta^*) \end{pmatrix}. \tag{14}$$

Also under H_1, $\Pr\left\{\hat{T}_n^A > c\right\} \to 1$ *for any $c > 0$.*

Remark: The asymptotic null distribution of $\hat{T}_n^A$ is non-standard and depends on the nuisance parameters. However, the critical values for testing H_0 of models equivalence can be obtained through simulations. Let $\hat{\pi}$ and $\hat{\kappa}$ be estimators of π and κ constructed using $\hat{W}_g$, $\hat{W}_h$, $\bar{g}\left(\hat{\theta}\right)$, and $\bar{h}\left(\hat{\beta}\right)$. Let $\hat{\Omega}$, $\hat{\Sigma}_1$, $\hat{\Sigma}_2$, and $\hat{A}$ be consistent estimators of the corresponding matrices. To obtain critical values, one simulates

$$\begin{pmatrix} X_{1r} \\ X_{2r} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\Omega} & \hat{\Sigma}_1 \\ \hat{\Sigma}_1' & \hat{\Sigma}_2 \end{pmatrix}\right)$$

independently across $r = 1, \ldots, R$, and computes $\xi_r = \left(\hat{\pi}' X_{1r} + 0.5\hat{\kappa}' \hat{A} X_{2r}\right)^2 / \|\hat{\pi}\|_{\hat{\Omega}}^2$. Then a simulated critical value for the asymptotic size α test is obtained as the $(1-\alpha)$-th sample quantile of $\{\xi_r\}_{r=1}^R$.

5 Extensions

5.1 Local power property

In this section, we study the local power properties of the proposed test based on T_n^A in (12) and the conventional RV test in a simple setup. In particular, we show that these tests are asymptotically equivalent under certain local alternatives, and argue that it is beneficial to explore approaches different from the conventional local power analysis. Since different tests typically show different global power properties (see, e.g., Hoeffding, 1965), our GNP optimality analysis for global power can be useful to explain the superior finite sample power properties of the approximate test observed in the simulation study in Section 6.

For simplicity, we assume that (i) models g and h are of the same dimension (i.e., $l_g = l_h$); (ii) the weights used to define the WE norms are same (i.e., $W_g = W_h = W$); and

(iii) the models are non-nested (i.e., $\Omega = Var\left(\left(g_i(\theta^*)', h_i(\beta^*)'\right)\right)$ is positive definite). We consider local alternatives (say, μ_{0n}) satisfying

$$E_n g_i(\theta^*) = E_n h_i(\beta^*) + \frac{c}{\sqrt{n}},$$

for some $c \neq 0$, where E_n is the expectation under μ_{0n}. In this case, the deviation from the null hypothesis H_0 is characterized by

$$\|E_n g_i(\theta^*)\|_W^2 - \|E_n h_i(\beta^*)\|_W^2 = \frac{c' W E_n h_i(\beta^*)}{\sqrt{n}} + \frac{\|c\|_W^2}{n}. \tag{15}$$

Thus, as n increases, the measure μ_{0n} approaches the set $\mathcal{P}_0$ (the set of measures consistent with the null hypothesis).

The RV test statistic is defined as

$$d_n = \sqrt{n}\left(\left\|\bar{g}\left(\hat{\theta}\right)\right\|_W^2 - \left\|\bar{h}\left(\hat{\beta}\right)\right\|_W^2\right) / (2\hat{\sigma}), \tag{16}$$

where $\hat{\sigma}$ is a consistent estimator of $\sigma^2 = \pi'\Omega\pi$ and π is defined in (14). Under analogous conditions to Assumption 1, the numerator of d_n satisfies

$$\begin{aligned}\sqrt{n}\left(\left\|\bar{g}\left(\hat{\theta}\right)\right\|_W^2 - \left\|\bar{h}\left(\hat{\beta}\right)\right\|_W^2\right) &= c' W E_n h_i(\beta^*) + 2\pi'\sqrt{n}\begin{pmatrix}\bar{g}\left(\hat{\theta}\right) - E_n g_i(\theta^*) \\ \bar{h}\left(\hat{\beta}\right) - E_n h_i(\beta^*)\end{pmatrix} + o_p(1) \\ &\xrightarrow{d} N\left(c' W \lim_{n\to\infty} E_n h_i(\beta^*), 4\sigma^2\right).\end{aligned}$$

Therefore, under the local alternatives satisfying (15), we see that

$$d_n^2 \xrightarrow{d} \chi_1^2\left(\left(c' W \lim_{n\to\infty} E_n h_i(\beta^*)\right)^2 / \left(4\sigma^2\right)\right),$$

where $\chi_1^2\left(\lambda^2\right)$ is the noncentral χ^2 distribution with one degree of freedom and the non-centrality parameter λ^2.

We now study the local power property of our approximate test statistic T_n^A. By adapting the proof of Theorem 3, the constraint $\|\eta_g\|_{W_g} - \|\eta_h\|_{W_h} = 0$ in the definition of T_n^A in (12) can be linearized as

$$2\pi'\begin{pmatrix} E_n g_i(\theta^*) - \eta_g \\ E_n h_i(\beta^*) - \eta_h \end{pmatrix} - \frac{c' W E_n h_i(\beta^*)}{\sqrt{n}} = o\left(n^{-1/2}\right).$$

Then, as in the proof of Theorem 3,

$$T_n^A = \min_y \|Z_n + \delta_n - y\|^2 + o_p(1) \quad \text{s.t. } \frac{1}{2} c' W E_n h_i(\beta^*) + \nu' y = 0$$

$$\xrightarrow{d} \left(\frac{\nu' Z}{\|\nu\|} + \frac{c' W \lim_{n\to\infty} E_n h_i(\beta^*)}{2\|\nu\|} \right)^2,$$

where Z_n, δ_n, and ν are defined in (25) in Appendix A.2, and $Z \sim N\left(0, I_{l_g+l_h}\right)$. Therefore, under the local alternatives satisfying (15), we see that

$$T_n^A \xrightarrow{d} \chi_1^2 \left(\left(c' W \lim_{n\to\infty} E_n h_i(\beta^*) \right)^2 / \left(4\|\nu\|^2 \right) \right).$$

Since $\nu = \Omega^{1/2}\pi$, it follows that the statistics d_n^2 and T_n^A are asymptotically equivalent under the local alternatives satisfying (15).

5.2 One-sided test

In this subsection, we describe how our approach can be extended to one-sided testing problems. In the case of one-sided testing, the researcher is interested in testing the null hypothesis that the fit of model g is at least as good as that of model h:

$$H_{0,g} : \|Eg_i(\theta^*)\|_{W_g}^2 \leq \|Eh_i(\beta^*)\|_{W_h}^2,$$

against the alternative hypothesis

$$H_h : \|Eg_i(\theta^*)\|_{W_g}^2 > \|Eh_i(\beta^*)\|_{W_h}^2.$$

Such testing problem can arise if, for example, the researcher is interested in showing that a newly proposed model h has a better fit than some benchmark model g. By applying the same argument as in Section 4.1 to the null hypothesis $H_{0,g}$, an approximate test statistic can be defined as

$$T_n^{A,g} = \min_{\left\{\eta_g\in\mathbb{R}^{l_g}, \eta_h\in\mathbb{R}^{l_h} : \|\eta_g\|_{W_g}^2 \leq \|\eta_h\|_{W_h}^2\right\}} -2n\ell(\eta_g, \eta_h),$$

where the function $\ell(\eta_g, \eta_h)$ is defined in (13).

The asymptotic behavior of $T_n^{A,g}$ can be described using the results of Gourieroux, Holly, and Monfort (1982). Let $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ denote the distribution of a mixed χ^2 random variable, where χ_0^2 denotes the point mass distribution at zero. We have the following result.

Theorem 5. *Suppose that Assumption 1 holds, and let $V \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. Then, under $H_{0,g}$, $\lim_{n\to\infty} \Pr\left\{T_n^{A,g} > c\right\} \leq \Pr\{V > c\}$ for any $c > 0$. Furthermore, under H_h, $\lim_{n\to\infty} \Pr\left\{T_n^{A,g} > c\right\} = 1$ for any $c > 0$.*

Let c_α be the $(1-\alpha)$-th quantile of $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. By this theorem, the test that rejects $H_{0,g}$ when $T_n^{A,g} > c_\alpha$ has the asymptotic size α.

5.3 Test for generalized EL-based measure of fit

Our testing approach is not limited to the case of the WE norm-based measures of fit. For example, suppose that we are interested in model comparison using the generalized EL-based measure of fit in (4), i.e.,

$$H_0 : \min_{\theta\in\Theta} \max_{\gamma_g\in\mathbb{R}^{l_g}} E\rho\left(\gamma_g' g_i(\theta)\right) = \min_{\beta\in B} \max_{\gamma_h\in\mathbb{R}^{l_h}} E\rho\left(\gamma_h' h_i(\beta)\right). \tag{17}$$

A discretized analog of (7) can be written as

$$\min_{\{p_i\}_{i=1}^n} -\frac{1}{n}\sum_{i=1}^n \log(np_i), \tag{18}$$

$$\text{s.t.} \quad p_i > 0, \quad \sum_{i=1}^n p_i = 1, \quad \min_{\theta\in\Theta} \max_{\gamma_g\in\mathbb{R}^{l_g}} \sum_{i=1}^n p_i\rho\left(\gamma_g' g_i(\theta)\right) = \min_{\beta\in B} \max_{\gamma_h\in\mathbb{R}^{l_h}} \sum_{i=1}^n p_i\rho\left(\gamma_h' h_i(\beta)\right).$$

As in the case of the WE measure of fit, this formulation is not practical because the last restriction is nonlinear in p_i's due to the minimax component. After fixing the parameters at $\hat\theta_\rho = \arg\min_{\theta\in\Theta}\max_{\gamma_g}\frac{1}{n}\sum_{i=1}^n \rho\left(\gamma_g' g_i(\theta)\right)$, $\hat\beta_\rho = \arg\min_{\beta\in B}\max_{\gamma_h}\frac{1}{n}\sum_{i=1}^n \rho\left(\gamma_h' h_i(\beta)\right)$, $\hat\gamma_g = \arg\max_{\gamma_g}\frac{1}{n}\sum_{i=1}^n \rho\left(\gamma_g' g_i\left(\hat\theta_\rho\right)\right)$, and $\hat\gamma_h = \arg\max_{\gamma_h}\sum_{i=1}^n p_i\rho\left(\gamma_h' h_i\left(\hat\beta_\rho\right)\right)$, the dual problem of (18) yields the following approximate test statistic for the null hypothesis in (17):

$$T_{n,\rho}^A = -2\max_{\lambda\in\mathbb{R}}\sum_{i=1}^n \log\left(1+\lambda\left(\rho\left(\hat\gamma_g' g_i\left(\hat\theta_\rho\right)\right) - \rho\left(\hat\gamma_h' h_i\left(\hat\beta_\rho\right)\right)\right)\right).$$

Similarly to the WE case, one can show that $T_{n,\rho}^A \xrightarrow{d} \chi_1^2$ provided that H_0 in (18) is true and $\Pr\left\{\rho\left(\gamma_g^{*\prime} g_i\left(\theta_\rho^*\right)\right) \neq \rho\left(\gamma_h^{*\prime} h_i\left(\beta_\rho^*\right)\right) : \mu_0\right\} > 0$, where $\theta_\rho^*, \beta_\rho^*, \gamma_g^*$, and γ_h^* denote the population analogs of $\hat\theta_\rho, \hat\beta_\rho, \hat\gamma_g$, and $\hat\gamma_h$ respectively (see Appendix A.5 for a sketch of the proof). One can also show that the test that rejects H_0 in (18) when $T_{n,\rho}^A > \chi_{1,\alpha}^2$ is consistent. Therefore, our approach to construct an approximate test statistic is not confined to the WE norm-based measure of fit.

For example, if we set $\rho(v) = -e^v$, the null hypothesis in (17) is the one considered in Kitamura (2000). While his statistic is based on the difference of the models' criterion functions, our test statistic is constructed using the KLIC between the empirical measure and the set of measures consistent with H_0.

Furthermore it is interesting to discuss the roles of quadratic approximations used to derive the asymptotic null distribution of $T_{n,\rho}^A$. First, a quadratic approximation is applied the logarithm in $T_{n,\rho}^A$ (see (26) in Appendix A.5). This approximation is commonly applied to derive the asymptotic properties of the EL-based statistic (see Owen, 1988; Qin and Lawless, 1994). Second, to derive the $\sqrt{n}$-consistency for $\hat\theta_\rho$ and $\hat\gamma_g$ (also for $\hat\beta_\rho$ and $\hat\gamma_h$) to the population analogs, we typically need a quadratic approximation of

$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\hat{\gamma}_g' g_i\left(\hat{\theta}_\rho\right)\right)$ around $\left(\hat{\gamma}_g,\hat{\theta}_\rho\right) = \left(\gamma_g^*,\theta_\rho^*\right)$ (see Domowitz and White, 1982; Kitamura, 2000). However, due to the first-order conditions for $\left(\gamma_g^*,\theta_\rho^*\right)$ (see (27) in Appendix A.5), the estimation error $\sqrt{n}\left(\hat{\theta}_\rho-\theta_\rho^*\right)$ does not contribute to the first-order asymptotics for $T_{n,\rho}^A$. Indeed, the convergence of $T_{n,\rho}^A$ to the χ_1^2 distribution is induced by the central limit theorem to $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\rho\left(\gamma_g^{*\prime}g_i\left(\theta_\rho^*\right)\right)-\rho\left(\gamma_h^{*\prime}h_i\left(\beta_\rho^*\right)\right)\right\}$ under H_0 implying $E\left[\rho\left(\gamma_g^{*\prime}g_i\left(\theta_\rho^*\right)\right)-\rho\left(\gamma_h^{*\prime}h_i\left(\beta_\rho^*\right)\right)\right] = 0$. For this term, a quadratic approximation is unnecessary (note that γ_g^* and γ_h^* are non-zero under misspecification and $\theta_\rho^*,\beta_\rho^*,\gamma_g^*$, and γ_h^* depend on the criterion function $\rho(\cdot)$).

Finally, we emphasize that the caveat pointed out by Hall and Pelletier (2011) still applies in this context, i.e., different choices of the criterion function ρ may yield different rankings.

6 Monte Carlo simulations

In this section, we evaluate the finite sample performance of our approximate test by simulations. Specifically, we compare the finite sample power properties of the following three tests: (i) approximate test based on T_n^A in (12), (ii) its infeasible version based on T_n^{*A} in (10), and (iii) the conventional RV test in (16).

The data generating process is similar to that of Hall and Pelletier (2007) and is based on the instrumental variable regression model:

$$y_i = x_i + z_{2i} + (1+\gamma)\,z_{4i} + u_i,$$
$$x_i = z_{1i} + z_{2i} + z_{3i} + z_{4i} + v_i,$$

for $i=1,\ldots,n$, where z_{1i}, z_{2i}, z_{3i}, z_{4i}, u_i, and v_i are independent standard normal random variables. We consider two misspecified moment restriction models g and h defined by $g_i(\theta) = (y_i-\theta x_i)(z_{1i},z_{2i})'$ and $h_i(\beta) = (y_i-\beta x_i)(z_{3i},z_{4i})'$. For the model comparison null hypothesis, we consider

$$H_0: \min_\theta \|Eg_i(\theta)\|^2 = \min_\beta \|Eh_i(\beta)\|^2,$$

i.e., the WE measures of fit with the weights $W_g = W_h = I_2$. In this setup, the null H_0 is satisfied when $\gamma=0$. For the alternative H_1, we consider the cases of γ =0.5, 1.0, 1.5, 2.0, 2.5, and 3.0. Note that for $\gamma>0$, the difference $\min_\beta\|Eh_i(\beta)\|^2 - \min_\theta\|Eg_i(\theta)\|^2$ is positive and increasing in γ. Thus, γ can be viewed as a parameter controlling the discrepancy between the null and alternative hypotheses. The sample size is $n=100$.

For each Monte Carlo repetition, we compute three test statistics: T_n^A in (12), its infeasible version T_n^{*A} in (10), and the RV test statistic in (16). To compute the infeasible statistic T_n^{*A}, we use the knowledge of $\theta^*=1.5$ and $\beta^*=1.5+0.5\gamma$ in this setup. To compute the RV statistic, we use $\hat{\sigma}^2 = \left\| \left(\bar{g}\left(\hat{\theta}\right)', -\bar{h}\left(\hat{\beta}\right)' \right) \right\|^2_{\hat{\Omega}}$, with

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} g_i\left(\hat{\theta}\right) - \bar{g}\left(\hat{\theta}\right) \\ h_i\left(\hat{\beta}\right) - \bar{h}\left(\hat{\beta}\right) \end{pmatrix} \begin{pmatrix} g_i\left(\hat{\theta}\right) - \bar{g}\left(\hat{\theta}\right) \\ h_i\left(\hat{\beta}\right) - \bar{h}\left(\hat{\beta}\right) \end{pmatrix}'.$$

In this setup, we have $d_n \stackrel{d}{\to} N(0,1)$ under H_0 because models g and h are non-nested in the sense of RV and satisfy their regularity conditions.

The size and power results (non-size-adjusted) based on 10,000 simulation repetitions are reported in Table 1 on page 23. According to the results for $\gamma = 0$ (H_0 is true), the RV test is under-sized, while the rejection frequencies of our infeasible and feasible tests are very close to the nominal levels. For $\gamma > 0$ (H_0 is false), the rejection frequencies of our tests are substantially higher than those of the RV test. Our tests are especially more powerful when the significance level is low. For example, when $\gamma = 2.0$ and $\alpha = 0.01$, the rejection frequencies of our tests are about 76%, while the RV test rejects H_0 only in 41% of the simulations. These simulation results are in agreement with our theoretical findings and confirm superior power properties of our approach. The simulations also show that estimation of unknown parameters has no significant effect on the size and power properties of our tests: in all cases, the rejection frequencies of the infeasible and feasible tests are remarkably close.

The size-adjusted power results are reported in Table 2 on page 24. These results are based on 100,000 simulation repetitions. To compute size-adjusted critical values, first, we generated 100,000 values of each of the statistics under H_0 ($\gamma = 0$). The size-adjusted critical value for a size α test is the $1 - \alpha$ quantile of the resulting empirical distribution.

When $\gamma = 0.5$, the three tests have same power as the differences in simulated rejection probabilities are below the precision of the simulations.[12] In all other cases, the differences in simulated rejection rates between the RV test and our tests are statistically significant, and the tests proposed in this paper dominate the RV test. The superior power properties of our approach are especially apparent in the case of larger values of γ and $\alpha = 0.01$. For example, when $\gamma = 2.5$ the 1% RV test rejects with probability 75.7%, while the simulated rejection probabilities of our infeasible and feasible tests are 80.3% and 80.7% respectively.

While the differences in size-adjusted power between our tests and the RV test are less stark than in the case of non-size-adjusted power, we would like to emphasize that the size adjustment is infeasible in practice and the non-size-adjusted results in Table 1 give a more accurate depiction of the tests performance in real applications. Given the fact that the size of our tests is very close to nominal, one might expect a substantial power gain with minimal size distortions by adopting the EL approach.

[12] When the simulated rejection probability is $\hat{\pi}$ and the number of simulation repetitions is R, the standard error for $\hat{\pi}$ is given by $\sqrt{\hat{\pi}(1-\hat{\pi})/R}$. Thus, in the case of $\alpha = 0.05$ and $\gamma = 0.05$, the standard error for the simulated power of the RV test is 0.0014.

Table 1: Rejection frequencies of the tests based on RV, T_n^{*A}, and T_n^A for different significance levels α and different values of γ (using 10,000 simulation repetitions)

α	RV test	Infeasible test (T_n^{*A})	Feasible test (T_n^A)
		$\gamma = 0.0$	
0.10	0.0717	0.1062	0.1092
0.05	0.0268	0.0540	0.0568
0.01	0.0013	0.0127	0.0141
		$\gamma = 0.5$	
0.10	0.3348	0.4110	0.4118
0.05	0.1915	0.2937	0.2952
0.01	0.0285	0.1262	0.1272
		$\gamma = 1.0$	
0.10	0.6811	0.7547	0.7501
0.05	0.5013	0.6482	0.6440
0.01	0.1495	0.4081	0.4071
		$\gamma = 1.5$	
0.10	0.8433	0.9001	0.8933
0.05	0.7033	0.8281	0.8267
0.01	0.2969	0.6359	0.6328
		$\gamma = 2.0$	
0.10	0.9096	0.9476	0.9453
0.05	0.8045	0.9054	0.9008
0.01	0.4055	0.7551	0.7580
		$\gamma = 2.5$	
0.10	0.9367	0.9686	0.9694
0.05	0.8546	0.9362	0.9369
0.01	0.4733	0.8164	0.8205
		$\gamma = 3.0$	
0.10	0.9520	0.9777	0.9791
0.05	0.8825	0.9529	0.9561
0.01	0.5176	0.8503	0.8583

Table 2: Size-adjusted power of the tests based on RV, T_n^{*A}, and T_n^A for different significance levels α and different values of γ (using 100,000 simulation repetitions)

α	RV test	Infeasible test (T_n^{*A})	Feasible test (T_n^A)
		$\gamma = 0.5$	
0.10	0.3993	0.3987	0.3979
0.05	0.2814	0.2795	0.2795
0.01	0.1114	0.1115	0.1132
		$\gamma = 1.0$	
0.10	0.7386	0.7507	0.7418
0.05	0.6227	0.6333	0.6284
0.01	0.3723	0.3837	0.3869
		$\gamma = 1.5$	
0.10	0.8809	0.8974	0.8893
0.05	0.8032	0.8238	0.8153
0.01	0.5778	0.6101	0.6112
		$\gamma = 2.0$	
0.10	0.9348	0.9482	0.9436
0.05	0.8786	0.9015	0.8958
0.01	0.6918	0.7367	0.7365
		$\gamma = 2.5$	
0.10	0.9582	0.9681	0.9664
0.05	0.9135	0.9351	0.9330
0.01	0.7537	0.8029	0.8066
		$\gamma = 3.0$	
0.10	0.9689	0.9765	0.9769
0.05	0.9329	0.9512	0.9517
0.01	0.7892	0.8388	0.8463

7 Conclusion

In this paper, we study global optimality in model comparison hypothesis testing for misspecified unconditional moment restriction models. Based on the generalized Neyman-Pearson optimality criterion, which focuses on the decay rates of the type I and II error probabilities under fixed distributions, we find an optimal test statistic that is defined by the Kullback-Leibler information criterion. We then propose a feasible approximation to the optimal test, and study its asymptotic properties for some examples. Simulation results show that our test has excellent finite sample properties and is more powerful than the existing Rivers-Vuong test.

A Proofs

A.1 Proof of Theorem 2

First, we check Definition 3 (a). Without loss of generality, we set as $c = 2$ in (6). Pick any $\delta > 0$ and $P_0 \in \mathcal{P}_0$. We start by showing that for each $\delta' \in (0, \delta/2)$,

$$cl\left(\Lambda_{1,\delta}^{\delta}\right) \subset \Lambda_{1,\delta'}^{\delta'}. \tag{19}$$

Pick any and $\nu \in cl\left(\Lambda_{1,\delta}^{\delta}\right)$. It is sufficient for (19) to show that

$$\inf_{\mu \in \mathcal{P}_0} I\left(\nu' \| \mu\right) > \alpha \quad \text{for each } \nu' \in cl\left(B\left(\nu, 2\delta'\right)\right). \tag{20}$$

Since $\nu \in cl\left(\Lambda_{1,\delta}^{\delta}\right)$, there exists $\omega \in \mathcal{M}$ such that $D_L(\nu, \omega) \leq \delta + (\delta - 2\delta')/2$ and $\inf_{\mu \in \mathcal{P}_0} I(\omega' \| \mu) > \alpha$ for each $\omega' \in cl(B(\omega, 2\delta))$. Thus, it is sufficient for (20) to show that $\nu' \in cl(B(\omega, 2\delta))$ for each $\nu' \in cl(B(\nu, 2\delta'))$. This can be shown by the triangle inequality:

$$D_L\left(\nu', \omega\right) \leq D_L\left(\nu', \nu\right) + D_L\left(\nu, \omega\right) \leq 2\delta' + \delta + \left(\delta - 2\delta'\right)/2 < 2\delta,$$

for each $\nu' \in cl(B(\nu, 2\delta'))$. Therefore, we obtain (19). Now, observe that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \Lambda_{1,\delta}^{\delta} : P_0\right\} \leq \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in cl\left(\Lambda_{1,\delta}^{\delta}\right) : P_0\right\} \\
&\leq - \inf_{P \in cl\left(\Lambda_{1,\delta}^{\delta}\right)} I\left(P \| P_0\right) \leq - \inf_{P \in \Lambda_{1,\delta'}^{\delta'}} I\left(P \| P_0\right) \leq -\alpha,
\end{aligned}$$

where the first inequality follows from a set inclusion relationship, the second inequality follows from Sanov's theorem, the third inequality follows from (19), and the last inequality follows from the definition of $\Lambda_{1,\delta'}^{\delta'}$. Therefore, the test Λ satisfies Definition 3 (a).

We now check Definition 3 (b). Without loss of generality, we set as $\bar{\delta} = 6\delta$. Pick any $\delta > 0$ and $P_1 \in \mathcal{M} \setminus \mathcal{P}_0$. We start by showing that

$$\Lambda_{0,2.1\delta}^{2.1\delta} \subset \Omega_0^{\delta}. \tag{21}$$

Suppose otherwise. Then there exists a sequence $\{\xi_m\}_{m\in\mathbb{N}}$ such that $\xi_m \in \Lambda_{0,2.1\delta}^{2.1\delta}$ and $\xi_m \in \Omega_1^{\delta}$ for all $m \in \mathbb{N}$. Since $\xi_m \in \Lambda_{0,2.1\delta}^{2.1\delta}$, there exists $\{\xi'_m\}_{m\in\mathbb{N}}$ such that $D_L(\xi_m, \xi'_m) < 4.2\delta$ and $\inf_{\mu\in\mathcal{P}_0} I(\xi'_m \| \mu) \leq \alpha$. The set $\{\xi \in \mathcal{M} : \inf_{\mu\in\mathcal{P}_0} I(\xi\|\mu) \leq \alpha\}$ is assumed to be compact, and therefore there exists a subsequence $\left\{\xi'_{m_k}\right\}_{k\in\mathbb{N}}$ such that $\xi'_{m_k} \to \xi' \in \{\xi \in \mathcal{M} : \inf_{\mu\in\mathcal{P}_0} I(\xi\|\mu) \leq \alpha\}$ as $k \to \infty$. Also, from $\xi_m \in \Omega_1^{\delta}$ and $D_L(\xi_m, \xi'_m) < 4.2\delta$ for all $m \in \mathbb{N}$, we have $\xi'_{m_k} \in \Omega_1^{5.2\delta}$ and thus $B\left(\xi'_{m_k}, \delta/2\right) \subset \Omega_1^{6\delta}$ for all $k \in \mathbb{N}$, which implies that the limit ξ' satisfies $B(\xi', \delta/4) \subset \Omega_1^{6\delta}$. Thus, Sanov's theorem implies

$$\begin{aligned}
&\sup_{P_0\in\mathcal{P}_0} \liminf_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \Omega_1^{6\delta} : P_0\right\} \geq \sup_{P_0\in\mathcal{P}_0} \liminf_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in B\left(\xi', \delta/4\right) : P_0\right\} \\
&\geq - \inf_{P_0\in\mathcal{P}_0} \inf_{P\in B(\xi',\delta/4)} I(P\|P_0) \geq -\alpha.
\end{aligned}$$

Since this contradicts with the requirement for Ω, we obtain (21). Now, observe that

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \Omega_0^{\delta} : P_1\right\} \geq \liminf_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \Lambda_{0,2.1\delta}^{2.1\delta} : P_1\right\} \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in int\left(\Lambda_{0,2.1\delta}^{2.1\delta}\right) : P_1\right\} \\
&\geq - \inf_{P\in int\left(\Lambda_{0,2.1\delta}^{2.1\delta}\right)} I(P\|P_1) \geq - \inf_{P\in\Lambda_{0,\delta}^{\delta}} I(P\|P_1) \\
&\geq \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\mu_n \in \Lambda_{0,\delta}^{\delta} : P_1\right\},
\end{aligned}$$

where the first inequality follows from (21), the second inequality follows from a set inclusion relationship, the third inequality follows from Sanov's theorem, the fourth inequality follows from (19), and the last inequality follows from Sanov's theorem. Therefore, the test Λ satisfies Definition 3 (b). ■

A.2 Proof of Theorem 3

The proof is an adaptation of that of Theorem 2.1 in Hall and La Scala (1990).

Proof for the property under H_0. Without loss of generality, we can assume that $\Omega = Var((g_i(\theta^*)', h_i(\beta^*)')')$ is positive definite. If $\text{rank}(\Omega) = r < l_g + l_h$, then as in Hall and La Scala (1990), one can express $l_g + l_h - r$ of the elements of $(g_i(\theta^*)', h_i(\beta^*)')$ as a linear combination of the remaining r elements with a positive definite variance matrix, and reformulate the problem using only those r elements. From Hall and Inoue (2003), we can see that $\sqrt{n}\left(\hat{\theta} - \theta^*\right) = O_p(1)$ and $\sqrt{n}\left(\hat{\beta} - \beta^*\right) = O_p(1)$ under Assumption 1. Thus, a standard Taylor expansion argument for EL (see, e.g., Hall and La Scala (1990)) yields a quadratic approximation:

$$-2n\ell\left(\eta_g,\eta_h\right)=n\left\|\Omega^{-1/2}\left(\begin{array}{c}\bar{g}\left(\hat{\theta}\right)-\eta_g\\ \bar{h}\left(\hat{\beta}\right)-\eta_h\end{array}\right)\right\|^2+o_p\left(1\right), \tag{22}$$

for any (η_g,η_h) and $C>0$ satisfying $\|Eg_i(\beta^*)-\eta_g\|\le Cn^{-1/2}$ and $\|Eh_i(\beta^*)-\eta_h\|\le Cn^{-1/2}$. On the other hand, under H_0 (i.e., $\|Eg_i(\beta^*)\|_{W_g}=\|Eh_i(\beta^*)\|_{W_h}$), a Taylor expansion of the constraint $\|\eta_g\|_{W_g}-\|\eta_h\|_{W_h}=0$ around $(\eta_g,\eta_h)=(Eg_i(\theta^*),Eh_i(\beta^*))$ yields

$$\left(\begin{array}{c}W_gEg_i\left(\theta^*\right)\\ -W_hEh_i\left(\beta^*\right)\end{array}\right)'\left(\begin{array}{c}Eg_i\left(\theta^*\right)-\eta_g\\ Eh_i\left(\beta^*\right)-\eta_h\end{array}\right)=o\left(n^{-1/2}\right), \tag{23}$$

for any (η_g,η_h) and $C>0$ satisfying $\|Eg_i(\beta^*)-\eta_g\|\le Cn^{-1/2}$ and $\|Eh_i(\beta^*)-\eta_h\|\le Cn^{-1/2}$. From (22) and (23) combined with the Taylor expansions of $\bar{g}\left(\hat{\theta}\right)$ and $\bar{h}\left(\hat{\beta}\right)$ around $\hat{\theta}=\theta^*$ and $\hat{\beta}=\beta^*$, respectively, the test statistic can be written as

$$T_n^A=\min_{y\in S_\nu^\perp}\left\|Z_n+\delta_n-y\right\|^2+o_p\left(1\right), \tag{24}$$

where

$$S_\nu^\perp=\left\{x\in\mathbb{R}^{l_g+l_h}:\nu'x=0\right\},\qquad \nu=\Omega^{1/2}\left(\begin{array}{c}W_gEg_i\left(\theta^*\right)\\ -W_hEh_i\left(\beta^*\right)\end{array}\right) \tag{25}$$

$$Z_n=\Omega^{-1/2}\sqrt{n}\left(\begin{array}{c}\bar{g}\left(\beta^*\right)-Eg_i\left(\theta^*\right)\\ \bar{h}\left(\beta^*\right)-Eh_i\left(\beta^*\right)\end{array}\right),\qquad \delta_n=\Omega^{-1/2}\sqrt{n}\left(\begin{array}{c}\frac{\partial\bar{g}(\theta^*)}{\partial\theta'}\left(\hat{\theta}-\theta^*\right)\\ \frac{\partial\bar{h}(\beta^*)}{\partial\beta'}\left(\hat{\beta}-\beta^*\right)\end{array}\right)$$

Now, let $Q_\nu=\nu\nu'/\|\nu\|^2$ be the projection matrix on $S_\nu=\{x:x=b\nu,b\in\mathbb{R}\}$. Then $(I-Q_v)Z_n\in S_\nu^\perp$. Also, from $v'\delta_n\xrightarrow{p}0$ (by the first-order conditions of $\hat{\theta}$ and $\hat{\beta}$), it holds $\text{plim}\delta_n\in S_\nu^\perp$. Therefore,

$$\begin{aligned}T_n^A&=\min_{y\in S_\nu^\perp}\left\|Q_\nu Z_n+\left(I-Q_\nu\right)Z_n+\delta_n-y\right\|^2+o_p(1)\\&=Z_n'Q_\nu Z_n+o_p\left(1\right)\\&\xrightarrow{d}\chi_1^2,\end{aligned}$$

where the result in the last line holds because $Z_n\to_d N\left(0,I_{l_g+l_h}\right)$ and $\text{rank}\left(Q_\nu\right)=1$.[13]

[13] When the both models are correctly specified, $\nu=0$, $S_\nu^\perp$ is $\mathbb{R}^{l_g+l_h}$, and $y=Z_n+\delta_n$ solves the minimization problem in (24). It follows then from (24) that $T_n^A=o_p(1)$.

Proof for the property under H_1. Pick any $c > 0$. Under H_1, we have

$$\begin{pmatrix} \bar{g}\left(\hat{\theta}\right) - \eta_g \\ \bar{h}\left(\hat{\beta}\right) - \eta_h \end{pmatrix} \xrightarrow{p} \begin{pmatrix} Eg_i\left(\theta^*\right) - \eta_g \\ Eh_i\left(\beta^*\right) - \eta_h \end{pmatrix} \neq 0,$$

for any (η_g, η_h) with $\|\eta_g\|_{W_g} = \|\eta_h\|_{W_h}$. Therefore, from (22), the conclusion follows. ■

A.3 Proof of Theorem 4

We only show the asymptotic null distribution. The property under H_1 is derived in the same manner as the proof of Theorem 3. The function $\ell_n(\eta_g, \eta_h)$ in (13) is now minimized under the constraint

$$0 = \eta_g' \hat{W}_g \eta_g - \eta_h' \hat{W}_h \eta_h = \eta_g' W_g \eta_g - \eta_h' W_h \eta_h + \eta_g' \left(\hat{W}_g - W_g\right)\eta_g - \eta_h'\left(\hat{W}_h - W_h\right)\eta_h.$$

By the same argument to derive (23), this constraint can be linearized as

$$\begin{pmatrix} W_g E g_i\left(\theta^*\right) \\ -W_h E h_i\left(\beta^*\right) \end{pmatrix}' \begin{pmatrix} Eg_i\left(\theta^*\right) - \eta_g \\ Eh_i\left(\beta^*\right) - \eta_h \end{pmatrix} + \frac{u_n}{2\sqrt{n}} = o_p\left(n^{-1/2}\right),$$

for any (η_g, η_h) and $C > 0$ satisfying $\|Eg_i(\beta^*) - \eta_g\| \le Cn^{-1/2}$ and $\|Eh_i(\beta^*) - \eta_h\| \le Cn^{-1/2}$, where $u_n = Eg_i(\theta^*)' \sqrt{n}\left(\hat{W}_g - W_g\right) Eg_i(\theta^*) - Eh_i(\beta^*)' \sqrt{n}\left(\hat{W}_h - W_h\right) Eh_i(\beta^*)$. Now, as in the proof of Theorem 3,

$$\begin{aligned} \hat{T}_n^A &= \min_{y:\nu' y + u_n/2 = 0} \|Z_n + \delta_n - y\|^2 + o_p(1) = \min_{y \in S_\nu^\perp} \left\| Z_n + \frac{u_n}{2\|\nu\|^2}\nu + \delta_n - y \right\|^2 + o_p(1) \\ &= \left(Z_n + \frac{u_n}{2\|\nu\|^2}\nu\right)' Q_\nu \left(Z_n + \frac{u_n}{2\|\nu\|^2}\nu\right) + o_p(1) = \frac{\left(\nu' Z_n + 0.5 u_n\right)^2}{\|\nu\|^2} + o_p(1). \end{aligned}$$

Assumption 2 and the central limit theorem imply

$$\begin{pmatrix} \nu' Z_n \\ u_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \pi' X_1 \\ \kappa' A X_2 \end{pmatrix}.$$

The result follows by the continuous mapping theorem. ■

A.4 Proof of Theorem 5

We prove only the first part of the theorem. By the same arguments as in the proof of Theorem 3, one can show that under $H_{0,g}$,

$$T_n^{A,g} \xrightarrow{d} \min_{\left\{y \in \mathbb{R}^{l_g + l_h} : \|Eg_i(\theta^*)\|_{W_g}^2 - \|Eh_i(\beta^*)\|_{W_h}^2 + \nu' y \le 0\right\}} \|Z - y\|^2$$

$$
\begin{aligned}
&\leq \min_{\left\{y\in\mathbb{R}^{l_g+l_h}:\nu' y\leq 0\right\}} \|Z-y\|^2 \\
&= \frac{1}{2}\chi_0^2+\frac{1}{2}\chi_1^2,
\end{aligned}
$$

where Z is a standard normal random vector, and ν is as defined in the proof of Theorem 3. The inequality in the second line holds because $\|Eg_i(\theta^*)\|_{W_g}^2 - \|Eh_i(\beta^*)\|_{W_h}^2 \leq 0$ under $H_{0,g}$, and the equality in the last line is by the results in Gourieroux, Holly, and Monfort (1982). ■

A.5 Sketch of proof for the null distribution of $T_{n,\rho}^A$

Under H_0 in (17), a similar quadratic approximation to derive (22) yields

$$
T_{n,\rho}^A = \frac{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\rho\left(\hat{\gamma}_g' g_i\left(\hat{\theta}_\rho\right)\right)-\rho\left(\hat{\gamma}_h' h_i\left(\hat{\beta}_\rho\right)\right)\right\}\right)^2}{\frac{1}{n}\sum_{i=1}^n\left\{\rho\left(\hat{\gamma}_g' g_i\left(\hat{\theta}_\rho\right)\right)-\rho\left(\hat{\gamma}_h' h_i\left(\hat{\beta}_\rho\right)\right)\right\}^2}+o_p(1). \tag{26}
$$

Define $\theta_\rho^* = \arg\min_{\theta\in\Theta}\max_{\gamma_g} E\rho\left(\gamma_g' g_i(\theta)\right)$ and $\gamma_g^* = \arg\max_{\gamma_g} E\rho\left(\gamma_g' g_i(\theta^*)\right)$, and let β_ρ^* and γ_h^* be defined similarly. Let $\rho_1(v) = d\rho(v)/dv$. We assume the $\sqrt{n}$-consistency of $\hat{\theta}_\rho$, $\hat{\beta}_\rho$, $\hat{\gamma}_g$, and $\hat{\gamma}_h$ to θ_ρ^*, β_ρ^*, γ_g^*, and γ_h^*, respectively (see, e.g., Domowitz and White, 1982, for regularity conditions). Under similar regularity conditions to Assumption 1, an expansion yields

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\rho\left(\hat{\gamma}_g' g_i\left(\hat{\theta}_\rho\right)\right)-\rho\left(\hat{\gamma}_h' h_i\left(\hat{\beta}_\rho\right)\right)\right\} \\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\rho\left(\gamma_g^{*\prime} g_i\left(\theta_\rho^*\right)\right)-\rho\left(\gamma_h^{*\prime} h_i\left(\beta_\rho^*\right)\right)\right\} \\
&+\frac{1}{n}\sum_{i=1}^n \rho_1\left(\gamma_g^{*\prime} g_i\left(\theta_\rho^*\right)\right) g_i\left(\theta_\rho^*\right)'\sqrt{n}\left(\hat{\gamma}_g-\gamma_g^*\right) \\
&+\frac{1}{n}\sum_{i=1}^n \rho_1\left(\gamma_g^{*\prime} g_i\left(\theta_\rho^*\right)\right)\gamma_g^{*\prime}\frac{\partial g_i\left(\theta_\rho^*\right)}{\partial\theta'}\sqrt{n}\left(\hat{\theta}_\rho-\theta_\rho^*\right) \\
&-\frac{1}{n}\sum_{i=1}^n \rho_1\left(\gamma_h^{*\prime} h_i\left(\beta_\rho^*\right)\right) h_i\left(\beta_\rho^*\right)'\sqrt{n}\left(\hat{\gamma}_h-\gamma_h^*\right) \\
&-\frac{1}{n}\sum_{i=1}^n \rho_1\left(\gamma_h^{*\prime} h_i\left(\beta_\rho^*\right)\right)\gamma_h^{*\prime}\frac{\partial h_i\left(\beta_\rho^*\right)}{\partial\beta'}\sqrt{n}\left(\hat{\beta}_\rho-\beta_\rho^*\right)+o_p(1) \\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\rho\left(\gamma_g^{*\prime} g_i\left(\theta_\rho^*\right)\right)-\rho\left(\gamma_h^{*\prime} h_i\left(\beta_\rho^*\right)\right)\right\}+o_p(1),
\end{aligned}
$$

where the second equality follows from the first-order conditions for θ_ρ^*, β_ρ^*, γ_g^*, and γ_h^*:

$$E\left[\rho_1\left(\gamma_g^{*\prime} g_i\left(\theta_\rho^*\right)\right) g_i\left(\theta_\rho^*\right)\right] = 0, \quad E\left[\rho_1\left(\gamma_g^{*\prime} g_i\left(\theta_\rho^*\right)\right) \gamma_g^{*\prime} \frac{\partial g_i\left(\theta_\rho^*\right)}{\partial \theta'}\right] = 0, \tag{27}$$

(similar results hold for β_ρ^* and γ_h^*). Therefore, under H_0 (i.e., $E[\rho(\gamma_g^{*\prime} g_i(\theta_\rho^*)) - \rho(\gamma_h^{*\prime} h_i(\beta_\rho^*))] = 0$), the central limit theorem implies $T_{n,\rho}^A \xrightarrow{d} \chi_1^2$. ■

References

Bates, C., and H. White (1985): "A Unified Theory of Consistent Estimation for Parametric Models," *Econometric Theory*, 1(2), 151–178.

Chaganty, N. R., and R. L. Karandikar (1996): "Some Properties of the Kullback-Leibler Number," *Sankhyā Series A*, 58, 69–80.

Christoffersen, P., J. Hahn, and A. Inoue (2001): "Testing and Comparing Value–at–Risk Measures," *Journal of Empirical Finance*, 8(3), 325–342.

Corradi, V., and N. R. Swanson (2007): "Evaluation of Dynamic Stochastic General Equilibrium Models Based on Distributional Comparison of Simulated and Historic Data," *Journal of Econometrics*, 136(2), 699–723.

Davidson, R., and J. G. MacKinnon (1981): "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49(3), 781–793.

Dembo, A., and O. Zeitouni (1998): *Large Deviations Techniques and Applications*. Springer, New York.

Deuschel, J. D., and D. W. Stroock (1989): *Large Deviations*. Academic Press, New York.

Domowitz, I., and H. White (1982): "Misspecified Models with Dependent Observations," *Journal of Econometrics*, 20(1), 35–58.

Dridi, R., A. Guay, and E. Renault (2007): "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136(2), 397–430.

Gallant, A. R., and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, Oxford.

Ghysels, E., and A. R. Hall (1990): "Testing Nonnested Euler Conditions with Quadrature-Based Methods of Approximation," *Journal of Econometrics*, 46(3), 273 – 308.

Gourieroux, C., A. Holly, and A. Monfort (1982): "Likelihood Ratio Test, Wlad Test, and Kuhn-Tucker Test in Linear Models With Inequality Constraints on the Regression Parameters," *Econometrica*, 50(1), 63–80.

Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

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

Hall, A. R., and A. Inoue (2003): "The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114(2), 361–394.

Hall, A. R., and D. Pelletier (2007): "Non-Nested Testing in Models Estimated via Generalized Method of Moments," Working Paper 011, North Carolina State University.

——— (2011): "Non-Nested Testing in Models Estimated via Generalized Method of Moments," *Econometric Theory*, forthcoming.

Hall, P., and B. La Scala (1990): "Methodology and Algorithms of Empirical Likelihood," *International Statistical Review*, 58(2), 109–127.

Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

Hansen, L. P., and R. Jagannathan (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 50(2), 557–590.

Hendry, D. F. (1979): "The Behaviour of Inconsistent Instrumental Variables Estimators in Dynamic Systems with Autocorrelated Errors," *Journal of Econometrics*, 9(3), 295–314.

Hoeffding, W. (1965): "Asymptotically Optimal Tests for Multinomial Distributions," *Annals of Mathematical Statistics*, 36(2), 369–408.

Kan, R., and C. Robotti (2009): "Model Comparison Using the Hansen-Jagannathan Distance," *Review of Financial Studies*, 22(9), 3449–3490.

Kan, R., C. Robotti, and J. A. Shanken (2009): "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology," NBER Working Paper 15047.

Kitamura, Y. (1997): "Empirical Likelihood Methods with Weakly Dependent Processes," *Annals of Statistics*, 25(5), 2084–2102.

——— (2000): "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood," Working Paper, University of Pennsylvania.

——— (2001): "Asymptotic Optimality of Empirical Likelihood For Testing Moment Restrictions," *Econometrica*, 69(6), 1661–1672.

——— (2003): "A Likelihood-Based Approach to the Analysis of a Class of Nested and Non-Nested Models," Working Paper, University of Pennsylvania.

——— (2007): "Empirical Likelihood Methods in Econometrics: Theory and Practice," in *Advances in Economics and Econometrics*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. 3, chap. 7. Cambridge University Press.

Kitamura, Y., A. Santos, and A. M. Shaikh (2009): "On the Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," Cowles Foundation Discussion Paper No. 1722, Yale University.

Kitamura, Y., and M. Stutzer (2002): "Connections Between Entropic and Linear Projections in Asset Pricing Estimation," *Journal of Econometrics*, 107(1-2), 159–174.

Leininger, W. (1984): "A Generalization of the Maximum Theorem," *Economics Letters*, 15, 309–313.

Maasoumi, E., and P. C. B. Phillips (1982): "On the Behavior of Inconsistent Instrumental Variables Estimators," *Journal of Econometrics*, 19, 183–201.

MacKinnon, J. G. (1983): "Model Specification Tests Against Non-Nested Alternatives," *Econometric Reviews*, 2(1), 85–110.

Newey, W., and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72(1), 219–255.

Otsu, T., and Y.-J. Whang (2008): "Testing for Non-nested Conditional Moment Restrictions via Conditional Empirical Likelihood," *Econometric Theory*, forthcoming.

Owen, A. B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75(2), 237–249.

Prescott, E. C. (1991): "Real Business Cycle Theory: What Have We Learned?," *Revista de Análisis Económico*, 6(2), 3–19.

Qin, J., and J. Lawless (1994): "Empirical likelihood and general estimating equations," *Annals of Statistics*, 22(1), 300–325.

Ramalho, J. J. S., and R. J. Smith (2002): "Generalized Empirical Likelihood Non-Nested Tests," *Journal of Econometrics*, 107(1-2), 99–125.

Rivers, D., and Q. Vuong (2002): "Model Selection Tests For Nonlinear Dynamic Models," *Econometrics Journal*, 5(1), 1–39.

SERFLING, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. John Wiley, New York.

SHI, X. (2009): "Model Selection Tests for Nonnested Moment Inequality Models," Working Paper.

SINGLETON, K. J. (1985): "Testing Specifications of Economic Agents' Intertemporal Optimum Problems in the Presence of Alternative Models," *Journal of Econometrics*, 30(1-2), 391–413.

SMITH, R. J. (1992): "Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments," *Econometrica*, 60(4), 973–980.

——— (1997): "Alternative Semi-parametric Likelihood Approaches to Generalised Method of Moments Estimation," *Economic Journal*, 107(441), 503–519.

VUONG, Q. H. (1989): "Likelihood Ratio Tests For Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333.

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

ZEITOUNI, O., AND M. GUTMAN (1991): "On Universal Hypotheses Testing via Large Deviations," *IEEE Transactions on Information Theory*, 37(2), 285–290.