Model Equivalence Tests
for Overidentifying Restrictions

Pascal Lavergne, Toulouse School of Economics

March 2015

Abstract

I propose a new theoretical framework to assess the approximate validity of overidentifying moment restrictions. Their approximate validity is evaluated by the divergence between the true probability measure and the closest measure that imposes the moment restrictions of interest. The divergence can be chosen as any of the Cressie-Read family. The considered alternative hypothesis states that the divergence is smaller than some user-chosen tolerance. Model equivalence tests are constructed for this hypothesis based on the minimum empirical divergence. These tests attain the local semiparametric power envelope of invariant tests. Three empirical applications illustrate their practical usefulness for providing evidence on the potential extent of misspecification.

Keywords: Hypothesis testing, Semiparametric models.
JEL Codes: C12, C14, C52.

I acknowledge financial support from SSHRC. This paper was presented at Econometric Study Group Conference, IAAE 2014, and at seminars in Toulouse, Mannheim, Marseille, and CREST-Paris. I thanks many colleagues and participants, and in particular Vadim Marmer and Laurent Davezies, for useful comments. Address correspondence: Pascal Lavergne, Toulouse School of Economics, 21 Allées de Brienne, 31000 Toulouse FRANCE. Email: pascal.lavergne@ut-capitole.fr
Although this may seem a paradox, all exact science is dominated by the idea of approximation.  

B. Russell (1931)

A realistic attitude . . . is that an economic model or a probability model is in fact only a more or less crude approximation to whatever might be the “true” relationships among the observed data . . . Consequently, it is necessary to view economic and/or probability models as misspecified to some greater or lesser extent.  

H. White (1996)

1 Introduction

Economic structural parameters are often identified and estimated through moment restrictions. For estimation, the Generalized Method of Moments (GMM) studied by Hansen (1982) is popular among practitioners. More recently, alternative methods have been studied, in particular Empirical Likelihood (EL), see Imbens (1993) and Qin and Lawless (1994), Exponential Tilting (ET), see Imbens (1993) and Kitamura and Stutzer (1997), and the Continuously Updated Estimator (CUE), see Hansen, Heaton, and Yaron (1996) and Antoine, Bonnal, and Renault (2007). As explained by Kitamura (2007), all these estimators rely on minimizing a divergence (or contrast) between the empirical distribution of the observations and one that imposes the moment restrictions. Smith (1997), Imbens, Spady, and Johnson (1998), and Newey and Smith (2004) consider a general class of Cressie-Read divergences that yield Generalized Empirical Likelihood (GEL) estimators.

When parameters are overidentified, it is usually of interest to assess the validity of overidentifying restrictions. For each of the above estimation methods, the objective function can serve as a basis for an overidentification test. Such a test may conclude against the moment restrictions, but can never provide positive evidence in favor of these restrictions. Hence the researcher can never conclude that they hold, even in an approximate sense. This is because a standard overidentification test considers as the null hypothesis the restrictions validity and aims to control the probability of falsely rejecting correct restrictions. However, what seems more crucial in practice is to control the probability of not rejecting a grossly misspecified model. This error is indeed the one that can have more adverse effects in applied economic analysis.

The goal of this work is to develop “classical” tests for assessing the approximate validity of overidentifying restrictions. The interest of approximate hypotheses has been
long recognized in statistics, see e.g. Hodges and Lehmann (1954). Leamer (1998) argues that “genuinely interesting hypotheses are neighborhoods, not points. No parameter is exactly equal to zero; many may be so close that we can act as if they were zero,” see also Good (1981) in statistics or McCloskey (1985) in economics among many others. Here the approximate validity of the moment condition is considered as the alternative hypothesis to reflect where the burden of proof is. This is known in biostatistics as equivalence testing, see Lehmann and Romano (2005) and the monograph of Wellek (2003). Applications of this approach are found for instance in Romano (2005) and Lavergne (2014) for restrictions on parameters, and in Rosenblatt (1962) and Dette and Munk (1998) for specification testing. With reference to equivalence testing in biostatistics, our tests are labeled model equivalence tests for overidentifying restrictions.

Practically, the approach recognizes that any model is misspecified to some extent, and aims at confirming that misspecification is relatively small. To develop such tests, a central issue is how to measure the extent of misspecification, or equivalently to evaluate the approximate validity of the moment conditions. Here we built on recent work on GEL estimation and we chose as a measure a theoretical Cressie-Read divergence. This choice is mainly motivated by invariance considerations. Indeed, any measure of validity (or lack of) should not vary if moment restrictions are reformulated in a different but equivalent way. Such a measure should also be invariant to any (potentially nonlinear) reparameterization. A Cressie-Read divergence fulfills these requirements. Moreover, it has a natural information-theoretic interpretation. As will be shown, any Cressie-Read divergence yields approximately the same theoretical measure of validity if a model is close to be well specified. Given a divergence between the true probability distribution and the “closest distribution” that imposes the moment conditions, we consider as the alternative hypothesis to be assessed that this divergence is smaller than some user-chosen tolerance. We label it the model equivalence hypothesis. We will discuss the practical choice of the tolerance, which is crucial for determining what is the acceptable amount of misspecification in a particular application.

Our generic model equivalence test is based on the corresponding empirical divergence. The alternative hypothesis is accepted for small values of the empirical divergence, and the critical value is not derived under the assumption that the moment restrictions are valid. Our model equivalence tests generalize the ones proposed in Lavergne (2014), who deals with restrictions on parameters in parametric models, in three directions. We consider
semiparametric models, general divergences, and overidentifying restrictions. The new tests have interesting properties, in particular they are by construction invariant to any transformation of the moment restrictions (by contrast a test based on a two-step GMM statistic would not necessarily be invariant). In addition, they attain the semiparametric power envelope of invariant tests.

We also show that the model equivalence hypothesis can be reformulated in terms of parameters. Imposing incorrect overidentifying restrictions yield an asymptotic bias in parameter estimation. The model equivalence hypothesis states that the induced bias is relatively small. The considered parameters include model parameters \( \theta \) as well as possible deviations of the moment restrictions from zero. Similarly, one can write the model equivalence hypothesis in terms of moments that provide overidentifying restrictions. It would be possible to test the latter hypothesis directly, but such a procedure would not share the desirable invariance properties of our model equivalence tests. However, we will show how these reformulations can be helpful in deciding for an appropriate tolerance. Finally, it must be stressed that any test that focuses on a particular feature or “prediction” cannot assess model validity. For instance, a divergence based on a subset of parameters only could be zero while the model is grossly misspecified. This is related to the well-known inconsistency of the Hausman test when focused on some specific parameter’s components, see e.g. Holly (1982). Hence such a focused test cannot provide a suitable basis for assessing the approximate validity of overidentifying restrictions, which is our goal here.

Our new tests allow to conclude that the model may be misspecified to an extent that is acceptable by the practitioner, as measured by the chosen tolerance. Since the seminal work by White (1982, 1996), it has been widely recognized that misspecification is the rule rather than the exception, and a growing literature aims at accounting for potential misspecification in inference. Recent work focuses on consequences on inference of local misspecifications of moment restrictions, e.g. instruments that locally violate exogeneity, and on the development of adapted inference methods, see Berkowitz, Caner, and Fang (2008), Berkowitz, Caner, and Fang (2012), Bugni, Canay, and Guggenberger (2012), Conley, Hansen, and Rossi (2012), Kraay (2012), Guggenberger (2012), Nevo and Rosen (2012), and Caner (2014), among others. Our model equivalence tests aim at assessing whether the moment restrictions are close to be valid and thus provide a complementary tool.
One may wonder whether and why a new approach is needed. A model equivalence test relies on a precise characterization of the approximate hypothesis under test. Compared to a usual overidentification test, it thus provides complementary information on the amount of potential misspecification of the model, as will be shown in our empirical applications. Considering as our alternative hypothesis one that states that the restrictions are “almost” valid roughly involve “flipping” the null and alternative hypotheses of a standard overidentification test. This is in line with the statistical principle that we should consider as the alternative hypothesis what we would like to show, so as to control the probability of falsely “confirming” the hypothesis of interest. Could confidence intervals or regions be used instead? As these are defined as sets of parameters values that cannot be rejected by a standard significance test, they do not provide a suitable answer. As will be seen in our illustrations of Section 4, the outcome of a model equivalence test cannot be reinterpreted as defining a confidence region of a special kind. Another possible approach would be to rely on power evaluation of overidentification tests. In particular, Andrews (1989) proposes approximations of the asymptotic inverse power function of Wald tests for restrictions on parameters as an aid to interpret non significant outcomes. While such an approach might be generalized to overidentification tests, this has not been investigated up to date.\footnote{Wald tests are not invariant to nonlinear transformations of restrictions under scrutiny, see e.g. Gregory and Veall (1985). Moreover, evaluating the asymptotic power of a significance test of given level does not directly provide evidence in favor of the approximate validity of the restrictions under consideration. Other issues surround post-experiment power calculations, as summarized by Hoenig and Heisey (2001).} To sum up, model equivalence tests for overidentifying restrictions deliver a new type of inference that is complementary to existing methods.

The paper is organized as follows. In Section 2, we develop a testing framework and a model equivalence test based on the chi-square divergence. This allows to discuss the main features of the approach and to provide alternative interpretations of our framework. In Section 3, we set up a more general framework considering a class of Cressie-Read divergences, that includes as special cases the ones used in EL or ET. We discuss the influence of the divergence and the choice of the tolerance for the test. In Section 4, we illustrate the usefulness of the new tests on three selected empirical examples. In Section 5, we study the theoretical properties of tests of model equivalence for overidentifying restrictions. Specifically, we derive the semiparametric envelope of tests that are invariant to transformations of the moment restrictions and we show that our tests reach this envelope.
Section 6 concludes. Section 7 contains the proofs of our results.

2 Test Based on the Chi-Square Divergence

2.1 Testing Framework

For a random variable $X \in \mathbb{R}^q$ with probability distribution $P$, we want to assess some implicit restrictions of the form

$$\exists \theta_0 \in \Theta \text{ such that } \mathbb{E}g(X, \theta_0) = 0,$$  \hspace{1cm} (2.1)

where $g(\cdot, \theta)$ is a $m$-vector function indexed by a finite-dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^p$, $p < m$. To do so, we can evaluate the divergence between $P$ and a measure that imposes these restrictions. Let us define the chi-square divergence (or contrast) between two measures $Q$ and $P$ as

$$D_2(Q, P) = \mathbb{E} \frac{1}{2} \left( \frac{dQ}{dP} - 1 \right)^2 = \frac{1}{2} \int \left( \frac{dQ}{dP} - 1 \right)^2 dP,$$

where $\frac{dQ}{dP}$ denotes the Radon-Nikodym derivative. Hence $D_2(Q, P) \geq 0$ with equality if and only if $Q = P$ $P$–almost surely. Twice the chi-square divergence measures the expected squared proportional difference between distributions and is thus an expected squared percentage. For a particular value of $\theta \in \Theta$, let us define $M_\theta = \{Q \text{ finite measure} : Q << P, \int dQ = 1, \int g(X, \theta) dQ = 0\}$ and $D_2(M_\theta, P) = \inf_{Q \in M_\theta} D_2(Q, P)$. A minimizer $Q_\theta$ of $D_2(Q, P)$ over $M_\theta$, if it exists, is labeled a projection of $P$ on $M_\theta$. Now let $\mathcal{M} = \cup_{\theta \in \Theta} M_\theta$. A minimizer $Q_\mathcal{M}$ of $D_2(Q, P)$ over $\mathcal{M}$ is a projection of $P$ on $\mathcal{M}$. The quantity

$$D_2(\mathcal{M}, P) = \inf_{\Theta} D_2(\mathcal{M}_\theta, P)$$  \hspace{1cm} (2.2)

provides a global measure of the approximate validity of the restrictions (2.1). By definition, this measure is invariant to any reparameterization and any transformation of the restrictions. In particular, for any $q \times q$ matrix $A(\theta)$ which is nonsingular for any $\theta$ with probability one, the moment restrictions (2.1) remains unaltered if $g(\cdot, \theta)$ is replaced by
A(\theta)g(\cdot, \theta), and so does \( D_2(\mathcal{M}, P) \). Moreover, a duality approach, as discussed e.g. by Kitamura (2007) and briefly outlined in Section 7.1, shows that

\[
D_2(\mathcal{M}, P) = \frac{1}{2} \min_{\Theta} \mathbb{E} \left( g'(X, \Theta) \right) \left[ \text{Var} \ g(X, \Theta) \right]^{-1} \mathbb{E} \left( g(X, \Theta) \right),
\]

(2.3)

see e.g. Antoine et al. (2007). This is the theoretical objective function used in CUE-GMM method. Hence the divergence evaluates the distance to zero of the moment restrictions, where the implicit norm depends on their variance, and has a pretty intuitive content.

To assess the approximate validity of our moment restrictions, we consider the alternative hypothesis that \( D_2(\mathcal{M}, P) \) is smaller than some tolerance chosen by the practitioner. That is, there is a measure imposing the moment restrictions which is close enough to the true probability measure. We write our alternative hypothesis as

\[
H_{1n} : 2D_2(\mathcal{M}, P) < \frac{\delta^2}{n}.
\]

This hypothesis is labeled the model equivalence hypothesis. It allows for some local misspecification of the moment restrictions, as apparent from (2.3). The null hypothesis is the complement of the alternative, that is

\[
H_{0n} : 2D_2(\mathcal{M}, P) \geq \frac{\delta^2}{n}.
\]

Considering a shrinking alternative hypothesis is a purely theoretical but useful device, as considered for model equivalence testing by Romano (2005) and Lavergne (2014). The vanishing tolerance \( \delta^2/n \) acknowledges that the tolerance is small in a substantive sense. In practice, as in subsequent illustrations, a small but fixed tolerance \( \Delta^2 \) is typically chosen, where \( \Delta \) can be seen as a percentage, so one can set \( \delta^2 = n\Delta^2 \) to run the test. But because the fixed tolerance is small, the asymptotics under a drifting tolerance will approximate the finite sample distribution of the test statistic better than the asymptotics under a fixed tolerance.

2.2 Testing Procedure

With at hand a random sample \( \{X_i, i = 1, \ldots, n\} \) from \( X \), the empirical divergence of interest is

\[
D_2(Q, P_n) = \mathbb{E}_n \left[ \frac{1}{2n} \left( \frac{dQ}{dP_n} - 1 \right)^2 \right] = \frac{1}{2n} \sum_{i=1}^{n} (Q(X_i) - 1)^2,
\]
where \( \mathbb{E}_n \) denotes expectation with respect to the empirical distribution \( P_n \). Let \( \mathcal{M}_{n, \theta} = \{ Q \text{ finite measure} : Q << P_n, \int dQ = 1, \int g(X, \theta) dQ = 0 \} \), \( \mathcal{M}_n = \cup_{\theta \in \Theta} \mathcal{M}_{n, \theta} \), and

\[
D_2(\mathcal{M}_n, P_n) = \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_{n, \theta}} D_2(Q, P_n).
\]

This quantity is the empirical equivalent of the theoretical divergence and thus provides a natural estimator of the latter. In addition, duality extends to the empirical chi-square divergence, so that

\[
D_2(\mathcal{M}_n, P_n) = \frac{1}{2} \min_{\theta} \mathbb{E}_n \left( g'(X, \theta) \right) \left[ \text{Var}_n g(X, \theta) \right]^{-1} \mathbb{E}_n \left( g(X, \theta) \right),
\]

where \( \text{Var}_n \) denotes the empirical variance. When the restrictions hold, we obtain as a by-product the CUE-GMM estimator of \( \theta_0 \) that fulfills (2.1). By contrast to standard two-step GMM, estimation is one-step and does not require a preliminary estimator. The empirical divergence is also invariant to any reparameterization and any transformation of the restrictions, while this may not be the case for the two-step GMM optimal objective function, see e.g. Hall and Inoue (2003).

The empirical divergence provides a natural basis for testing \( H_0n \) against \( H_1n \). When the theoretical divergence \( 2D_2(\mathcal{M}, P) = \frac{\delta^2}{n} \), \( 2n D_2(\mathcal{M}_n, P_n) \) converges in distribution to a \( \chi^2_r(\delta^2) \), the non-central chi-square with \( r = m - p \) degrees of freedom with noncentrality parameter \( \delta^2 \). The model equivalence test is then defined as

\[
\pi_n = \mathbb{I} \left[ 2n D_2(\mathcal{M}_n, P_n) < c_{\alpha, r, \delta^2} \right],
\]

where \( c_{\alpha, r, \delta^2} \) is the \( \alpha \)-quantile of a \( \chi^2_r(\delta^2) \). The test thus concludes that overidentifying restrictions are approximately valid if the test statistic \( 2n D_2(\mathcal{M}_n, P_n) \) is relatively small. This stands in contrast to the standard overidentification test, that rejects the exact validity of overidentifying restrictions for large values of the test statistic, and for which the critical value is the \( 1 - \alpha \) quantile of a central chi-square distribution. This is because our model equivalence test does not assume that overidentifying restrictions hold under the null hypothesis, as the test aims at confirming that these restrictions approximately hold. While critical values are non-standard, they can be readily obtained from most statistical softwares, and selected values are reported in Lavergne (2014).

The main properties of the test are easily derived. First, it is invariant to reparameterization and to transformation of the moment restrictions. Second, when \( 2D_2(\mathcal{M}, P) \) is
large, which corresponds to grossly misspecified restrictions, the test will fail to reject \( H_{0n} \) in favor of model equivalence. This can be deduced from the convergence of \( D_2(M_n, P_n) \) to the theoretical divergence \( D_2(M, P) \), see Broniatowski and Keziou (2012, Theorem 5.6). In Section 5, we will study further the asymptotic optimality properties of the test.

If we do not wish to choose a tolerance at the outset, we may let it vary for a given level of the test. Formally, let

\[
\delta_{\text{inf}}^2(\alpha) = \inf \{ \delta^2 > 0 : 2n \ D_2(M_n, P_n) < c_{\alpha, r, \delta^2} \}.
\]  

(2.5)

This provides a useful benchmark against which practitioners may be able to decide a posteriori whether \( \Delta_{\text{inf}}^2(\alpha) = \delta_{\text{inf}}^2(\alpha)/n \) is too large to consider that the restrictions are approximately valid, or is sufficiently small to conclude that they are. We will illustrate in Section 4 how this can help reaching a conclusion on the model approximate validity.

### 2.3 Alternative Formulations of Model Equivalence

We now show how to formulate and interpret the model equivalence hypothesis in terms of parameters. Such alternative formulations are appealing from an empirical viewpoint. Specifically, they can be obtained using a useful intuition from Newey and McFadden (1994). For a \( p \times m \) matrix \( L \) with full rank \( p \), consider the partition of \( g(\cdot, \theta) \) into a \( p \)-vector \( g_1(\cdot, \theta) = Lg(\cdot, \theta) \) and the remaining \( (m - p) \) vector \( g_2(\cdot, \theta) = Mg(\cdot, \theta) \), where \( [L, M] \) is full rank. Let \( \lambda = (\theta', \nu')' \in \Lambda = \Theta \times \mathbb{R}^{m-p} \), and define

\[
h(X, \lambda) = \begin{bmatrix} g_1(X, \theta) \\ g_2(X, \theta) - \nu \end{bmatrix}.
\]

For any \( \lambda \in \Lambda \), let \( M_\lambda = \{ Q \text{ finite measure} : Q << P, \int dQ = 1, \int h(X, \lambda) dQ = 0 \} \), and \( M_\Lambda = \cup_{\lambda \in \Lambda} M_\lambda \). On the one hand, under standard assumptions, there exists a unique \( \theta^* \) such that \( \mathbb{E} g_1(X, \theta^*) = 0 \), so that \( \mathbb{E} h(X, \lambda^*) = 0 \) for \( \lambda^* = (\theta^*, \nu^*)' = \mathbb{E} g'_2(X, \theta^*) \)' and \( D(M, P) = 0 \). On the other hand, if we restrict \( \nu \) to be zero, the problem boils down to the one of interest and

\[
\inf_{\Theta \times \{0\}} D_2(M_\lambda, P) = \inf_{\Theta} D_2(M_\theta, P).
\]
Let us label $\lambda^*_R = \left(\hat{\theta}_0, \theta^*\right)$ the solution to the above problem and define the following divergences

$$D_H(M, \tilde{P}^{(n)}) = \frac{1}{2} \frac{1}{2} (\lambda^* - \lambda^*_R)^\prime J (\lambda^* - \lambda^*_R)$$

and

$$D_W(M, P) = \frac{1}{2} \frac{1}{2} \mathbb{E} \frac{1}{2} g_2(X, \theta^*) \Sigma^{-1} \mathbb{E} g_2(X, \theta^*),$$

where $J^{-1}$ and $\Sigma$ are the semiparametric efficiency bounds on the $\sqrt{n}$-variances for estimating $\lambda^*$ and $\nu^*$, respectively. We will show that these divergences are both asymptotically equivalent to $D_2(M, P)$ in the following sense.

**Definition 1** Two divergence measures $d_i$, $i = 1, 2$, are locally equivalent under a drifting sequence of probability distributions $\tilde{P}^{(n)}$, $n \geq 1$, if whenever $d_i(M, \tilde{P}^{(n)}) = o(1)$ or $d_j(M, \tilde{P}^{(n)}) = o(1)$, we have $d_i(M, \tilde{P}^{(n)}) = d_j(M, \tilde{P}^{(n)})(1 + o(1))$.

Let us introduce the following assumptions.

**Assumption A** (i) $\Theta$ is compact; (ii) $\text{Var} \ g(X, \theta)$ is positive definite for any $\theta \in \Theta$; (iii) For any $p \times m$ matrix $L$ with full rank $p$, there exists a unique solution $\theta^*$ to the equations $L \mathbb{E} g(X, \theta) = 0$; (iv) $\tilde{\theta}_0 = \arg \inf_\Theta D_1(M, \tilde{P}^{(n)})$ exists and is unique, and $\nabla_\theta \mathbb{E} g(X, \tilde{\theta}_0)$ is full rank.

**Assumption B** Each component of the function $g(\cdot, \theta)$ is twice continuously differentiable in $\theta$ over $\Theta$.

**Lemma 2.1** Under any drifting sequence of probability distributions $\tilde{P}^{(n)}$ such that Assumptions A and B hold, $D_2(M, \tilde{P}^{(n)})$, $D_H(M, \tilde{P}^{(n)})$, and $D_W(M, \tilde{P}^{(n)})$ are asymptotically equivalent.

Our result entails that the alternative hypothesis $H_{1n}$ is asymptotically equivalent to

$$(\lambda^* - \lambda^*_R)^\prime J (\lambda^* - \lambda^*_R) < \frac{\delta^2}{n}. \quad (2.6)$$

This formulation uses a divergence that compares the parameters values $\lambda^*$ and $\lambda^*_R$. Evaluating the closeness to zero of overidentifying restrictions through the chi-square divergence is thus equivalent to evaluate the consequences on parameters of these restrictions. One possible interpretation of this formulation is as follows. The difference $\lambda^* - \lambda^*_R$ is the
asymptotic bias that is potentially introduced by imposing overidentification restrictions. Hence the model equivalence hypothesis states that this bias is of order \( n^{-1/2} \) and small enough (compared to statistical variability) as measured by the tolerance. This bias is scaled by \( J^{1/2} \), so that biases are comparable across elements of \( \lambda \). However, it is important to consider the whole extended parameter vector \( \lambda \). A divergence that would be based on the parameters in \( \theta \) only, or on a subset of them, can fail to provide an accurate measure of the validity of the moment restrictions, and may be zero while the model is grossly misspecified. This is basically similar to the well-known inconsistency of the Hausman test when focused on some specific parameter’s components, see e.g. Holly (1982).

The model equivalence hypothesis is also asymptotically equivalent to

\[
\mathbb{E} g_2(X, \theta^*) \Sigma^{-1} \mathbb{E} g_2(X, \theta^*) < \frac{\delta^2}{n}.
\]  

(2.7)

This second alternative formulation uses a divergence that focuses on the closeness to zero of \( m - p \) overidentification restrictions. For instance, if there is one degree of overidentification only, i.e. \( m - p = 1 \), then the above expression becomes

\[
|\mathbb{E} g_2(X, \theta^*)| < \frac{\delta \sigma}{\sqrt{n}},
\]

where \( \sigma^2 \) is the semiparametric bound on the \( \sqrt{n} \)-variance for estimating \( \mathbb{E} g_2(X, \theta^*) \). With a consistent estimator of \( \sigma \), one can then evaluate the content of the model equivalence hypothesis in terms of closeness to zero of the overidentification restriction. Though this last formulation is simple and intuitive, it must be kept in mind that direct tests of this hypothesis would generally not be invariant. We could nevertheless use this asymptotically equivalent formulation for interpretative purposes, as we will illustrate in Section 4.

3 Cressie-Read Divergence-Based Testing

We here detail the more general procedures based on general Cressie-Read divergences and we discuss their relationship with the test described in the previous section.

3.1 General Model Equivalence Test

As done by Smith (1997), Imbens et al. (1998), Newey and Smith (2004), and Kitamura (2007), we focus here on the class of divergences based on the Cressie and Read (1984)
family of functions
\[ \varphi_\gamma(x) = \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma (\gamma - 1)}, \quad \gamma \in \mathbb{R} \setminus \{0, 1\}, \]
\[ \varphi_1(x) = x \log x - x + 1, \]
\[ \varphi_0(x) = -\log x + x - 1. \]
Moreover, if \( \varphi_\gamma(\cdot) \) is not defined on \(( -\infty, 0)\), as for \( \gamma = 0 \), or when it is not convex on \(( -\infty, 0)\) as \( \varphi_3(x) \), we set it to \(+\infty\) on \(( -\infty, 0)\). Hence, all considered functions are strictly convex, positive, and twice differentiable on their domain. The way we wrote the Cressie-Read family of functions slightly differs from most of the econometric literature, but yields the normalization \( \varphi_\gamma(1) = 0 \), \( \varphi'_\gamma(1) = 0 \), and \( \varphi''_\gamma(1) = 1 \), so that all functions behave similarly around 1 up to second-order. The cases \( \gamma = 1 \) and 0 correspond to Kullback-Leibler-type divergences, \( \gamma = 1/2 \) yields the Hellinger divergence, and \( \gamma = 2 \) the chi-square divergence considered above.

For each \( \gamma \), the Cressie-Read divergence between two measures \( Q \) and \( P \) is defined as
\[ D_\gamma(Q,P) = \mathbb{E} \varphi_\gamma \left( \frac{dQ}{dP} \right) = \int \varphi_\gamma \left( \frac{dQ}{dP} \right) dP. \]
The quantity \( D_\gamma(\mathcal{M},P) = \inf_\Theta D_\gamma(\mathcal{M}_\theta,P) \) thus provides an alternative global measure of the validity of the moments restrictions (2.1). The model equivalence hypothesis based on \( D_\gamma(\cdot,\cdot) \) then writes
\[ H_{1n} : 2 D_\gamma(\mathcal{M},P) < \frac{\delta^2}{n}, \]
and the null hypothesis is
\[ H_{0n} : 2 D_\gamma(\mathcal{M},P) \geq \frac{\delta^2}{n}. \]
The corresponding empirical divergence is
\[ D_\gamma(\mathcal{M}_n,P_n) = \inf_\Theta \inf_{Q \in \mathcal{M}_n,\theta} D_\gamma(Q,P_n). \]
For \( \gamma = 1 \), respectively \( \gamma = 0 \), one obtains as a by-product the exponential tilting (ET) estimator, respectively the empirical likelihood (EL) estimator. The model equivalence test writes
\[ \pi_n = \mathbb{I} \left[ 2 n D_\gamma(\mathcal{M}_n,P_n) < c_{\alpha,\gamma,\delta^2} \right], \]
with the same critical values as the test based on the chi-square divergence. Irrespective of the choice of the divergence, the test retain the same basic characteristics than the test
based on the chi-square divergence. In particular, it remains invariant to any transforma-
tion of the moment restrictions. But because of the degree of freedom in the choice of the
specific divergence, there is a multiplicity of implied model equivalence hypotheses and
tests. This issue is investigated in what follows.

3.2 Choice of the Divergence

We now show that all Cressie-Read divergences are asymptotically equivalent for locally
misspecified models, so that the choice of the divergence should not matter much in prac-
tice. This also sheds some light on the practical choice of the tolerance.

Assumption C For any \( \theta \in \Theta \), \( D_\gamma (M_\theta, P) < \infty \).

Lemma 3.1 For any \( \gamma \), under any drifting sequence of probability distributions \( \tilde{P}^{(n)} \) such
that Assumptions A, B, and C hold, \( D_\gamma (M, \tilde{P}^{(n)}) \) and \( D_2 (M, \tilde{P}^{(n)}) \) are asymptotically
equivalent.

Our result entails that there is no “best” divergence to construct a testing framework.
Indeed, it makes no difference asymptotically in the definition of the model equivalence
hypothesis \( H_{1n} \), while there may be some supplementary (theoretical or practical) reason
to favor a specific divergence in a particular application. The alternative interpretations of
model equivalence for the chi-square divergence mentioned in Section 2.3 extend to other
Cressie-Read divergences, since they all are close to the chi-square divergence for a small
enough tolerance.

To show the asymptotic equivalence between different Cressie-Read divergences, we
use duality, see Kitamura (2007) and Section 7.1. The strength of the duality principle is
that dual optimization is finite-dimensional and concave. For duality to apply, one needs
a projection to exist, which is ensured by Assumption C. Basically, this requires that for
each \( \theta \) a measure \( Q \in \mathcal{M}_\theta \) exists such that \( \frac{dQ}{dP}(x) \) lies in the interior of the support of
\( \varphi_\gamma (\cdot) \). The projection of \( P \) on \( \mathcal{M}_\theta \) is then essentially unique, see Keziou and Broniatowski
(2006) for more detailed conditions on the existence and unicity of projections. Our
technical assumption may seem pretty innocuous in practice. Indeed, one can always
restrict the parameter space to the set of \( \theta \) for which a finite empirical divergence obtains.
However it may not be so when moment restrictions are misspecified. Take any function
\( \varphi_\gamma (\cdot) \) with domain \((0, \infty)\), such as the ones used for EL or ET. Hence the measure \( Q \)
should be a probability measure with the same support as $P$. In case of misspecification, such a measure may not exist. Issues of GEL estimation methods under misspecification have been documented in the literature. In particular, Schennach (2007) shows that the EL estimator can have an atypical behavior when moment restrictions are invalid, as a projection does not generally exist when the functions in $g(\cdot, \cdot)$ are unbounded. Sueishi (2013) points out that under misspecification there may exist no probability measure in $\mathcal{M}$ with a finite divergence $D_1(\mathcal{M}, P)$. By contrast, because $\varphi_2(\cdot)$ has domain $\mathbb{R}$, and since $\mathcal{M}_\theta$ includes signed measures, a solution always exists when minimizing the chi-square divergence.

As the equivalence result of Lemma 3.1 is asymptotic, it seems useful to evaluate its practical relevance. To gain some insights, we considered the simple situation where $X$ is univariate, there is no unknown parameter, and only one restriction $\mathbb{E}(\theta - X) = 0$ is imposed for selected values of $\theta$. We then computed divergences corresponding to $\gamma = 2, 1,$ and $0$. We chose $X$ with a Beta($\alpha, \beta$) distribution and we let the two parameters of the distribution vary to consider distributions with different skewness and kurtosis. Values of $\theta$ were chosen so that twice the chi-square divergence equal $(0.05)^2$, $(0.1)^2$, $(0.2)^2$, and $(0.3)^2$. Table 1 reports the corresponding divergences and relative differences. When $X \sim \text{Beta}(1, 1)$ is uniformly distributed, values of the divergences are extremely close to each other. The other symmetric case Beta(2, 2) yields similar findings. In asymmetric cases, divergences still pretty much agree for a tolerance of $(5\%)^2$ and $(10\%)^2$, and relative differences are respectively of 5% and 10% at most. When increasing the tolerance, sizeable differences appear. For a tolerance of $(20\%)^2$ or $(30\%)^2$, relative differences can reach 20% and 25%, respectively. This small simulation experiment indicates that divergences can differ substantially when misspecification is large, while they do appear to be pretty close for a small misspecification, corresponding to a tolerance up to $(10\%)^2$. This will guide us in our subsequent applications.

\footnote{Exact results can be easily determined for the chi-square divergence. For $\gamma = 1$ and 0, I used simulations to determine divergence values. Namely I ran 1000 replications on 100000 observations and computed the average divergence.}
4 Empirical Illustrations

We here apply our model equivalence tests to three selected empirical problems. This will help us to discuss the choice of the tolerance and the interpretation of the outcomes. All computations used the R package gmm, see Chaussé (2010).

4.1 Social Interactions

Graham (2008) shows how social interactions can be identified through conditional variance restrictions. He applies this strategy to assess the role of peer spillovers in learning using data from the class size reduction experiment Project STAR. His model yields conditional restrictions of the form

$$
E \left[ \rho(Z_c, \tau^2(W_{1c}), \gamma_0^2) \right] | W_{1c}, W_{2c} = 0
$$

where $Z_c$ are observations related to classroom $c$, $\tau^2(W_{1c}) = W_{1c}' \beta_0$ represents conditional heterogeneity in teacher effectiveness as a function of classroom-level covariates $W_{1c}$, $\gamma_0$ is the peers effect parameter (where $\gamma_0 = 1$ corresponds to no spillover), and $W_{2c}$ denotes class size. I focus on results concerning math test scores as reported in Graham (2008, Table 1, Column 1). In this application, the classroom-level covariates $W_{1c}$ are school dummy variables as well as a binary variable indicating whether classroom is of the regular with a full time teaching aide type, while $W_{2c}$ is binary indicating whether class size is small (13 to 17 students) as opposed to regular (22 to 25 students). Graham (2008) based estimation on the unconditional moments

$$
E \left[ W_{c}' \rho(Z_c, \tau^2(W_{1c}), \gamma_0^2) \right] = 0,
$$

where $W_c = (W_{1c}, W_{2c})$. To assess the approximate validity of the social interactions model, I use unconditional moments of the above type, where $W_c$ additionally includes some interactions between binary variables. I consider two such interactions, namely interactions of a dummy for whether a classroom is in one of the 48 larger schools with the small and regular-with-aide class type dummies. Graham (2008) argues that such interactions terms are of particular interest if within-class-type student sorting or student-teaching matching in large schools is a potential concern. The standard two-step GMM

---

4 Considering all interactions terms of school dummies with small and regular dummies would yield a large number of restrictions with respect to the sample size.
overidentification test statistic is 1.08 and does not reject the null hypothesis that the overidentifying restrictions hold. In terms of spillovers, the (GEL) estimated value of $\gamma_0^2$ is about 3.07, which is a little bit lower than the value of 3.47 reported by Graham (2008), and the p-value of a significance test of $\gamma_0^2 = 1$ (the null of no spillover) is always less than 1%.

The model equivalence tests for $\gamma = 2, 1, \text{and} 0$, provide evidence in favor of the approximate validity of the model. The results gathered in Table 3 closely agree. For $\Delta^2 = (0.1)^2$, p-values are around 13%. Thus for a significance level just above 10%, model equivalence at a tolerance $\Delta^2 = (0.1)^2$ can be accepted. The minimum tolerance that would yield to accept model equivalence for a 5% level is around $(13\%)^2$. To gain further insight, we rely on the alternative formulation of the model equivalence hypothesis

$$\mathbb{E} g_2'(X, \theta) \Sigma^{-1} \mathbb{E} g_2(X, \theta) < \Delta^2.$$ 

For ease of interpretation, we focused on correlations between the error and interactions terms. Setting $\Delta^2 = (13.24\%)^2$ and estimating the matrix $\Sigma$ (based on CUE results) yields an estimated set of correlations that can be confirmed by our test.\(^4\) This set, by definition an ellipse centered at $(0, 0)$, is represented in Figure 1. It includes correlations that are both around 4% or less. Hence the model equivalence tests seem to indicate that the social interactions model is approximately adequate for the Project STAR data.

It is interesting to contrast these findings with the ones that obtain from a more standard approach based on confidence regions. From estimation results, one can readily evaluate the 95% confidence region for the correlations between errors and interaction terms. This region is also represented in Figure 1. The confidence ellipse is centered at the empirical correlations. It is slightly wider than the model equivalence set and includes larger correlations values. Crucially, it does not include the point where both correlations are zero. This illustrates that confidence regions and model equivalence tests provide different information about the problem at hand.

4.2 Demand for Differentiated Products

In a recent paper, Nevo and Rosen (2012) consider inference in the presence of “imperfect instruments,” that is ones that are correlated with the error term of the model, and show

\(^4\)Strictly speaking, this is the largest set of correlations that is not confirmed by the test, but by a slight abuse of language, I refer to it as the smallest set that is confirmed.
how one may obtain bounds on structural parameters. They illustrate the usefulness of their method on a logit demand model for differentiated products. Market shares $s_{jt}$ for product $j$ in market $t$ are expressed as

$$\log s_{jt} - \log s_{0t} = p_{jt} \beta + w'_{jt} \Gamma + \varepsilon_{jt},$$

where $w_{jt}$, $p_{jt}$, and $\varepsilon_{jt}$ are observable characteristics, price, and unobservable characteristics of product $j$ in market $t$, and $s_{0t}$ is the market share of outside good in market $t$. Though one can control for unobserved product characteristics that are fixed over time by using product fixed effects, price is still likely endogenous. The standard approach for dealing with endogeneity of price in this setting is to use prices of the product on other markets as instrumental variables, see Hausman, Leonard, and Zona (1994) and Nevo (2001). However, some concerns may still linger on the instruments validity.

I used data from Nevo and Rosen (2012) on the ready-to-eat cereal industry at the brand-quarter-MSA (metropolitan statistical area) level. Observations are from twenty quarters on the top 25 brands (in terms of market share) for the San Francisco and Boston markets. The key variables observed for each product, market, and quarter combination are quantity sold, total revenue, and brand-level advertising. For each market, Boston and San Francisco, two instruments are used: the average price on the other markets in the New England region for Boston and northern California for San Francisco, and the average price in the other city. The number of overidentifying restrictions is thus one. Results from model equivalence tests for $\gamma = 2, 1, \text{and} 0$ are gathered in Table 3. The three test statistics closely agree. For model equivalence at $\Delta^2 = (0.1)^2$, p-values are around 90%, so that all tests fail to accept model equivalence for this tolerance. Note that the two-step GMM test statistic equals 19.53, which agrees with the GEL statistics, so a standard overidentification test would reject model validity.

Using the results of Section 2.3, we can rewrite the model equivalence hypothesis $H_{1n}$ through the divergence $D_W$ applied to one overidentification restriction only. We focused on the correlation between the error term and one specific instrument providing overidentification. Therefore we evaluated the content of our hypothesis written as

$$|\mathbb{E} g_2(X, \theta^*)| < \sigma \Delta,$$

where $\mathbb{E} g_2(X, \theta^*)$ is the correlation of the error term and the average price in the other city, and $\sigma$ is its $\sqrt{n}$ variance. In practice, we estimated the model using CUE, and from

---

5For additional information on the data source and the ready-to-eat cereal industry, see Nevo (2001).
the resulting estimates, we estimated the corresponding \( \sigma \) as 1.166. Our results from the model equivalence tests thus indicate that there is no statistical evidence that the correlation between the error term and the average price in the other city is less than \( \sigma \Delta = 1.166 \times 0.1 = 11.66\% \) in absolute value. Moreover, the minimum tolerance that would yield the reverse decision for a 5% level, as defined by (2.5), is around \((20\%)^2\) for all three tests. This can be interpreted as not finding statistical evidence that the correlation of the error term with the average price in other city is less than \(1.166 \times 0.203 = 23.67\%\). Such an outcome gives support to concerns on instruments exogeneity.

By contrast, the confidence interval on the same correlation coefficient is pretty much uninformative. Though the estimated correlation is basically zero, the standard error is so large that the confidence interval includes any possible value for this coefficient.

### 4.3 Nonlinearities in Growth Regression

I consider here a cross-country growth regression in the spirit of Mankiw and al. (1992) using data on 86 countries averaged over the 1960’s, 1970’s and 1980’s from King and Levine (1993) and further studied by Liu and Stengos (1999). Explanatory variables include GDP60, the 1960 level of GDP; POP, population growth (to which 0.05 is added to account for depreciation rate and technological change); SEC, the enrollment rate in secondary schools; INV, the share of output allocated to investment; and fixed time effects. The Solow model assumes a Cobb-Douglas aggregate technology, which yields a linear regression of growth on \( \log(INV) \), \( \log(POP) \), and \( \log(SEC) \). There is more uncertainty about the relationship to the initial GDP level. Liu and Stengos (1999) argue that the relation is actually nonlinear in the initial GDP level and in human capital based on the outcome of a joint semiparametric specification test.

I used the proposed model equivalence tests to check whether the regression is approximately linear in the initial level of GDP and human capital. The considered restrictions are \( E(UW) = 0 \), where \( U \) is the error term of the linear model, \( W \) contains each explanatory variable, and polynomials terms from order two to four of GDP and human capital. I first consider both variables of interest simultaneously, so that there is six overidentifying restrictions. Results in Table 4 show that the tests for \( \gamma = 2, 1, \) and 0 closely agree. For model equivalence at \( \Delta^2 = (0.1)^2 \), p-values are around 85%, and model equivalence is rejected. The minimum tolerance \( \Delta_{\text{inf}}^2(5\%) \) that would yield the reverse decision for a 5% level is around \((30\%)^2\), which indicates a sizeable misspecification. Hence our model
equivalence tests results agree with the findings of Liu and Stengos (1999) that there are some unaccounted nonlinearities in the simple Solow model.

I then consider nonlinearity in initial GDP and human capital separately. In each case, there are three overidentifying restrictions. For GDP60, and when considering model equivalence at $\Delta^2 = (0.1)^2$, p-values are greater than 90%. The minimum tolerance is again around (30%)^2. Thus the tests fail to accept the quasi-absence of nonlinearities in initial GDP at usual levels. For SEC, the picture is strikingly different. When considering model equivalence at $\Delta^2 = (0.1)^2$, all p-values are around 1%. Moreover, the minimum tolerances $\Delta^2_{\inf}(5\%)$ is zero for all three tests, because all test statistics are smaller than the critical value $c_{0.05,3.0}$. This constitutes strong evidence in favor of approximate linearity of growth with respect to human capital, which is accepted at level 5% regardless of how small the tolerance is. It is noteworthy that, by contrast, a confidence region for $E_{g2}(X,\theta^*)$ cannot be arbitrarily small, so our model equivalence hypothesis is not a confidence region of a special kind. Our finding that the model is approximately linear in log(SEC) does not actually contradict Liu and Stengos (1999). Indeed, their separable semiparametric model appears to be only slightly non-linear in log(SEC), as seen in their Figure 2, with a large confidence band that does not exclude a linear relation.

5 Asymptotic Properties

To analyze the properties of our tests, we rely on the concept of semiparametric power envelope. We restrict to tests that are invariant to linear transformations of the moment restrictions and of the parameters. We consider a sufficiently rich family of parametric distributions for the unknown data generating process. Namely, we use the framework of Section 2.3 and we focus on a sequence of probability distributions that are differentiable in quadratic mean. We then rely on the local asymptotic normality of the likelihood ratio and the asymptotic equivalent experiments setting, see Le Cam and Lo Yang (2000) and van der Vaart (1998). Using results in Lavergne (2014), we determine an upper bound for the power of any test that is invariant to orthogonal transformations of the restrictions. We then show that our tests, which are invariant, attain this bound. Formally, we consider the following family of probability distributions.

**Definition 2** \(\mathcal{P}\) is a family of probability distributions \(P_\lambda, \lambda \in \Lambda\), with common support and such that \(E_{P_\lambda} h(X,\lambda) = 0\). The corresponding density (or probability mass function) is
differentiable with respect to $\lambda$ for any $x$, and the density and its derivatives are dominated over $\Lambda$ by an integrable function. The family $P$ is differentiable in quadratic mean and the limiting information matrix is $J = H'V^{-1}H$, where $H = \nabla_{\lambda'} E_{P_{\lambda}} h(X, \lambda)$, and $V = \text{Var}_{P_{\lambda}} h(X, \lambda)$. It contains at least one distribution with $\lambda = (\hat{\theta}', 0')'$, where $\hat{\theta} \in \hat{\Theta}$.

Such a family of distributions can generally be constructed as a multinomial distribution, see Chamberlain (1987) who uses such a construction to study asymptotic efficiency bounds. In specific models, one can consider adapted family of distributions, see Gourieroux and Monfort (1989, Chap. 23). It is also possible to consider a family of distributions indexed by a parameter of higher dimension, but this would not affect the main analysis.

The following result shows that the model equivalence tests attain the local asymptotic power envelope of tests of $H_0$ against $H_1$ for any parametric sub-family of models $P$. Here local means that we are considering parameters value around $\lambda = (\hat{\theta}', 0')'$. However, the result is independent of the specific value of $\hat{\theta}$ or the precise form of the distributions $P_{\lambda}$. We consider the supplementary Assumption D given in Section 7.3. This corresponds to the technical conditions in Broniatowski and Keziou (2012) that allow to study asymptotics of GEL estimators under misspecification, see Newey and Smith (2004) for the corresponding assumption for a well specified model.

**Theorem 5.1** Suppose $X_1, \ldots, X_n$ are i.i.d. according to $P_{\lambda} \in P$ as defined above, and that Assumptions A, B, C, and D hold.

(A) Let $\varphi_n$ be a pointwise asymptotically level $\alpha$ tests sequence, that is
\[
\limsup_{n \to \infty} E_{P_{\lambda}} (\varphi_n) \leq \alpha \quad \forall P_{\lambda} \in H_0 \cap P.
\]
Let $M > 0$ arbitrary large and $\mathcal{N}(\bar{\lambda}, M) = \{\bar{\lambda} + n^{-1/2} \Upsilon, \ U \in \mathbb{R}^m, \|\Upsilon\| \leq M\}$. If $\varphi_n$ is invariant to orthogonal transformations of the parameters and of the moment restrictions, then for all $\nu^2 < \delta^2$
\[
\limsup_{n \to \infty} E_{P_{\lambda}} (\varphi_n) \leq \Pr \left[ \chi^2_r (\nu^2) < c_{\alpha, r, \delta^2} \right] \quad \forall P_{\lambda} \in \partial H_1 (\nu) \cap P, \quad \lambda \in \mathcal{N}(\bar{\lambda}, M), \quad (5.9)
\]
where $\partial H_1 (\nu) = \{P_{\lambda} : 2 D_\gamma (M, P_{\lambda}) = \nu^2 / n\}$.

(B) The tests sequence $\pi_n$ is pointwise asymptotically level $\alpha$ for any $P_{\lambda} \in H_0 \cap P$ with $\lambda \in \mathcal{N}(\bar{\lambda}, M)$, is invariant to orthogonal transformations of the parameters and of the moment restrictions, and is such that for all $\nu^2 < \delta^2$
\[
\limsup_{n \to \infty} E_{P_{\lambda}} (\pi_n) = \Pr \left[ \chi^2_r (\nu^2) < c_{\alpha, r, \delta^2} \right] \quad \forall P_{\lambda} \in \partial H_1 (\nu) \cap P, \quad \lambda \in \mathcal{N}(\bar{\lambda}, M).
\]
The introduction of the set $\partial H_{1n}(\nu)$ allows to focus on alternatives distant from the null hypothesis for which power is not trivial. Our result shows that our model equivalence test attains the power envelope of tests of $H_{0n}$ that are invariant to orthogonal transformations. But tests that are also invariant to possibly nonlinear transformations cannot be more powerful. Hence our test asymptotically reaches the semiparametric power envelope of invariant tests.

6 Conclusion

We have proposed a new theoretical framework to assess the approximate validity of overidentifying moment restrictions. Approximate validity is evaluated through a Cressie-Read divergence between the true probability measure and the closest measure that imposes the moment restrictions of interest. The considered alternative hypothesis states that the divergence is smaller than some user-chosen tolerance. A model equivalence test is built on the corresponding empirical divergence, and attains the local semiparametric power envelope of invariant tests. Using three empirical applications, we have illustrated the usefulness of model equivalence testing and we have shown that it provides complementary information on potential misspecification compared to standard procedures.

7 Proofs

We use the following notations. For a real-valued function $l(x, \cdot)$, $\nabla l(x, \cdot)$ and $\nabla^2 l(x, \cdot)$ respectively denote the column vector of first partial derivatives and the matrix of second derivatives with respect to its second vector-valued argument. We use indices for derivatives with respect to specific arguments.
7.1 Preliminaries: Duality

Let $\psi_\gamma (\cdot)$ be the so-called convex conjugate of $\varphi_\gamma (\cdot)$, defined as $\psi_\gamma (y) = \sup_x \{ xy - \varphi_\gamma (x) \}$.

For the Cressie-Read family of functions, the convex conjugates are

$$
\psi_\gamma (y) = \gamma^{-1} \left[ (\gamma y - y + 1)^{\gamma-1} - 1 \right], \quad \gamma \in \mathbb{R} \setminus \{0, 1\}
$$

$$
\psi_1 (y) = \exp(y) - 1,
$$

$$
\psi_0(y) = -\log(1-y),
$$

where the domain may vary depending on $\gamma$. By definition, the convex conjugate is strictly convex on its domain, and due to our definition, $\psi_\gamma (0) = 0$, $\psi'_\gamma (0) = 1$, and $\psi''_\gamma (0) = 1$.

For $t \in \mathbb{R}^{m+1}$ let $m_\gamma (X, \theta, t) = t_0 - \psi_\gamma (t_0 + \sum_{i=1}^{m+1} t_i g_i (X, \theta))$. Provided it applies, duality implies that

$$
D_\gamma (\mathcal{M}, P) = \inf_\Theta \sup_{t \in \mathbb{R}^{m+1}} \mathbb{E} m_\gamma (X, \theta, t) \tag{7.10}
$$

and

$$
D_\gamma (\mathcal{M}_n, P_n) = \inf_\Theta \sup_{t \in \mathbb{R}^{m+1}} \mathbb{E}_n m_\gamma (X, \theta, t). \tag{7.11}
$$

We now detail some key properties that will be used in our proofs. We let $\tilde{g} (X, \theta) = (\mathbb{I} (X \in \mathbb{R}^p), g' (X, \theta))^t$ so that $m_\gamma (X, \theta, t) = t_0 - \psi_\gamma (t'_0 \tilde{g} (X, \theta))$, where $t = (t_0, t_1, \ldots t_m)^t$.

Properties of $\mathbb{E} m_\gamma (\cdot, \cdot, \cdot)$

1. $\mathbb{E} m_\gamma (X, \cdot, \cdot)$ is twice continuously differentiable in $t \in \mathcal{T}_0$ and in $\theta$.

   This comes from Assumption B and the differentiability of Cressie-Read divergences.

2. It is also strictly concave in $t$ for all $\theta$ since $\psi(\cdot)$ is strictly convex.

3. Denoting $\delta_0 = (1, 0, \ldots 0)^t$, derivatives are

   $$
   \nabla \mathbb{E} m_\gamma (X, \theta, t) \equiv \begin{bmatrix}
   \nabla_\theta \mathbb{E} m_\gamma (X, \theta, t) \\
   \nabla_t \mathbb{E} m_\gamma (X, \theta, t)
   \end{bmatrix} = \mathbb{E} \begin{bmatrix}
   -\psi'_\gamma (t'_\tilde{g} (X, \theta)) \nabla_\theta t'_\tilde{g} (X, \theta) \\
   \delta_0 - \psi'_\gamma (t'_\tilde{g} (X, \theta)) \tilde{g} (X, \theta)
   \end{bmatrix}
   $$

   $$
   \nabla^2_{\theta\theta} \mathbb{E} m_\gamma (X, \theta, t) = \mathbb{E} \begin{bmatrix}
   -\psi''_\gamma (t'_\tilde{g} (X, \theta)) \nabla_\theta t'_\tilde{g} (X, \theta) \\
   \nabla_\theta t'_\tilde{g} (X, \theta) \nabla_\theta t'_\tilde{g} (X, \theta)
   \end{bmatrix},
   $$

   $$
   \nabla^2_{\theta t} \mathbb{E} m_\gamma (X, \theta, t) = \mathbb{E} \begin{bmatrix}
   -\psi''_\gamma (t'_\tilde{g} (X, \theta)) \nabla_\theta t'_\tilde{g} (X, \theta) \\
   \nabla_\theta t'_\tilde{g} (X, \theta) \nabla_\theta t'_\tilde{g} (X, \theta)
   \end{bmatrix},
   $$

   $$
   \nabla^2_{tt} \mathbb{E} m_\gamma (X, \theta, t) = \mathbb{E} \begin{bmatrix}
   -\psi''_\gamma (t'_\tilde{g} (X, \theta)) \tilde{g} (X, \theta) \tilde{g} (X, \theta)
   \end{bmatrix}.
   $$
From these results, $E m_\gamma(X, \theta, 0) = 0$,

$$\nabla E m_\gamma(X, \theta, 0) = \begin{bmatrix} 0 \\ 0 \\ -E g(X, \theta) \end{bmatrix}$$

$$\nabla^2 E m_\gamma(X, \theta, 0) = \begin{bmatrix} 0 & -E \nabla_\theta g(X, \theta) \\
\cdot & -E \tilde{g}(X, \theta) \tilde{g}'(X, \theta) \end{bmatrix}.$$

**Properties of $\bar{t}(\cdot)$** Recall $\bar{t}(\cdot) = \sup_{T_\theta} E m_\gamma(X, \theta, t)$.

1. The function $\bar{t}(\cdot)$ is well-defined.

   Existence for any $\theta$ is ensured by Assumptions A and C. By Assumption B, $\text{Var} g(X, \theta)$ is positive definite, and hence the functions in $g(X, \theta)$ are linearly independent, so uniqueness is ensured, see e.g. Keziou and Broniatowski (2006).

2. The function $\bar{t}(\cdot)$ is continuous and twice differentiable on $\Theta$ by the properties of $\psi(\cdot)$ and $g(X, \cdot)$.

3. $\bar{t}(\cdot)$ admits at most one root.

   Indeed, $\bar{t}(\theta) = 0 \Rightarrow \sup_{T_\theta} E m_\gamma(X, \theta, t) = 0 \Rightarrow D_\gamma(M_\theta, P) = 0 \Rightarrow E g(X, \theta) = 0 \Rightarrow \theta = \theta^*$ for a unique $\theta^*$ by Assumption A.

4. Conversely, if there exists $\theta^*$ such that $E g(X, \theta^*) = 0$, then $\bar{t}(\theta^*) = 0$. This is because on the one hand, $E m_\gamma(X, \theta^*, \bar{t}(\theta^*)) = \sup_{T_\theta} E m_\gamma(X, \theta^*, t) = 0$, and on the other hand, $E m_\gamma(X, \theta^*, 0) = 0$, $\nabla_t E m_\gamma(X, \theta^*, 0) = 0$, and $E m_\gamma(X, \theta^*, t)$ is strictly concave in $t$.

**7.2 Proof of Lemmas 2.1 and 3.1**

We show the two lemmas in a compact way. Let $\tilde{h}(X, \lambda) = \langle 1(X \in \mathbb{R}^p), h'(X, \lambda) \rangle$ and $m_\gamma(X, \lambda, t) = t_0 - \psi_\gamma \left( t' \tilde{h}(X, \lambda) \right)$, where $t = (t_0, t_1, \ldots, t_m)$. Under Assumptions A, B, and C, there is a unique $\lambda^*$ such that

$$0 = \inf_{\lambda} D_\gamma(M_\lambda, P) = \inf_{\lambda} \sup_{t} E m_\gamma(X, \lambda, t) = \sup_{t} E m_\gamma(X, \lambda^*, t) = E m_\gamma(X, \lambda^*, 0).$$

Moreover, there exist unique $\lambda^*_R$ and $t^*_R = \bar{t}(\lambda^*_R)$ such that

$$D_\gamma(M, P) = \inf_{\Theta \times 0} \sup_{t} E m_\gamma(X, \lambda, t) = \sup_{t} E m_\gamma(X, \lambda^*_R, t) = E m_\gamma(X, \lambda^*_R, t^*_R). \quad (7.12)$$
(i). If \( D_\gamma(\mathcal{M}, P) = o(1) \), then \( 0 = \mathbb{E} m_\gamma(X, \lambda^*_R, 0) \leq \mathbb{E} m_\gamma(X, \lambda^*_R, t^*_R) = o(1) \), and it follows that \( \| t^*_R \| = o(1) \) since \( \mathbb{E} m_\gamma(X, \lambda^*_R, t) \) is strictly concave in \( t \). Since \( \bar{t}(\lambda^*) = \mathbf{0} \), and \( \bar{t}(\cdot) \) admits only one root, it must be that \( \| \lambda^*_R - \lambda^* \| = o(1) \). By a Taylor expansion of \( \mathbb{E} m_\gamma(X, \lambda, t) \) and using the continuity of \( \nabla^2 \mathbb{E} m_\gamma(X, \lambda, t) \) for \( \| t \| = o(1) \), we obtain that uniformly in \( (\lambda, t) \) in a \( o(1) \) neighborhood of \( (\lambda^*, 0) \)

\[
\mathbb{E} m_\gamma(X, \lambda, t) = \left[ - (\lambda - \lambda^*)' \nabla_\lambda \mathbb{E} \tilde{h}(X, \lambda^*) t - \frac{1}{2} t' \nabla^2 \mathbb{E} \tilde{h}(X, \lambda^*) h'(X, \lambda^*) t \right] (1 + o(1)). \tag{7.13}
\]

We can then solve for \( \bar{t}(\lambda) \) to get

\[
\sup_t \mathbb{E} m_\gamma(X, \lambda, t) = \frac{1}{2} (\lambda - \lambda^*)' J (\lambda - \lambda^*) (1 + o(1)), \tag{7.14}
\]

with \( J = J(\lambda^*) = H(\lambda^*)' V(\lambda^*)^{-1} H(\lambda^*) \), \( H(\lambda^*) = \nabla_\lambda \mathbb{E} h(X, \lambda^*) \), and \( V(\lambda^*) = \text{Var} h(X, \lambda^*) = \text{Var} g(X, \theta^*) \). Hence

\[
\mathbb{E} m_\gamma(X, \lambda^*_R, \bar{t}(\lambda^*_R)) = \frac{1}{2} (\lambda^*_R - \lambda^*)' J (\lambda^*_R - \lambda^*) (1 + o(1)) = D_\lambda H(\mathcal{M}, P)(1 + o(1)).
\]

(ii). Solving (7.14) for \( \lambda^*_R \) under the constraint \( R' \lambda = [\mathbf{0}, I_{m-p}] \lambda = \mathbf{0} \) yields

\[
\lambda^*_R = J^{-1/2} [I - P] J^{1/2} \lambda^*(1 + o(1)), \quad D_\lambda(\mathcal{M}, P) = \frac{1}{2} \lambda^* J^{1/2} P J^{1/2} \lambda^*(1 + o(1)) = \frac{1}{2} \lambda^* \Sigma^{-1} \lambda^*(1 + o(1)) = D_X(\mathcal{M}, P)(1 + o(1)),
\]

where

\[
\Sigma = R' J^{-1} R \quad \text{and} \quad P = J^{-1/2} R [R' J^{-1} R]^{-1} R' J^{-1/2}. \tag{7.15}
\]

(iii). If \( D_\lambda H(\mathcal{M}, P) = o(1) \), then it follows that \( \| \lambda^*_R - \lambda^* \| = o(1) \) from Assumption A, and thus \( t^*_R = o(1) \). Using (7.14) and (7.12), this implies that \( D_\lambda(\mathcal{M}, P) = D_X(\mathcal{M}, P)(1 + o(1)) \) by (i).

(iv). If \( D_X(\mathcal{M}, P) = o(1) \), there exists \( \theta^* \) such that \( \| \mathbb{E} g(X, \theta^*) \| = o(1) \), so that

\[
0 \leq D_2(\mathcal{M}, P) \leq \frac{1}{2} \mathbb{E} \left( g'(X, \theta^*) \right) [\text{Var} g(X, \theta^*)]^{-1} \mathbb{E} \left( g(X, \theta^*) \right) = o(1).
\]

This implies in turn that \( D_\lambda H(\mathcal{M}, P) = o(1) \) and thus that \( D_\lambda(\mathcal{M}, P) = D_X(\mathcal{M}, P)(1 + o(1)) \) by (i) and (ii).
7.3 Proof of Theorem 5.1

The following assumption is needed to establish asymptotic normality of the GEL estimators using duality under misspecification, see Broniatowski and Keziou (2012). Newey and Smith (2004) impose weaker regularity conditions, but deal with well specified models.

**Assumption D** (i) $\mathbb{E} \sup_{\theta \in \Theta} \|g(X, \theta)\|^\alpha < \infty$ for some $\alpha > 2$

(ii) Let $T_\theta = \{ t \in \mathbb{R}^{1+m} : \mathbb{E} |\psi_\gamma (t_0 + \sum_{i=1}^m t_i g_i(X, \theta))| < \infty \}$.

Then $\hat{\theta}_0 = \arg\inf_{\theta} \sup_{T_\theta} \mathbb{E} m_\gamma(X, \theta, t)$ exists, is unique, and belongs to $\Theta$. Moreover, for some neighborhood $N_{\hat{\theta}_0}$ of $\hat{\theta}_0$, $\mathbb{E} \sup_{\theta \in N_{\hat{\theta}_0}} \|\nabla g(X, \theta)\| < \infty$.

(iii) Let $\bar{d}(\theta) = \sup_{T_\theta} \mathbb{E} m_\gamma(X, \theta, t)$. Then $\mathbb{E} \sup_{\theta \in \Theta} \sup_{t \in \bar{d}(\theta)} |m_\gamma(X, \theta, t)| < \infty$, where $N_{\bar{d}(\theta)} \subset T_\theta$ is a compact set such that $\bar{d}(\theta) \in N_{\bar{d}(\theta)}$.

(i). Recall that with $J = J(\lambda^*) = H(\lambda^*) V(\lambda^*)^{-1} H(\lambda^*)$, $H(\lambda^*) = \nabla \lambda \mathbb{E} h(X, \lambda^*)$, and $V(\lambda^*) = \text{Var} h(X, \lambda^*) = \text{Var} g(X, \theta^*)$. The proof of Lemma 3.1 yields that $2D_\gamma(M, P) = \lambda^* J^{1/2} P J^{1/2} \lambda^* (1 + o(1))$, uniformly in $\lambda^* \in \mathcal{N}(\bar{\lambda}, M)$, where $P$ is defined in (7.15). Moreover, and also uniformly in $\lambda^* \in \mathcal{N}(\bar{\lambda}, M)$, we have $J = J(\bar{\lambda}) + o(1) = \bar{J} + o(1)$, and similarly $\bar{P} = \bar{P} + o(1)$ with self-explanatory notations. Since $\bar{P} \bar{J}^{1/2} = \bar{J}^{-1/2} R [\bar{R'} \bar{J}^{-1} R']^{-1} R' \bar{\lambda} = \mathbf{0}$,

\[
2D_\gamma(M, P) = \lambda^* J^{1/2} P \bar{J}^{1/2} \lambda^* (1 + o(1)) = (\lambda^* - \bar{\lambda})' \bar{J}^{1/2} P \bar{J}^{1/2} (\lambda^* - \bar{\lambda}) (1 + o(1)) \approx n^{-1} \gamma' J^{1/2} P \bar{J}^{1/2} \gamma (1 + o(1)). \tag{7.16}
\]

Let $\hat{\lambda}$ be the minimum empirical divergence estimator of $\lambda^*$, that is the argument minimizing $2 \inf_\Lambda D_\gamma(M_{\lambda^*}, P_n)$. Using a reasoning similar to Lemma 3.1’s proof for the empirical problem yields

\[
2n D(M_n, P_n) \approx n \hat{\lambda}' J_n^{1/2} P_n J_n^{1/2} \hat{\lambda} (1 + o_p(1)) \tag{7.17}
\]

with $P_n = J_n^{-1/2} R [\bar{R'} J_n^{-1} R]^{-1} R' J_n^{-1/2}$, $J_n = H_n V_n^{-1} H_n$, $H_n = \nabla \lambda \mathbb{E}_n h(X, \hat{\lambda})$, and $V_n = \text{Var}_n g(X, \hat{\theta})$.

(ii). If we assume correct specification of the moment restrictions, that is $\lambda = \bar{\lambda} = (\bar{\theta}, \mathbf{0})$, standard tools, see e.g. Newey and Smith (2004, Theorem 3.2) or Broniatowski and Keziou (2012, Theorem 5.6), yield that under Assumptions A, B, C, and D,

\[
\sqrt{n} (\hat{\lambda} - \bar{\lambda}) = -J^{-1} \bar{H}' \bar{V}^{-1} \sqrt{n} \mathbb{E}_n h(X, \bar{\lambda}) \xrightarrow{d} N(0, \bar{J}^{-1}),
\]

where $\bar{J} = J(\bar{\lambda})$, and similarly for $\bar{H}$ and $\bar{V}$. Moreover, $J_n = \bar{J} + o_p(1)$ and $P_n = \bar{P} + o(1)$. Let us now look at the behavior of $\hat{\lambda}$ under local misspecification. Local asymptotic
normality of the log-likelihood ratio, which follows from the assumption that the model is differentiable in quadratic mean over \( \Lambda \), see van der Vaart (1998, Theorem 7.2), yields

\[
n^{1/2} \ln \prod_{i=1}^{n} \frac{f(X_i; \lambda)}{f(X_i; \bar{\lambda})} = (\lambda - \bar{\lambda})' \Delta_n - (\lambda - \bar{\lambda})' \bar{J} (\lambda - \bar{\lambda}) / 2 + o_p(1) \quad \forall \lambda,
\]

with

\[
\Delta_n = n^{-1/2} \sum_{i=1}^{n} \nabla_{\lambda} \log f(X_i; \lambda) \xrightarrow{d} N(0, J^{-1}),
\]

\[
\bar{J} = \mathbb{E} \nabla_{\lambda} \log f(X; \bar{\lambda}) \nabla_{\lambda}' \log f(X; \bar{\lambda}) = J^{-1} \bar{H}.
\]

Since \( \mathbb{E} h(X, \bar{\lambda}) = 0 \), total differentiation yields

\[
\text{Cov} \left( h(X, \bar{\lambda}), \nabla_{\lambda} \log f(X; \bar{\lambda}) \right) = -\nabla_{\lambda} \mathbb{E} h(X, \bar{\lambda}).
\]

Hence,

\[
\text{Cov} \left( \sqrt{n} (\bar{\lambda} - \hat{\lambda}), \Delta_n \right) = -n J^{-1} \bar{H} \bar{V}^{-1} \text{Cov} \left( \mathbb{E} n h(X, \bar{\lambda}), \mathbb{E} n \nabla_{\lambda} \log f(X; \bar{\lambda}) \right)
\]

\[
= -n J^{-1} \bar{H} \bar{V}^{-1} \bar{H} = -I_m. \tag{7.18}
\]

Therefore by Le Cam’s third Lemma, see e.g. van der Vaart (1998), we obtain that under the sequences of distributions corresponding to \( \lambda = \bar{\lambda} + \frac{1}{2} \Upsilon \),

\[
\tau_n \equiv \sqrt{n} (\bar{\lambda} - \hat{\lambda}) \equiv Z + o_p(1),
\]

where \( Z \sim N(-\Upsilon, J^{-1}) \). As a consequence,

\[
n (\bar{\lambda} - \hat{\lambda})' J_n^{1/2} P_n J_n^{1/2} (\bar{\lambda} - \hat{\lambda})' = Z' J^{-1/2} \bar{P} \bar{J}^{1/2} Z + o_p(1).
\]

(iii). Since the sequence of distributions converges to a limiting normal experiment \( Z \) with unknown mean \(-\Upsilon\) and known covariance matrix \( J^{-1} \), it follows that we can approximate pointwise the power of any test \( \varphi_n \) by the power of a test in the limit experiment, see van der Vaart (1998, Theorem 15.1) and Lehmann and Romano (2005, Theorem 13.4.1). Now we apply the following result.

**Lemma 7.1 (Lavergne (2014, Lemma 4.2))** Consider testing

\[
H_0 : \mu' \Omega^{-1/2} P \Omega^{-1/2} \mu \geq \delta^2 \quad \text{against} \quad H_1 : \mu' \Omega^{-1/2} P \Omega^{-1/2} \mu < \delta^2,
\]

where \( P \) is a known orthogonal projection matrix of rank \( r \), from one observation \( Z \in \mathbb{R}^p \) distributed as a multivariate normal \( N(\mu, \Omega) \) with unknown mean \( \mu \) and known nonsingular
covariance matrix $\Omega$. Then the test $\pi(z)$ that rejects $H_0$ when $Z'\Omega^{-1/2}P\Omega^{-1/2}Z < c_{\alpha,r,\delta^2}$ is of level $\alpha$. For any $\nu^2 < \delta^2$, the test is maximin among $\alpha$-level tests of $H_0$ against $H_1(\nu): \mu'\Omega^{-1/2}P\Omega^{-1/2}\mu \leq \nu^2$ with guaranteed power $\Pr[\chi^2(\nu^2) < c_{\alpha,r,\delta^2}]$.

In our case, the test writes $\pi(Z) = \mathbb{1}[Z'\hat{J}^{1/2}\hat{P}\hat{J}^{1/2}Z < c_{\alpha,r,\delta^2}]$. Since the test is maximin, it is necessarily admissible and unbiased. Moreover, as it is independent of $\nu^2$, it must be most powerful against $\Upsilon = 0$. Finally, as it is invariant to orthogonal transformations of the parameter space, it must be UMP invariant.

(iv). For $\lambda \in \mathcal{N}(\hat{\lambda}, M)$, the model equivalence test $\pi_n$ is asymptotically equivalent to $\pi(\tau_n)$, where $\pi(\cdot)$ is the test defined above and $\tau_n = \sqrt{n}(\hat{\lambda} - \lambda)$. It thus remains to check that $\pi_n$ has the same local asymptotic properties as the optimal test $\pi(Z)$ in the limiting experiment.

We have $\mathbb{E}\pi_n = \mathbb{E}\pi(\tau_n) + o(1)$ pointwise in $\Upsilon \in \mathbb{R}^m$. Also $n\tau_n'J_n^{1/2}P_nJ_n^{1/2}\tau_n$ is for any $\Upsilon$ asymptotically distributed as a non-central $\chi^2_{m-r}(\Upsilon J^{1/2}P\hat{J}^{1/2}\Upsilon)$, see Rao and Mitra (1972). As $\pi(\tau_n)$ rejects $H_0(n)$ when $\tau_nJ_n^{1/2}P_nJ_n^{1/2}\tau_n < c_{\alpha,r,\delta^2}$,

$$
\mathbb{E}\hat{\lambda}^{-1/2}\Upsilon \pi(\tau_n) = \mathbb{P}\hat{\lambda}^{-1/2}\Upsilon \left[\tau_n'J^{1/2}P\hat{J}^{1/2}\tau_n < c_{\alpha,r,\delta^2}\right] \rightarrow \mathbb{P}\left[\chi^2_\nu(\Upsilon J^{1/2}P\hat{J}^{1/2}\Upsilon) < c_{\alpha,r,\delta^2}\right].
$$

Hence, $\pi(\tau_n)$ and thus $\pi_n$ are locally pointwise asymptotic level $\alpha$.

The proof of Lemma 7.1 in Lavergne (2014) shows that $\pi(Z)$ is a $\alpha$-level Bayes test of

$$
H_0: \Upsilon'\hat{J}^{1/2}\hat{P}\hat{J}^{1/2}\Upsilon \geq \delta^2 \quad \text{against} \quad H_1(\nu): \Upsilon'\hat{J}^{1/2}\hat{P}\hat{J}^{1/2}\Upsilon \leq \nu^2
$$

for $\nu^2 < \delta^2$ under least favorable a priori measures, which are respectively the uniform measure $Q_\nu$ on the domain $S(\delta)$ such that $\Upsilon'\hat{J}^{1/2}\hat{P}\hat{J}^{1/2}\Upsilon = \delta^2$ and the uniform measure $Q_\nu$ defined similarly. Now

$$
\mathbb{E}_{Q_\nu}\pi(\tau_n) = \int_{S(\nu)} \mathbb{E}\pi(\tau_n) dQ_\nu \rightarrow \mathbb{E}_{Q_\nu}\pi(Z)
$$

by the Lebesgue dominated convergence theorem, so that $\pi(\tau_n)$ and thus $\pi_n$ are also asymptotically Bayesian level $\alpha$ for the same a priori measures. For any other test sequence $\varphi_n$ of asymptotically Bayesian level $\alpha$,

$$
\limsup_{n \to \infty} \inf_{H_1(\nu)} \mathbb{E}\varphi_n \leq \limsup_{n \to \infty} \mathbb{E}_{Q_\nu}\varphi_n \leq \limsup_{n \to \infty} \mathbb{E}_{Q_\nu}\pi(\tau_n).
$$

But $\limsup_{n \to \infty} \mathbb{E}_{Q_\nu}\pi(\tau_n) = \mathbb{E}_{Q_\nu}\pi(Z) = \inf_{H_1(\nu)} \mathbb{E}\pi(Z) = \lim_{n \to \infty} \inf_{H_1(\nu)} \mathbb{E}\pi(\tau_n)$. Gathering results,

$$
\liminf_{n \to \infty} \left( \inf_{H_1(\nu)} \mathbb{E}\pi(\tau_n) - \inf_{H_1(\nu)} \mathbb{E}\varphi_n \right) \geq 0,
$$

27
which shows that $\pi(\tau_n)$ and thus $\pi_n$ are locally asymptotically maximin.

Consider an invariant test sequence $\varphi_n$ of pointwise asymptotic level $\alpha$. Then for any $\nu$ and any $\Upsilon$ such that $\Upsilon J_1 / 2 \bar{P} J_1 / 2 \Upsilon = \nu^2$

$$\limsup_{n \to \infty} \mathbb{E} \lambda_n^{-1/2} \Upsilon \varphi_n \leq \limsup_{n \to \infty} \mathbb{E} Q_n \varphi_n \leq \limsup_{n \to \infty} \mathbb{E} Q_n \pi(\tau_n) = \lim_{n \to \infty} \mathbb{E} \lambda_n^{-1/2} \Upsilon \pi(\tau_n),$$

so that $\pi(\tau_n)$ and thus $\pi_n$ have maximum asymptotic local power among invariant tests.

Since the power of $\pi(\tau_n)$ converges to a bounded function which is continuous in $\Upsilon$, limits of extrema on $H_1(\nu)$ equal limits of extrema on $H_{1n}(\nu)$: $2D_{\gamma}(\mathcal{M}, P) < \nu^2/n$, using (7.16). Hence the same local asymptotic properties hold for $\pi(\tau_n)$ and thus $\pi_n$ as tests of $H_{0n}$ against $H_{1n}(\nu)$.

References


Table 1: Divergences for $X \sim \text{Beta}(\alpha, \beta)$ and $g(X, \theta) = \theta - X$

<table>
<thead>
<tr>
<th>$\alpha, \beta$</th>
<th>$\sqrt{\text{Var}(X)} \sqrt{2 D_2(Q_{\gamma}^*, P)}$</th>
<th>$\sqrt{2 D_1(Q_{\gamma}^*, P)}$</th>
<th>$100 \frac{D_1(Q_{\gamma}^<em>, P)}{D_2(Q_{\gamma}^</em>, P)} - 1$</th>
<th>$\sqrt{2 D_0(Q_{\gamma}^*, P)}$</th>
<th>$100 \frac{D_0(Q_{\gamma}^<em>, P)}{D_2(Q_{\gamma}^</em>, P)} - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1, 1)$</td>
<td>0.050</td>
<td>0.050</td>
<td>0.948</td>
<td>0.050</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.100</td>
<td>0.197</td>
<td>0.100</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.200</td>
<td>0.454</td>
<td>0.200</td>
<td>0.453</td>
</tr>
<tr>
<td></td>
<td>0.300</td>
<td>0.302</td>
<td>1.005</td>
<td>0.301</td>
<td>1.002</td>
</tr>
<tr>
<td>$(2, 2)$</td>
<td>0.050</td>
<td>0.050</td>
<td>0.923</td>
<td>0.050</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.100</td>
<td>0.021</td>
<td>0.100</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.200</td>
<td>0.208</td>
<td>0.200</td>
<td>0.367</td>
</tr>
<tr>
<td></td>
<td>0.300</td>
<td>0.301</td>
<td>0.567</td>
<td>0.299</td>
<td>0.734</td>
</tr>
<tr>
<td>$(2, 4)$</td>
<td>0.050</td>
<td>0.050</td>
<td>0.310</td>
<td>0.050</td>
<td>0.540</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.099</td>
<td>1.658</td>
<td>0.098</td>
<td>3.395</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.197</td>
<td>2.960</td>
<td>0.193</td>
<td>6.771</td>
</tr>
<tr>
<td></td>
<td>0.300</td>
<td>0.294</td>
<td>3.961</td>
<td>0.284</td>
<td>10.226</td>
</tr>
<tr>
<td>$(1/2, 1)$</td>
<td>0.050</td>
<td>0.050</td>
<td>1.691</td>
<td>0.049</td>
<td>2.719</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.099</td>
<td>2.093</td>
<td>0.098</td>
<td>4.074</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.197</td>
<td>3.417</td>
<td>0.193</td>
<td>7.127</td>
</tr>
<tr>
<td></td>
<td>0.300</td>
<td>0.292</td>
<td>5.135</td>
<td>0.284</td>
<td>10.322</td>
</tr>
<tr>
<td>$(1/2, 2)$</td>
<td>0.050</td>
<td>0.049</td>
<td>2.179</td>
<td>0.049</td>
<td>4.201</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.098</td>
<td>3.608</td>
<td>0.096</td>
<td>7.496</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.193</td>
<td>7.254</td>
<td>0.185</td>
<td>14.472</td>
</tr>
<tr>
<td></td>
<td>0.300</td>
<td>0.284</td>
<td>10.269</td>
<td>0.268</td>
<td>20.439</td>
</tr>
<tr>
<td>$(1/2, 3)$</td>
<td>0.050</td>
<td>0.049</td>
<td>2.575</td>
<td>0.049</td>
<td>5.127</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.098</td>
<td>4.717</td>
<td>0.095</td>
<td>9.601</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.191</td>
<td>9.123</td>
<td>0.181</td>
<td>18.193</td>
</tr>
<tr>
<td></td>
<td>0.300</td>
<td>0.280</td>
<td>12.703</td>
<td>0.259</td>
<td>25.548</td>
</tr>
</tbody>
</table>

Table 2: Equivalence tests results for social interactions model

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>Test statistic</th>
<th>$\Delta^2(0.1)^2$</th>
<th>$\delta^2_{\text{inf}}(5%)$</th>
<th>$\Delta^2_{\text{inf}}(5%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.108</td>
<td>0.127</td>
<td>5.557</td>
<td>(13.24%)^2</td>
</tr>
<tr>
<td>1</td>
<td>1.139</td>
<td>0.131</td>
<td>5.649</td>
<td>(13.35%)^2</td>
</tr>
<tr>
<td>0</td>
<td>1.157</td>
<td>0.133</td>
<td>5.70</td>
<td>(13.41%)^2</td>
</tr>
</tbody>
</table>
Figure 1: Equivalence hypothesis and confidence region in terms of correlations

Table 3: Equivalence tests results for demand model

<table>
<thead>
<tr>
<th></th>
<th>$\gamma = 2$</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test statistic</td>
<td>22.50</td>
<td>20.59</td>
<td>20.67</td>
</tr>
<tr>
<td>P-value ($\Delta^2 = (0.1)^2$)</td>
<td>0.94</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>$\delta_{\inf}^2$(5%)</td>
<td>40.81</td>
<td>38.22</td>
<td>38.43</td>
</tr>
<tr>
<td>$\Delta_{\inf}^2$(5%)</td>
<td>$(20.30%)^2$</td>
<td>$(19.65%)^2$</td>
<td>$(19.70%)^2$</td>
</tr>
</tbody>
</table>

Table 4: Equivalence tests results for growth regression

<table>
<thead>
<tr>
<th></th>
<th>$\gamma = 2$</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinearities in initial GDP and human capital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test statistic</td>
<td>13.37</td>
<td>13.60</td>
<td>13.35</td>
</tr>
<tr>
<td>P-value ($\Delta^2 = (0.1)^2$)</td>
<td>0.85</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>$\delta_{\inf}^2$(5%)</td>
<td>22.02</td>
<td>22.37</td>
<td>21.99</td>
</tr>
<tr>
<td>$\Delta_{\inf}^2$(5%)</td>
<td>$(29.22%)^2$</td>
<td>$(29.44%)^2$</td>
<td>$(29.20%)^2$</td>
</tr>
<tr>
<td>Nonlinearities in initial GDP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test statistic</td>
<td>12.38</td>
<td>12.39</td>
<td>11.80</td>
</tr>
<tr>
<td>P-value ($\Delta^2 = (0.1)^2$)</td>
<td>0.93</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>$\delta_{\inf}^2$(5%)</td>
<td>24.22</td>
<td>24.25</td>
<td>23.36</td>
</tr>
<tr>
<td>$\Delta_{\inf}^2$(5%)</td>
<td>$(30.64%)^2$</td>
<td>$(30.66%)^2$</td>
<td>$(30.09%)^2$</td>
</tr>
<tr>
<td>Nonlinearities in human capital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test statistic</td>
<td>0.278</td>
<td>0.292</td>
<td>0.305</td>
</tr>
<tr>
<td>P-value ($\Delta^2 = (0.1)^2$)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>$\delta_{\inf}^2$(5%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Delta_{\inf}^2$(5%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>