EFFICIENT CONDITIONALLY SIMILAR-ON-THE-BOUNDARY TESTS

José Luis Montiel Olea[1]

This paper presents a new class of tests for hypothesis testing problems with a notable feature: a *boundary-sufficient statistic*. Examples include testing in Linear Instrumental Variables Regression, testing in a class of weakly identified Generalized Method of Moments models, and testing for dynamic effects in a Structural Vector Autoregression identified using external instruments. The new tests minimize a weighted sum of the average rates of Type I and Type II error (*average risk*), while controlling the conditional rejection probability on the boundary of the null hypothesis; in this sense they are *efficient conditionally similar on the boundary* (ECS). ECS tests are *admissible* within the class of procedures that control the rate of Type I error by *conditioning* on the boundary-sufficient statistic. Moreover, they verify an important finite-sample optimality property: admissibility within the class of *all* tests, provided the boundary-sufficient statistic is boundedly complete. The theory developed in this paper yields novel—analytically optimal—tests for the examples mentioned above. This paper also shows that the Anderson and Rubin (1949) test for Linear Instrumental Variables Regression and Stock and Wright's (2000) S-tests for the Generalized Method of Moments framework are ECS in a wide variety of "just-identified" models.

1. INTRODUCTION

THERE IS A STANDARD APPROACH IN ECONOMETRICS to test statistical hypotheses in the presence of nuisance parameters. First, one finds a point estimate for the parameter of interest. Second, one approximates the distribution of the estimator's sampling error. Third, one estimates the relevant nuisance parameters. The standard test is implemented by comparing the estimator's *null* sampling error (i.e., $\widehat{\theta} - \theta_0$) with the quantiles of the estimated distribution. Despite its prevalence, there is now a large body of work—both empirical and theoretical—documenting problems with this practice in the context of several widely used models. Three important examples are: *Linear Instrumental Variables Regression* (IV) [Nelson and Startz (1990), Bound, Jaeger, and Baker (1995), Staiger and Stock (1997), Stock, Wright, and Yogo (2002)]; the *Generalized Method of Moments* (GMM) [Andersen and

[1] I am deeply indebted to my main advisors, Gary Chamberlain and James Stock, for their continuous guidance, support, patience, and encouragement. I owe special thanks to Alberto Abadie, Isaiah Andrews, Matías Cattaneo, Denis Chetverikov, Eduardo Dávila, Yuhta Ishii, Maximilian Kasy, Michal Kolesar, Timothy McQuade, Lauren Merrill, Mikkel Plagborg-Møller, Daniel Pollmann, and Guillaume Pouliot for helpful comments and suggestions. All errors are my own. E-mail: montiel@fas.harvard.edu. This version: October 31st, 2012.

Sørensen (1996), Hansen, Heaton, and Yaron (1996), Stock and Wright (2000), Mavroeidis, Plagborg-Møller, and Stock (2012)]; and *Structural Vector Autoregressions identified using external instruments* (SVAR) [Montiel Olea, Stock, and Watson (2012)]. The main practical concern is that the actual rate of Type I error for standard testing procedures (e.g., Wald tests) can be very different from the nominal target and dramatically changes with the values of the nuisance parameters.

This paper studies point and one-sided testing problems in IV, GMM, and SVARs from a different perspective. I analyze a general class of parametric hypothesis testing problems with a key characteristic: a *boundary-sufficient statistic*. Broadly speaking, a statistic X_2 is boundary sufficient if any movement along the set of null parameter values that are the closest to the alternative hypothesis—i.e., the boundary of the null—affects the distribution of the data (X_1, X_2) only through its effect on X_2. I show that this property arises naturally in the limiting experiments associated with the three aforementioned examples.[1] In each of these cases, it is possible to control the rate of Type I error by conditioning on the corresponding boundary-sufficient statistic: an adjusted vector of Ordinary Least Squares reduced-form first-stage coefficients in the IV model, an adjusted derivative of the sample moment condition in GMM, and a linear transformation of the sample covariance between the reduced form errors in the vector autoregression and the external instruments used to identify the SVAR.[2]

The main theoretical contribution of this paper is a new class of tests for hypothesis testing problems with a boundary-sufficient statistic. The *Efficient Conditionally Similar-on-the-Boundary tests* (henceforth, ECS tests) are minimizers of a weighted sum of the average rates of Type I and Type II error subject to a *conditional similarity-on-the-boundary constraint*. Their main claim for optimality, albeit decision-theoretic, is of a very applied nature: there is no other test—among those that condition the accept/reject decision on the realizations of a boundary-sufficient statistic—with smaller rates of Type I and Type II error. That is, ECS tests are *admissible* within the class of conditionally similar-on-the-boundary tests. Neither Moreira's (2003) Conditional Likelihood Ratio (CLR) for testing a point hypothesis in IV nor Kleibergen's (2007) extensions of the CLR to GMM have been shown to

[1] I use the phrase *limiting experiment* in the modern sense of Müller (2011) and not in the classical sense of Le Cam (1986). Thus, a limiting experiment refers to a statistical model derived from a set of weak convergence assumptions. Section 5 presents a detailed explanation of the concept.

[2] The remarkable paper of Moreira (2003) introduced the idea of "conditioning" as a device to control the rate of Type I error in Structural Equations Models.

satisfy this property.

The admissibility result can be further strengthened. This paper shows that ECS tests are admissible within the class of *all* tests, provided the boundary-sufficient statistic is *boundedly complete* (as defined by Lehmann and Romano (2005)) and the rates of Type I and Type II error vary continuously over the parameter space. These assumptions are satisfied in several IV, GMM, and SVAR settings. The result is relevant for applied econometrics. For instance, neither the Two-Stage Least Squares (TSLS) nor the Limited Information Maximum Likelihood (LIML) Wald tests are known to be admissible, not even in the context of a Gaussian, independent, homoskedastic model.[3] Hence, even if practitioners do not regard similarity-on-the-boundary as a desirable property—which the widespread use of the TSLS Wald test in IV regression suggests—there is still a strong justification to use ECS procedures, for it is not possible to find a non-similar test with better rates of Type I and Type II error.

The theory developed in this paper provides new insights about hypotheses testing in IV, GMM, and SVARs. There are five main results with an emphasis on point testing—in which case, ECS tests are simply maximizers of weighted average power (for a full-support *prior*) subject to a conditional similarity constraint. First, I show that the Anderson and Rubin (1949) test (henceforth, AR) is ECS in just-identified IV models with Gaussian reduced-form errors, independent observations, fixed instruments, and an arbitrary number of endogenous regressors. Furthermore, a robust version of the AR test is shown to be ECS in the limiting experiment of *weakly just-identified* IV models with heteroskedastic, autocorrelated, and/or clustered data. The priors over the structural parameters of the IV model (denoted β and Π) for which the AR test maximizes weighted average power have an interesting property: there are no other priors for which the implied distribution over the reduced-form parameters ($\Pi\beta$ and Π) is Gaussian, centered at zero, and with the same covariance matrix as the distribution of their sample counterparts.

Second, I derive new ECS tests—for point and one-sided null hypotheses—in the over-identified IV model studied by Andrews, Moreira, and Stock (2006) and

[3] Consider the linear IV regression model with a single endogenous regressor (β) under the following assumptions. Suppose that the instruments are non-stochastic (fixed) and suppose that the reduced-form errors are independent and identically distributed as a bivariate Gaussian random vector with known covariance matrix. To the best of my knowledge, there are no finite-sample optimality claims available for either the TSLS or the LIML Wald tests. In other words, there is no theoretical support for the use of the test that rejects $\mathbf{H}_0 : \beta = \beta_0$ in favor of $\mathbf{H}_1 : \beta \neq \beta_0$ for large values of $(\widehat{\beta}_{\text{TSLS}} - \beta_0)^2/(\widehat{\text{var}}(\widehat{\beta}_{\text{TSLS}} - \beta_0))$.

Chamberlain (2007). The ECS test for the point hypothesis problem enjoys basic optimality properties that neither CLR nor the TSLS (LIML) Wald tests have been shown to satisfy. The "conditional" critical region of the new test—which can be expressed in terms of the AR and the Lagrange Multiplier (LM) statistics—admits a simple interpretation: if the LM is below (above) its conventional χ^2_1 critical value, the ECS test automatically adjusts upwards (downwards) the χ^2_k threshold for the AR. The magnitude of the adjustment depends on the value of the boundary-sufficient statistic and the ECS test rejects the null hypothesis whenever the AR exceeds the adjusted critical value. This procedure is also ECS in models in which the reduced-form ordinary least-squares coefficients exhibit a "Kronecker" asymptotic covariance matrix, for example, the proportional heteroskedasticity/autocorrelation models used in Montiel Olea and Pflueger (2012).

Third, I derive a limiting experiment for GMM models with one scalar parameter and m moment conditions. The statistical experiment is derived by considering a set of Gaussian weak convergence assumptions for both the sample moment condition and its derivative. I provide a set of sufficient conditions under which the GMM S-test of Stock and Wright (2000) is ECS in the limiting experiment.

Fourth, I present general ECS tests for over-identified GMM models in which the strength of identification is controlled by a finite-dimensional nuisance parameter. The tests are specialized to non-homoskedastic and/or serially correlated weakly identified IV with one endogenous regressor. In this context, the implementation of the ECS test requires two numerical exercises. First, numerical integration is required to compute an *integrated likelihood* in the ECS test statistic. Second, Monte-Carlo methods are used to compute the quantiles of the empirical distribution of the ECS test statistic, conditional on the boundary-sufficient statistic.

Finally, I derive ECS tests for the limiting experiment of SVARs identified by external instruments, as defined in Montiel Olea, Stock, and Watson (2012). The external instruments are random variables correlated with a target shock i, uncorrelated with the other structural shocks in the model, and excluded from the vector autoregression. The object of interest is the dynamic effect of the structural shock i over variable j at horizon h. The fifth result in this paper shows that the test used by Montiel, Stock, and Watson (2012) to build confidence intervals for dynamic effects is ECS—provided there is only one external instrument for the target shock i. The ECS test rejects for large values of the sample covariance between the instrument and a linear combination of the reduced-form shocks in the vector autoregression.

This paper also presents an ECS test for the over-identified SVAR model.

The remainder of this paper is organized as follows. Section 2 presents the basic elements of a parametric testing problem (sample space, parameter space, statistical model, test, Type I/II error, risk, and admissibility) and the main regularity assumptions (which I denote TC1, TC2, C). Section 3 defines *boundary sufficiency*, which is the key concept in this paper. Throughout both sections, a Gaussian "quasi-shift" model is used to ilustrate the main concepts and assumptions. Section 4 presents the ECS tests, their main theoretical properties, and the main result concerning their implementation. Section 5 derives ECS tests for each of the examples discussed in the introduction. Section 6 presents a summary of the main results and concludes. All proofs are collected in the Appendix.

2. BASIC DEFINITONS AND ASSUMPTIONS

Section 2.1 presents the three basic elements of a parametric testing problem: sample space, parameter space, and statistical model. This section also defines the *boundary of the null hypothesis* ($\text{Bd}\Theta_0$) and presents Assumptions TC1 and TC2, both of which impose restrictions on the types of null hypotheses under consideration. A simple example (Gaussian quasi-shift model) is used to ilustrate the concepts and assumptions.

Section 2.2 defines the rates of Type I/Type II error of a test ϕ, both of which are summarized by the risk function, $R(\phi, \theta)$. Just as in classical decision theory, risk is used to define the optimality criterion for test selection: *admissibility*. Section 2.2 also introduces Assumption C, which imposes a continuity restriction on the rates of Type I and Type II error.

The main definitions in this section follow Chamberlain (2007); Chapters 2 and 5 in Ferguson (1967); and Chapter 4 in Linnik (1968).

NOTATION PRELIMINARIES: Let $\mathcal{B}(\mathbb{R}^n)$ denote the Borel σ-algebra on $\mathbb{R}^n$. For any set $\mathcal{S} \in \mathcal{B}(\mathbb{R}^n)$, let $\mathcal{B}(\mathbb{R}^n)_\mathcal{S}$ denote the sub-space σ-algebra.[4] *Measurability* of the function $f : \mathcal{S} \to \mathbb{R}$ is always relative to the measurable spaces $(\mathcal{S}, \mathcal{B}(\mathbb{R}^n)_\mathcal{S})$-$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The integral of f with respect to the lebesgue measure in $\mathbb{R}^n$ is denoted by $\int_\mathcal{S} f(s)ds$. Integration with respect to a different measure μ is denoted $\int_\mathcal{S} f(s)d\mu(s)$. All the $\mathbb{R}^m$-valued random variables in this paper are assumed to be absolutely continuous

[4] That is, $\mathcal{B}(\mathbb{R}^n)_\mathcal{S} \equiv \{\mathcal{S} \cap F \mid F \in \mathcal{B}(\mathbb{R}^n)\}$.

with respect to the lebesgue measure in $\mathbb{R}^m$, unless otherwise noted. Thus, random variables with discrete support are ruled out.

2.1. ***Basic Elements of a Parametric Testing Problem***

Sample Space, Parameter Space, and Statistical Model: There is a random variable X that takes values in the *sample space* $\mathbf{X} \subseteq \mathbb{R}^s$. There is a *parameter space* $\mathbf{\Theta} \subseteq \mathbb{R}^p$ whose elements $\theta \in \mathbf{\Theta}$ are used to index a set of probability density functions over the sample space, $X \sim f(x, \theta)$. The collection $\{\, f(\cdot, \theta) \,\}_{\theta \in \mathbf{\Theta}}$ is called a *statistical model*. The mapping $f : \mathbf{X} \times \mathbf{\Theta} \to \mathbb{R}_+$ is called the *likelihood function*. It is assumed that the sample space has a product space structure. Consequently, X can be written as a random vector (X_1, X_2) with realizations $(x_1, x_2) \in \mathbb{R}^{s_1} \times \mathbb{R}^{s_2}$, $s_1 + s_2 = s$.

Null hypothesis: Let Θ_0 be a strict subset of the parameter space. There is a null hypothesis $\mathbf{H}_0$ that states $X \sim f(x, \theta)$ for some $\theta \in \Theta_0$. The hypothesis testing problem is abbreviated $\mathbf{H}_0 : \theta \in \Theta_0$ vs. $\mathbf{H}_1 : \theta \in \Theta_1 \equiv \mathbf{\Theta} \backslash \Theta_0$, and it is denoted by the tuple $(\mathbf{X}, \mathbf{\Theta}, f, \Theta_0)$.

Boundary of the Null Hypothesis : The set $\mathrm{Bd}\Theta_0$ plays an important role in this paper. For the sake of formality, I present a general topological definition of this set. Let $\mathcal{T}$ be the subspace topology on $\mathbf{\Theta} \subseteq \mathbb{R}^p$ and let τ_θ denote an open neighborhood of $\theta \in \mathbf{\Theta}$; i.e., $\theta \in \tau_\theta$ and $\tau_\theta \in \mathcal{T}$. Define

$$\mathrm{Bd}\Theta_0 \equiv \{\theta \in \mathbf{\Theta} \mid \tau_\theta \cap \Theta_0 \neq \emptyset \quad \text{and} \quad \tau_\theta \cap \Theta_1 \neq \emptyset, \, \forall \, \tau_\theta \in \mathcal{T}\}.$$ [5]

Intuitively, the boundary of the *null set* Θ_0 contains those elements of the null that are the closest to the alternative.[6]

Assumptions concerning the structure of the null: All the hypotheses testing problems considered in this paper satisfy the following assumptions:

[5] The topological boundary of $\mathcal{A} \subseteq \mathbf{\Theta}$ is usually defined as the intersection of two sets: the closure of $\mathcal{A}$ and the closure of $\mathbf{\Theta} \backslash A$; see Munkres (2000) pp. 95, 102 (Exercise 19). The definition presented here is based on the characterization of closure provided in Munkres (2000), Theorem 17.5a, p. 96.

[6] If θ belongs to the boundary, any open ball τ_θ contains an element of the alternative hypothesis. In this sense, there is always a "nearby" element of Θ_1. If, however, θ belongs to $\Theta_0 \backslash \mathrm{Bd}\Theta_0$, then the latter statement no longer holds: there is a neighborhood of θ_0 that does not contain elements of Θ_1.

Assumption TC1: $\#\left(\mathrm{Bd}\Theta_0\right) > 1$.[7]

Assumption TC2: Θ_0 is closed relative to $(\boldsymbol{\Theta}, \mathcal{T})$.

Assumptions TC1 and TC2 imply that Θ_0 is composite: $\mathrm{Bd}\Theta_0 \subseteq \Theta_0$ and therefore the null set is not a singleton.[8] Not all hypothesis testing problems with a composite null satisfy Assumption TC1. For instance, in a one-dimensional Gaussian location model with parameter μ, the hypothesis $\mathbf{H}_0 : \mu \leq 0$ is closed and composite. However, the boundary of the null contains only one point: $\mu = 0$. The main property used in this paper, *boundary sufficiency*, is only defined for models in which $\mathrm{Bd}\Theta_0$ has more than one element.

Example—Gaussian Quasi-Shift Model: This parametric testing problem—which is intrinsically connected with a just-identified instrumental variable regression (See Section 5.1)—ilustrates the concepts discussed thus far. Let the sample space $\mathbf{X} \equiv (\mathbf{X}_1, \mathbf{X}_2) = \mathbb{R}^n \times \mathbb{R}^{n^2}$, $n \in \mathbb{N}$. Let $\boldsymbol{\mu}_1$ be an $n \times 1$ vector and let $\boldsymbol{\mu}_2 = [\mu_{21}, \mu_{22}, \ldots \mu_{2n}]$ be a $n \times n$ matrix, not necessarily of full rank. Let the parameter space be given by

$$\{\mathrm{vec}(\boldsymbol{\mu_1}, \mu_{21} \ldots \mu_{2n}) : \boldsymbol{\mu}_1 \in \mathbb{R}^n \text{ and } \ \mu_{2i} \in \mathbb{R}^n \; \forall \, i = 1 \ldots n\}.$$

Consider the statistical model:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2} \begin{pmatrix} \boldsymbol{\mu}_2 \boldsymbol{\mu}_1 & , & \mathbb{I}_{n+n^2} \\ \mathrm{vec}(\boldsymbol{\mu}_2) & & \end{pmatrix}$$

and the testing problems:

$$\mathbf{H}_0 : \boldsymbol{\mu}_1 = \mathbf{0} \quad vs. \quad \mathbf{H}_1 : \boldsymbol{\mu}_1 \neq 0 \quad (\text{“Point-null”})$$

or

$$\mathbf{H}_0 : \boldsymbol{\mu}_1 \leq \mathbf{0} \quad vs. \quad \mathbf{H}_1 : \boldsymbol{\mu}_1 \nleq 0 \quad (\text{“One-sided”}).$$

Boundary of the null in the Gaussian Quasi-shift Model: In the point-null testing

[7] $\#A$ is defined as the cardinality of the set A.
[8] TC should be read as *topologically composite*.

problem $\text{Bd}\Theta_0$=Θ_0, the boundary of the null *is the null hypothesis itself*. In the one-sided problem, the set $\text{Bd}\Theta_0$ contains the set of parameter values $\boldsymbol{\mu}_1 \leq 0$ for which at least one of the components is equal to zero.

Assumption TC1: The parameter $\boldsymbol{\mu}_2$ is a nuisance parameter on $\text{Bd}\Theta_0$. Therefore, neither of the testing problems considered above have a set $\text{Bd}\Theta_0$ with only one element. Therefore, Assumption TC1 is verified.

Assumption TC2: The null set in both testing problems is a closed set relative to the standard topology in $\mathbb{R}^{n+n^2}$.

2.2. ***Tests, Type I/Type II Error and Risk Function***

TESTS: A *test* is a measurable mapping

$$\phi : \mathbf{X} \rightarrow [0,1],$$

where the scalar $\phi(x)$ is interpreted as the probability of rejecting $\mathbf{H}_0$ (in favor of $\mathbf{H}_1$) after a realization x of X. Therefore, a test is a summary of the decision of whether to accept or reject $\mathbf{H}_0$ for all *data sets*, x, in the sample space. Let $\mathcal{C}$ denote the class of all tests.

TYPE I AND TYPE II ERROR: Fix a test ϕ. The *rate of Type I error* of test ϕ at $\theta \in \Theta_0$ is defined as

$$\mathbb{E}_\theta[\phi(X)] = \int_{\mathbf{X}} \phi(x) f(x,\theta) dx.$$

This rate refers to the probability of rejecting the null hypothesis when the true parameter belongs to the null set. Likewise, the *rate of Type II error* of ϕ at $\theta \in \Theta_1$ is defined as

$$1 - \mathbb{E}_\theta[\phi(X)] = 1 - \int_{\mathbf{X}} \phi(x) f(x,\theta) dx.$$

When both $\mathbf{H}_0$ and $\mathbf{H}_1$ are composite, the Type I and Type II errors vary over Θ_0 and Θ_1. These variations are summarized by the risk function, defined as

$$R(\phi,\theta) \equiv \begin{cases} \mathbb{E}_\theta[\phi(X)] & \text{if} \quad \theta \in \Theta_0 \\ \\ 1 - \mathbb{E}_\theta[\phi(X)] & \text{if} \quad \theta \in \Theta_1. \end{cases}$$

Admissibility: The optimality criterion used in this paper is that of admissibility. This classical decision theoretic concept provides a natural ordering over tests based on the risk function. Let $\mathcal{C}^* \subseteq \mathcal{C}$ be a class of tests that contain ϕ. Let ϕ' be an arbitrary element of $\mathcal{C}^*$.

Definition 1: (Ferguson (1967), p. 54) The test ϕ is **admissible** within the class $\mathcal{C}^*$ if there is no $\phi' \in \mathcal{C}^*$ such that $R(\phi', \theta) \leq R(\phi, \theta)$ for all $\theta \in \mathbf{\Theta}$, with strict inequality for at least one $\theta \in \mathbf{\Theta}$.[9]

Tests that are inadmissible within a class $\mathcal{C}^*$ can be improved (that is, smaller rates of Type I and Type II error can be achieved) all over the parameter space. Thus, admissibility is a minimal requirement that a test must satisty.

Assumptions on the behavior of the Risk Function: The behavior of $R(\phi, \theta)$ is restricted by imposing a regularity assumption on the statistical models under study:

Assumption C: For any measurable set $\mathcal{F} \in \mathcal{B}(\mathbb{R}^s)_{\mathbf{X}}$, the real-valued function $P_{\mathcal{F}}(\theta) \equiv \int_{\mathcal{F}} f(x, \theta) dx$ is continuous in θ, for every $\theta \in \mathbf{\Theta}$.

Assumption C implies that for any test ϕ, $\mathbb{E}_\theta[\phi(X)]$ is a continuous function of θ. Therefore, this paper only considers problems in which the risk function of any test is continuous on both $\text{Int}(\Theta_0)$ and $\text{Int}(\Theta_1)$.[10] A sufficient condition for Assumption C is the continuity (in θ) of $f(x, \theta)$, for each $x \in \mathbf{X}$. See Lemma 5.1 in Wald (1950), p. 133; or Theorem 10 in Berger (1985), p. 545.

3. TESTING PROBLEMS WITH A BOUNDARY-SUFFICIENT STATISTIC

This paper focuses on the study of testing problems with a *boundary-sufficient statistic*. This statistical property is common to Linear IV, weakly identified GMM, Structural VARs, and some other problems with nuisance parameters; for example,

[9] Define an "ordering" over tests as a binary relation $\succ$ in the space of all tests that verifies two properties. The first one is asymmetry: $\phi \succ \phi' \implies \phi' \not\succ \phi$. The second one is transitivity: $\phi \succ \phi'$ and $\phi' \succ \phi''$ implies $\phi \succ \phi''$. Admissibility induces an ordering through the "weakly dominated" binary relation: a test ϕ' *weakly dominates* ϕ if $R(\phi', \theta) \leq R(\phi, \theta)$ with strict inequality for at least one $\theta \in \mathbf{\Theta}$.

[10] $\text{Int}(\Theta_i)$, $i = \{0, 1\}$, denotes the topological interior of the set Θ_i. That is, $\text{Int}(\Theta_i) \equiv \Theta_i \backslash \text{Bd}\Theta_i$.

the Linear Regression Model with a sign restriction in Elliott, Müller, and Watson (2012) and the predictive regression model with nearly integrated regressors studied in Stock and Watson (1996), Jansson and Moreira (2006), and Elliott et al. (2012).

This section introduces the notion of a boundary-sufficient statistic and a boundary conditional likelihood. These concepts are further illustrated using the Gaussian quasi-shift experiment.

BOUNDARY SUFFICIENCY: Boundary sufficiency is intuitively described as follows. Let $f(x_1, x_2, \theta)$ be a statistical model for the elements of the product sample space $\mathbf{X}$. The statistic X_2 is boundary sufficient if movements of θ along the boundary of Θ_0 affect the distribution of the data (X_1, X_2) *only* through its effect on X_2. Formally, this is captured by requiring the likelihood to satisfy the following decomposition.

DEFINITION 2: The statistic X_2 is ***boundary sufficient*** for the testing problem $(\mathbf{X}, \boldsymbol{\Theta}, f, \Theta_0)$ if

$$f(x_1, x_2, \theta) = g(x_1, x_2)h(x_2, \theta) \quad \text{for every} \quad \theta \in \text{Bd}\,\Theta_0,$$

where $g(\cdot, x_2)$ is a probability density function with support given by the set $\mathbf{X}_1(x_2) \equiv \{x_1 \in \mathbf{X}_1 | (x_1, x_2) \in \mathbf{X}\}$ and $\{h(x_2, \theta)\}_{\theta \in \text{Bd}\Theta_0}$ is a statistical model for the random variable X_2.

REMARK 1 Section 5 in this paper shows that in the IV model the boundary-sufficient statistic relates to the OLS estimate of the first-stage coefficient; in GMM, X_2 is a function of the derivative of the sample moment condition; in SVARs it is a function of the correlations of the reduced-form VAR errors and external instruments used to identify the structural shocks.

BOUNDARY CONDITIONAL LIKELIHOOD: In general, $g(x_1, x_2)$ corresponds to the density of the conditional distribution of X_1 given X_2, which does not depend on the element $\theta \in \text{Bd}\Theta_0$ at which the likelihood is evaluated. In light of this observation, $g(x_1, x_2)$ is denoted as $f_{\text{Bd}}(x_1 \mid x_2)$; and it is called the *boundary conditional* *likelihood.*[11]

Boundary Sufficiency in the Gaussian Quasi-Shift Model: For simplicity, consider the point-null problem $\mathbf{H}_0 : \boldsymbol{\mu}_1 = 0$. The Gaussian quasi-shift model evaluated at the boundary of null ($\boldsymbol{\mu}_\mathbf{1} = 0$) becomes

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2}\left(\begin{matrix} \mathbf{0} \\ \text{vec}(\boldsymbol{\mu}_2) \end{matrix} \quad , \quad \mathbb{I}_{n+n^2} \right).$$

Consequently, any movement along $\text{Bd}\Theta_0$—which corresponds to a change in $\boldsymbol{\mu}_2$ while keeping $\boldsymbol{\mu}_1 = 0$—affects (X_1, X_2) only through its effect in the location parameter of X_2. Hence, X_2 is a boundary-sufficient statistic. The boundary conditional likelihood is given by the density of $X_1 \sim \mathcal{N}_n(\mathbf{0}, \mathbb{I}_n)$.

The theory developed in this paper is general enough to include "point-null" and some "one-sided" hypothesis testing problems with nuisance parameters. For instance, when $\mathbf{H}_0 : \boldsymbol{\mu}_1 \leq 0$ and $n = 1$, the statistical model satisfies boundary sufficiency.

4. MAIN RESULTS

This paper provides a systematic approach to generate admissible tests within the class of *conditionally similar-on-the-boundary tests*; that is, testing procedures that control the rate of Type I error on the boundary of the null hypothesis by conditioning the accept/reject decision on the realizations of a boundary-sufficient statistic, X_2 (see Definition 3 below). The new tests derived are shown to be admissible within the class of all tests whenever X_2 is *boundedly complete* (see Definition 5 below) . The latter property is verified in IV and also in some GMM, SVARs models).

This section starts by presenting the class of *Efficient Conditionally Similar Tests* (subsequently abbreviated ECS tests), which are defined as minimizers of average risk in the class of tests that are *conditionally similar on the boundary*. The main results of this section are Theorems 1 and 2.

Theorem 1a shows that the optimization problem defining ECS tests has a solution. Theorem 1b shows that the ECS tests are admissible within the class of all procedures

[11] Note that for any $\theta \in \text{Bd}\Theta_0$:

$$f(x_1, x_2; \theta) \Big/ \int_{\mathbf{X}_1(x_2)} f(x_1, x_2; \theta) dx_1 = f(x_1, x_2; \theta)/h(x_2; \theta) = g(x_1, x_2).$$

that control Type I error by means of a boundary-sufficient statistic. Theorem 1c extends the admissibility result to the class of all tests, provided the boundary-sufficient statistic is also boundedly complete. Theorem 2 provides the basis for the implementation of ECS tests. Under certain regularity conditions, the new tests are implemented by comparing

a) The ratio of "weighted difference of integrated likelihoods" relative to the boundary conditional likelihood, against;

b) A *critical value function* that depends on the boundary-sufficient statistic.

4.1. ***ECS Tests***

Let X_2 be a boundary-sufficient statistic and let $h(x_2, \theta)$ denote the probability density function of X_2 parameterized by $\theta \in \text{Bd}\Theta_0$.

Definition 3: A test ϕ is α-***conditionally similar on the boundary of the null*** (abbreviated α-csb) if

$$\mathbb{E}_\theta[\phi(X_1, X_2) \mid X_2] = \alpha$$

for all $\theta \in \text{Bd}\Theta_0$, and for all $x_2 \in \mathbf{X}_2$ except perhaps in a set having probability zero under all distributions $\{h(x_2, \theta)\}_{\theta \in \text{Bd}\Theta_0}$.

Let $\mathcal{C}_{X_2}(\alpha\text{-csb})$ denote the class of all α-csb tests. The law of iterated expectations implies that an α-csb test is α-*similar on the boundary* (α-sb), this is:

$$\mathbb{E}_\theta[\phi(X)] = \alpha, \quad \forall\, \theta \in \text{Bd}\Theta_0.$$

Similarity and conditional similarity are classical concepts in statistical decision theory.[12] However, to the best of my knowledge, there are no general results concerning the construction of admissible similar or conditionally similar tests in the presence of a boundary-sufficient statistic.

[12]Similarity was first introduced by Neyman (1935) and it has been extensively studied by Linnik (1968). Neyman does not use the word *"similarity"* in his paper. Instead he refers to a critical region whose area is well-determined by the (composite) hypothesis to verify (*ensamble critique d'aire* 'α' *bien déterminée par l'hypothèse à verifier*). Linnik (1968) refers to such regions as α-similar regions.

Efficient Conditionally Similar Tests: Let $p_i(\theta)$ denote a full-support probability density function over $\text{Int}\Theta_i$, for $i = \{0, 1\}$ and let $\tau \in (0, 1)$.[13]

Definition 4: A test ϕ^* is α-**Efficient Conditionally Similar on the Boundary** if

$$\phi^* \in M(\tau, p_1, p_0)$$

where

(4.1.1)

$$M(\tau, p_1, p_0) \equiv \underset{\phi \in \mathcal{C}_{X_2}(\alpha\text{-csb})}{\arg\min} \; \tau \int_{\text{int } \Theta_1} R(\phi, \theta) p_1(\theta) d\theta + (1-\tau) \int_{\text{Int } \Theta_0} R(\phi, \theta) p_0(\theta) d\theta.$$

Ecs tests are built in the following way: a full-support prior on the interior of the alternative set Θ_1 is used to construct an average rate of Type II error. Likewise, a full support prior on the interior of the null set Θ_0 is used to construct an average rate of Type I error. The test that minimizes the weighted sum (with parameters τ and $1-\tau$) of average Type II and Type I errors in $\mathcal{C}_{X_2}(\alpha\text{-csb})$ is defined as an ecs test.

Note that ecs tests minimize average risk. For $\theta \notin \text{Int}\Theta_i$, set $p_i(\theta) = 0$. The function

$$p^*(\theta) \equiv \tau p_1(\theta) + (1 - \tau) p_0(\theta)$$

defines a full-support probability density function on $\text{Int } \Theta_0 \cup \text{Int } \Theta_1$. Therefore, ϕ^* is an ecs test if it minimizes average risk, that is,

$$\phi^* \in \underset{\phi \in \mathcal{C}_{X_2}(\alpha\text{-csb})}{\arg\min} \int_{\text{Int } \Theta_0 \cup \text{Int } \Theta_1} R(\phi, \theta) p^*(\theta) d\theta.$$

Priors for the Gaussian Quasi-Shift Model: Consider again the point-null problem $\mathbf{H}_0 : \boldsymbol{\mu}_1 = 0$. The interior of the alternative hypothesis is the alternative hypothesis itself. The interior of the null hypothesis is empty. Let

$$(z_1, z_2, \ldots z_{n+n^2})' \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \lambda^2 \mathbb{I}_{n+n^2}),$$

where $\lambda > 0$ is a scalar parameter used to index the priors under study. Consider

[13] A probability density function $p(\theta)$ is said to have full-support on an open set $A \subseteq \boldsymbol{\Theta}$ if $\int_a p(\theta) d\theta > 0$ for all open $a \subseteq A$, and $\int_a p(\theta) d\theta = 0$ for all $a \in \text{Int}(\boldsymbol{\Theta} \backslash A)$.

the following distribution over the parameters of the model:

$$\boldsymbol{\mu}_2 = \begin{bmatrix} z_{n+1} & z_{n+n+1} & \cdots & z_{n^2+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n+n} & z_{n+n+n} & \cdots & z_{n+n^2} \end{bmatrix}$$

and

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2^{-1}(z_1, \ldots z_n)'.$$

Note that $\boldsymbol{\mu}_2$ has the distribution of a $n \times n$ random matrix of i.i.d. normal random variables with variance λ^2. Therefore, the inverse $\boldsymbol{\mu}_2^{-1}$ exists with probability one. The prior over the parameters $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ is obtained as a transformation of the multivariate normal vector $(z_1, \ldots z_{n+n^2})$.

Remark 2 In the point-null problem, the ECS tests are simply maximizers of weighted average power subject to a conditional similarity-on-the-boundary constraint.

4.2. ***Theorem 1***

Let $(\mathbf{X}, \Theta, f, \Theta_0)$ be a hypothesis testing problem with a product sample space $(\mathbf{X}_1, \mathbf{X}_2)$. Let $\mathcal{G}$ be a collection of bounded measurable functions, $g : \mathbf{X} \to \mathbb{R}$. Theorem 1 provides a general approach to generating admissible tests within subclasses of the form:

$$\mathcal{C}(\alpha\text{-}\mathcal{G}) \equiv \{\phi \in \mathcal{C} \mid \mathbb{E}_\theta[\,(\phi(x) - \alpha)g(x)\,] = 0 \quad \forall\, \theta \in \text{Bd}\Theta_0,\; g \in \mathcal{G}\}.$$

The suggestion is as follows. First, compute the average rates of Type I/Type II errors with respect to full-support priors. Second, trade off the average rates of Type I/Type II error using a *strictly monotone function* $\mathbf{W} : \mathbb{R}^2 \to \mathbb{R}$, while imposing the constraints in $\mathcal{C}(\alpha\text{-}\mathcal{G})$.[14]

Note that ECS tests are a particular case of this approach. The set of constraints defining the class $\mathcal{C}_{X_2}(\alpha\text{-csb})$ is given by the collection of all indicator functions of the form:

$$\mathcal{G}_{X_2} = \Big\{g \mid g(x_1, x_2) = 1 \text{ if } x_2 \in \mathcal{F} \text{ and } g(x_1, x_2) = 0 \;\text{ if } x_2 \notin \mathcal{F};\; \mathcal{F} \in \mathcal{B}(X_2)\Big\}$$

[14] W(x,y) is strictly monotone if whenever $x \le x'$, $y \le y'$ (with either $x < x'$ or $y < y'$), then $W(x, y) < W(x', y')$.

where X_2 is a boundary-sufficient statistic (see Corollary 2 to Lemma 1 in Appendix A) and $\mathcal{B}(X_2)$ is the smallest σ-algebra generated by X_2. In addition, ECS tests use a linear trade-off function $\mathbf{W}(x, y) = \tau x + (1 - \tau)y$.

Theorem 1 *Let $p_i(\theta)$ denote a full-support probability density function over Int Θ_i, for $i = \{0, 1\}$, and let $\mathbf{W} : \mathbb{R}^2 \to \mathbb{R}$ be a continuous, strictly monotone function. Define*

$$M(\mathbf{W}, p_1, p_0, \mathcal{G}) \equiv \underset{\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})}{\arg\min} \mathbf{W}\Big(\int_{int\,\Theta_1} R(\phi, \theta) p_1(\theta) d\theta \,,\, \int_{Int\,\Theta_0} R(\phi, \theta) p_0(\theta) d\theta \Big).$$

T1a: *Suppose that the sample space $\mathbf{X}$ is topologically separable. Under Assumptions TC1, TC2, and Assumption C,*

$$M(\mathbf{W}, p_1, p_0, \mathcal{G}) \neq \emptyset.$$

T1b: *Suppose that $g^* \in \mathcal{G}$, where $g^*(x) = 1$ for all $x \in \mathbf{X}$. Under Assumption TC2 and Assumption C,*

$$\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}) \implies \phi^* \text{ is admissible in } \mathcal{C}(\alpha\text{-}\mathcal{G}).$$

T1c: *Suppose that $\mathcal{G}^* \equiv \{g^*\}$. Under Assumption TC1, TC2, and Assumption C,*

$$\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}^*) \implies \phi^* \text{ is admissible in } \mathcal{C}.$$

See Appendix A.2 for the proof of Theorem 1.

Remark 3 The statement of Theorem 1 is general enough to include testing problems where similarity on the boundary is of interest, but in which there is no boundary-sufficient statistic.[15] The equality $\mathcal{C}(\alpha\text{-}\mathcal{G}^*) = \mathcal{C}(\alpha\text{-sb})$ implies that such settings are covered by Theorem 1. A very interesting implication of T1a and T1c applied to $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$ is the following: *similarity-on-the-boundary is compatible with admissibility* in hypothesis testing problems satisfying Assumptions TC1, TC2, and C. That is to say, there *exists* an admissible test that is similar on the boundary. The test can be obtained as an element of $M(\mathbf{W}, p_1, p_0, \mathcal{G}^*)$.[16]

[15] For example, moment inequality models.

[16] Topological separability of $\mathbf{X}$ holds automatically in all the applications, since the sample space

REMARK 4 Theorem 1 is easily applied to ECS tests by simply noting that $\mathcal{C}(\alpha\text{-}\mathcal{G}_{X_2})$ is equal to $\mathcal{C}_{X_2}(\alpha\text{-csb})$. Theorem 1a guarantees that the ECS tests—which are average risk minimizers—are well-defined. Theorem 1b implies ECS tests are admissible in the class $\mathcal{C}_{X_2}(\alpha\text{-csb})$. The admissibility result is extended to all tests using the following condition.

DEFINITION 5: (Lehmann and Romano (2005) p. 115) Let $m : \mathbf{X}_2 \to \mathbb{R}$ be an arbitrary bounded measurable function. A boundary-sufficient statistic is boundedly complete if

$$\mathbb{E}_{h(\cdot,\theta)}[m(X_2)] = 0, \quad \forall\, \theta \in \text{Bd}\Theta_0 \implies m(X_2) = 0.$$

except, perhaps, in a set that has zero measure under every element of $\{h(\cdot,\theta)\}_{\theta\in\text{Bd}\Theta_0}$.

From Theorem 4.3.2 in Lehmann and Romano (2005) it follows that if X_2 is boundedly complete then $\mathcal{C}(\alpha\text{-}\mathcal{G}^*) = \mathcal{C}(\alpha\text{-}\mathcal{G}_{X_2})$, or equivalently, $\mathcal{C}(\alpha\text{-sb}) = \mathcal{C}(\alpha\text{-csb})$.[17] Then T1c implies that whenever the sufficient statistic X_2 is boundedly complete the ECS tests are admissible within the class of all tests.

BOUNDED COMPLETENESS IN THE GAUSSIAN QUASI-SHIFT MODEL: In Section 3 it was shown that the gaussian quasi-shift model evaluated at the boundary of null ($\boldsymbol{\mu_1} = 0$) becomes

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2}\left(\begin{matrix} \mathbf{0} \\ \text{vec}(\boldsymbol{\mu}_2) \end{matrix} \quad , \quad \mathbb{I}_{n+n^2} \right).$$

Hence, the boundary-sufficient statistic is X_2 and its distribution evaluated at the boundary is given by the statistical model

$$X_2 \sim \mathcal{N}_{n^2}(\text{vec}(\boldsymbol{\mu}_2), \mathbb{I}_{n^2}), \quad \text{vec}(\boldsymbol{\mu}_2) \in \mathbb{R}^{n^2}.$$

Theorem 4.3.1 in Lehmann and Romano (2005) provides a sufficient condition to guarantee that the family of distributions above is complete, and thus, boundedly complete. In this case, it is sufficient to show that the parameter space contains an n^2 dimensional rectangle. That is, the parameter space contains a set of the form:

$$I(a,b) = \{(\mu_{21,1}, \ldots \mu_{21,n}, \mu_{22,1}, \ldots \mu_{22,n}, \ldots \mu_{2n,n})' \in \mathbb{R}^{n^2} \mid a < \mu_{2i,j} < b, \forall\, i,j\},$$

under consideration is $\mathbb{R}^s$.

[17]See also the Lehmann and Scheffe's Theorem (Linnik (1968) Chapter 4, p. 67.

for $a, b \in \mathbb{R}, a < b$.

REMARK 5 The proof of Theorem 1 uses a novel result concerning the class of α-conditionally (or unconditionally) similar on the boundary tests: *compactness in the weak* topology* (Lemma 1, Appendix 1). Minimizers of average risk over a compact set $\mathcal{D}$ (this is, Bayes tests in $\mathcal{D}$) play an important role in *essentially complete class* theorems (see for example, Ferguson (1967), Theorem 2.10.3; Le Cam (1986), Chapter 2, Theorem 1). This is a relevant observation in light of Theorem 1a, which associates different admissible tests with different choices of $\mathbf{W}$. An important fact is that ECS tests (plus properly defined limits) form an essentially complete class in the class of α-conditionally similar-on-the-boundary tests.

4.3. *Theorem 2*

Section 4.2 presented the main theoretical properties of ECS tests. This section focuses on their implementation. ECS tests were defined as the solution to a minimization problem over a space of functions. Under boundary sufficiency and some regularity conditions, a closed form solution for this problem is available.

INTEGRATED LIKELIHOODS: Let $p_i(\theta)$ denote a full-support probability density function over $\text{Int}\Theta_i$, for $i = \{0, 1\}$ and let $\{f(x_1, x_2, \theta)\}_{\theta \in \boldsymbol{\Theta}}$ be a statistical model. Define the *null and alternative integrated likelihoods* as

$$f_i^*(x_1, x_2) \equiv \int_{\text{Int}\ \Theta_i} f(x_1, x_2, \theta) p_i(\theta) d\theta, \quad i = \{0, 1\}.$$

In cases where $\text{Int}\Theta_0 = \emptyset$ (i.e., $\Theta_0 = \text{Bd}\Theta_0$), set $f_0^*(x_1, x_2) = 0$.

The key insight of this section is the following. Fubini's Theorem implies that ϕ^* is an ECS test if and only if ϕ^* solves the problem:

$$\min_{\phi \in \mathcal{C}} \quad \tau \int_X (1 - \phi(x)) f_1^*(x) dx + (1 - \tau) \int_X \phi(x) f_0^*(x) dx$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) f_{\text{Bd}}(x_1 | x_2) dx_1 = \alpha.$$

The product structure of $\mathbf{X}$ and the linearity of the integral can be used to further simplify the objective function:

$$\max_{\phi \in \mathcal{C}} \quad \int_{X_2} \Big(\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2)[\, \tau f_1^*(x_1, x_2) - (1-\tau) f_0^*(x_1, x_2)\,] dx_1 \Big) dx_2.$$

Hence, it is possible to solve the optimization problem over the functional space by considering the following collection of problems

$$\max_{\phi \in \mathcal{C}} \int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2)[\, \tau f_1^*(x_1, x_2) - (1-\tau) f_0^*(x_1, x_2)\,] dx_1$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) f_{\mathrm{Bd}}(x_1 | x_2) dx_1 = \alpha,$$

which can be solved using the Generalized Neyman Pearson Lemma in Ferguson (1967) p. 204. The previous arguments lead to the following definitions.

ECS TEST STATISTIC: Let $\tau \in (0,1)$. Define

$$z_{\mathrm{ECS}}(x_1, x_2; p_1, p_0, \tau) \equiv \Big[\tau f_1^*(x_1, x_2) - (1-\tau) f_0^*(x_1, x_2)\Big] / f_{\mathrm{Bd}}(x_1 \mid x_2).$$

CRITICAL VALUE FUNCTION: For each $x_2 \in \mathbf{X}_2$, define

$$c(x_2; \alpha) \equiv \operatorname*{arg\,min}_{q \in \mathbf{X}_1(x_2)} \mathbb{E}_{f_{\mathrm{Bd}}(X_1 | x_2)} \Big[\rho_{(1-\alpha)} \Big(z_{\mathrm{ECS}}(X_1, x_2; p_1, p_0, \tau) - q \Big) \Big],$$

where $\rho_{1-\alpha}(\cdot)$ is the "check function" defined by $\rho_{1-\alpha}(u) = u[(1-\alpha) - \mathbb{1}\{u < 0\}]$. For each x_2, $c(x_2; \alpha)$ corresponds to the conditional (1-α) quantile of the random variable $z_{\mathrm{ECS}}(X_1, x_2; p_1, p_0, \tau)$.

REGULARITY ASSUMPTIONS: The implementation result requires regularity conditions on the integrated and boundary conditional likelihood, but also on the critical value function. A function $g : X \times Y \to \mathbb{R}$ is *separately continuous* if $g(\cdot, y) : X \to \mathbb{R}$ is continuous for all $y \in Y$ and $g(x, \cdot) : Y \to \mathbb{R}$ is continuous for all $x \in X$; see Rudin (2005) p.52.

ASSUMPTION R1: $f_i^*(x_1, x_2)$, $i = \{0, 1\}$, and $f_{\mathrm{Bd}}(x_1 | x_2)$ are *separately continuous.*

Assumption R2: The function $c(\cdot;\alpha) : \mathbf{X}_2 \to \mathbb{R}$ is measurable.

Theorem 2 *Let* $\mathbf{X}$ *be separable. Suppose Assumption R1 and R2 hold. Then* $\phi \in M(\tau, p_1, p_0)$ *if and only if* $\phi(x)$ *equals the test*

$$\phi^*(x_1, x_2) = \mathbb{1}\{z_{ecs}(x_1, x_2; p_1, p_0, \tau) - c(x_2; \alpha) > 0\},$$

except, perhaps, in a set of lebesgue-measure zero in $\mathbf{X}$*.*

See Appendix A.3 for the proof of Theorem 2.

Theorem 2 formalizes the arguments presented at the beginning of this section. In principle, the function $\phi(x_1, x_2)$ that aggregates the "accept-reject" decisions of each conditional optimization problem need not be a well-defined test (i.e., a measurable mapping from $\mathbf{X}$ to $[0,1]$). Assumptions R1 and R2 provide a set of sufficient conditions under which the measurability condition is verified.

"Point" ecs test for the Gaussian Quasi-shift model: Consider the testing problem $\mathbf{H}_0 : \boldsymbol{\mu}_1 = 0$ vs. $\mathbf{H}_1 : \boldsymbol{\mu}_1 = 0$. Just as before, let

$$(z_1, z_2, \ldots z_{n+n^2})' \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \lambda^2 \mathbb{I}_{n+n^2}),$$

where $\lambda > 0$ is a scalar parameter used to index the priors under study. I compute the integrated likelihood $f_1^*(x_1, x_2)$ under the following priors:

$$\boldsymbol{\mu}_2 = \begin{bmatrix} z_{n+1} & z_{n+n+1} & \cdots & z_{n^2+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n+n} & z_{n+n+n} & \cdots & z_{n+n^2} \end{bmatrix}$$

and

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2^{-1}(z_1, \ldots z_n)'.$$

In Section A.6 of the Appendix I show that

$$f_1^*(x_1, x_2) = c_1 \exp\Big(-\frac{x_1' x_1}{2}\Big) \exp\Big(\frac{\lambda^2}{2(1+\lambda^2)} x_1' x_1\Big) \exp\Big(\frac{-x_2' x_2}{2}\Big) \exp\Big(\frac{\lambda^2}{2(1+\lambda^2)} x_2' x_2\Big),$$

where c_1 is a non-negative constant that does not depend on (x_1, x_2). Since:

$$f_{\text{Bd}}(x_1|x_2) = c_2 \exp\Big(-\frac{1}{2}x_1'x_1\Big),$$

the ECS test statistic in Theorem 2 is given by:

$$z(x_1,x_2) \equiv \frac{c_1}{c_2}\exp\Big(\frac{1}{2}\frac{\lambda^2}{1+\lambda^2}x_1'x_1\Big)\exp\Big(-\frac{1}{2}x_2'x_2\Big)\exp\Big(\frac{1}{2}\frac{\lambda^2}{1+\lambda^2}x_2'x_2\Big).$$

Consider the critical value function

$$c(x_2;\alpha) \equiv \frac{c_1}{c_2}\exp\Big(\frac{1}{2}\frac{\lambda^2}{1+\lambda^2}\chi^2_{n,1-\alpha}\Big)\exp\Big(-\frac{1}{2}x_2'x_2\Big)\exp\Big(\frac{1}{2}\frac{\lambda^2}{1+\lambda^2}x_2'x_2\Big),$$

where $\chi^2_{n,1-\alpha}$ is the $1-\alpha$ quantile of a central χ^2_n random variable. Note that

$$\mathbb{P}_{f_{\text{Bd}}(x_1|x_2)}(z(x_1,x_2) > c(x_2;\alpha)) = \alpha.$$

Since the function $c(x_2;\alpha)$ is measurable, as it is continuous in x_2, Theorem 2 implies that the ECS test rejects if and only if:

$$x_1'x_1 > \chi^2_{n,1-\alpha},$$

regardless of the parameter λ^2.

Remark 6 In the next section I will show that the just-identified IV model with n endogenous regressors can be reduced to a Gaussian quasi-shift experiment with correlated components; that is

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2}\left(\begin{array}{c} \boldsymbol{\mu}_1\mu_2 \\ \text{vec}(\boldsymbol{\mu}_2) \end{array} , \quad W \right).$$

The parameter $\boldsymbol{\mu_1}$ corresponds to the coefficients of the n (right-hand) endogenous regressors and $\boldsymbol{\mu_2}$ is the matrix of first-stage coefficients. The random variables X_1 and X_2 are the (standarized and orthogonalized) Ordinary Least Squares estimators of the second-stage and first-stage reduced-form coefficients, respectively. I will show that the α-ECS test for the just-identified IV model rejects the null hypothesis if $x_1'x_1 > \chi^2_{n,1-\alpha}$, which corresponds to the Anderson and Rubin (1949) test.

5. EXAMPLES

This section derives ECS tests in three testing problems that are common in econometric practice: Linear Instrumental Variables regression (IV), a class of weakly identified GMM models with a scalar parameter (wGMM), and Structural Vector Autoregressions (SVARs).

Each example is presented using the following structure:

(a) Econometric model of interest and description of the testing problem;
(b) Distributional assumptions;
(c) Statistical model;
(d) Boundary sufficiency;
(e) ECS tests and the priors used to generate them.

5.1. ***Linear Instrumental Variables Regression (IV)***

This section considers three different set-ups for linear IV regression. First, I study a just-identifed IV model with non-stochastic instruments and i.i.d. normal reduced-form errors with a known covariance matrix. I show that the Anderson and Rubin (1949) test (subsequently abbreviated, AR) is ECS (RESULT 1). The priors that generate the AR have an interesting property: the implied distribution for the reduced form parameters have the same law—up to location—as their Ordinary Least-Squares (OLS) estimates.

Second, using the same distributional assumptions as above I study Chamberlain's (2007) canonical representation of an over-identified IV model with a single endogenous regressor, β. The canonical model has parameters (ρ, ϕ, ω), where ρ is a non-negative scalar measuring the strength of the instruments, ϕ is normalized to be a point on the unit circle, and ω is an element on the $(k-1)$ sphere that represents the instruments' direction. I derive a new ECS test for the point hypothesis problem, $\beta = \beta_0$, or equivalently $\phi_1 = 0$ (RESULT 2). The priors for ϕ and ω are uniform on their domain and $\rho \sim \sqrt{\lambda^2 \chi^2_k}$, where λ^2 is a free parameter for the researcher.[18] The ECS test for these priors depends on the maximal invariant in Andrews et al. (2006). The test has optimality properties that neither Moreira's (2003) CLR, nor the Wald tests based on TSLS or LIML have been shown to satisfy; namely admissibility in the class of all tests and efficiency in the class of similar tests.

[18] Chamberlain (2007) shows that the choice of prior distributions for ϕ and ω arise from a solution to a minimax problem.

Third, I study a just-identified IV model using the "local-to-zero" framework of Staiger and Stock (1997). A weak convergence result for the reduced-form OLS coefficients provides a limiting experiment—in the sense of Müller (2011)—that is convenient for the study of just-identified IV models with heteroskedastic, autocorrelated, and/or clustered data, all of which are common features in applied work. Once-again, a "robust" version of the AR test—which incorporates the asymptotic variance of the OLS coefficients—is shown to be ECS.

5.1.1. ***General Gaussian IV model***

a) ECONOMETRIC MODEL: Consider a linear IV model in reduced form matrix notation with n endogenous regressors and $k \geq n$ instruments. The notation follows the simultaneous equations framework of Moreira (2003),

$$\begin{aligned} y_1 &= Z\Pi\beta + v_1, \\ Y_2 &= Z\Pi \quad + V_2. \end{aligned}$$

The structural parameter of interest is $\beta \in \mathbb{R}^n$, while $\Pi \in \mathbb{R}^{k\times n}$ denotes the unknown matrix of first-stage coefficients. The sample size is T and the econometrician observes the data set $\{y_{1t}, Y_{2t}, Z_t\}_{t=1}^T$. I denote observations of the outcome variable, the n endogenous regressors, and the vector of instruments by y_{1t}, Y_{2t} and Z_t, respectively. The unobserved reduced form errors have realizations v_{1t} and V_{2t}. I stack the realized variables in matrices $y_1 \in \mathbb{R}^T$, $Y_2 \in \mathbb{R}^{T\times n}$, $Z \in \mathbb{R}^{T\times k}$, $v_1 \in \mathbb{R}^T$, $V_2 \in \mathbb{R}^{T\times n}$. The testing problem of interest is

$$\mathbf{H}_0 : \beta = \beta_0 \quad vs. \quad \mathbf{H}_1 : \beta \neq \beta_0.$$

b) DISTRIBUTIONAL ASSUMPTIONS: Assume that the T rows of the $T \times (n + 1)$ matrix of reduced-form errors $V = [v_1, V_2]$ are i.i.d. normally distributed with mean zero and known nonsingular covariance matrix $\Omega = [\omega_{ij}]$. This is,

$$\text{vec}(V) \sim \mathcal{N}_{T(n+1)}(\mathbf{0}, \Omega \otimes \mathbb{I}_T),$$

where " $\otimes$ " denotes the Kronecker product. For simplicity, assume that Z is nonstochastic.

c) STATISTICAL MODEL: Let $Y = [y_1, Y_2]$. Under the normality assumption for V, the sufficient statistics for the IV model are the reduced-form ordinary least-squares (OLS) estimates of $\Pi\beta$ and Π,

$$\widehat{\gamma}_{\text{OLS}} \equiv \text{vec}(\,(Z'Z)^{-1}Z'Y\,) \sim \mathcal{N}_{k+nk}\left(\begin{pmatrix} \Pi\beta \\ \text{vec}(\Pi) \end{pmatrix}, \quad \Omega \otimes (Z'Z)^{-1} \right).$$

d) BOUNDARY SUFFICIENCY: The boundary of the null hypothesis $\mathbf{H}_0 : \beta = \beta_0$ is the null hypothesis itself

$$\text{Bd}\Theta_0 = \{(\beta, \Pi) \in \mathbb{R}^n \times \mathbb{R}^{k\times n} \mid \beta = \beta_0\}.$$

In Appendix A.4, I show that the following rotation and standarization of the OLS coefficients:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2})\widehat{\gamma}_{\text{OLS}}$$

yields a statistical model in which γ_2^* is a boundary-sufficient statistic. The transformation follows Moreira (2003), p. 1030 with

$$C_0 \equiv \begin{pmatrix} (b_0'\Omega b_0)^{-1/2}b_0' \\ \\ (A_0'\Omega^{-1}A_0)^{-1/2}A_0'\Omega^{-1} \end{pmatrix},$$

$$b_0 = [1, -\beta_0']' \quad A_0 = [\beta_0, \mathbb{I}_n]'.$$

Intuitively, γ_2^* corresponds to the standarized and normalized coefficients from the first-stage regressions.

5.1.2. ***Just-identified Gaussian Model***

e1) PRIORS FOR THE JUST-IDENTIFIED MODEL AND ECS TEST: An IV model is just-identified if $k = n$. Consider the following multivariate normal vector

$$\gamma \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \Omega \otimes (Z'Z)^{-1}).$$

Write $\gamma = (\gamma_1', \gamma_{21}', \gamma_{22}', \ldots \gamma_{2n}')'$ where γ_1 is $n \times 1$ and γ_{2i} is $n \times 1$ for all $i = 1, \ldots n$. Consider the following prior distribution over the parameters (β, Π), which are natural extensions of the priors used in the Gaussian quasi-shift model in the previous section:

$$\Pi = [\gamma_{21}, \gamma_{22}, \ldots \gamma_{2n}],$$

$$\beta = [\gamma_{21}, \gamma_{22}, \ldots \gamma_{2n}]^{-1}\gamma_1.$$

Note that Π has the distribution of a Gaussian random matrix, and that Π is full rank with probability one, provided the covariance matrix $\Omega \otimes (Z'Z)^{-1}$ is nonsingular. In light of this observation, the distribution for β is well-defined. There is an interesting feature about the distribution selected: it is the unique distribution for which the reduced form coefficients $(\Pi\beta, \Pi)$ have the same random behavior—up to location—as their sample counterparts,

$$\begin{pmatrix} \Pi\beta \\ \text{vec}(\Pi) \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \text{vec}(\gamma_{21}, \gamma_{22}, \ldots \gamma_{2n}) \end{pmatrix} \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \Omega \otimes (Z'Z)^{-1}).$$

Result 1 *The α-ECS test for the problem $\boldsymbol{H}_0 : \beta = \beta_0$ vs. $\boldsymbol{H}_1 : \beta \neq \beta_0$ in a just-identified IV model with n endogenous regressors and the priors in $\boldsymbol{e}1$) rejects the null if*

$$\gamma_1^{*\prime}\gamma_1^* = (y_1 - Y_2\beta_0)'Z(Z'Z)^{-1}Z'(y_1 - Y_2\beta_0)/(b_0'\Omega b_0) > \chi^2_{n,1-\alpha}.$$

Hence, the Anderson and Rubin (1949) test is efficient conditionally similar on the boundary.

See Appendix A.5 for the proof of Result 1.

Remark 7 Note that γ_2^* is boundedly complete. This follows from the fact that the collection of distributions

$$\mathcal{N}_{n^2}\Big(\text{vec}[(Z'Z)^{1/2}\Pi\,(A_0'\Omega^{-1}A_0)^{1/2}]\,,\,\mathbb{I}_{n^2}\Big), \quad \Pi \in \mathbb{R}^{n\times n},$$

is complete. Consequently, Result 1 provides a new sense of optimality for the AR test; namely, *efficiency*: the test maximizes weighted average power—with respect to the full-support priors described above—among all similar tests. This observation complements previous results in the literature. For instance, Theorem 3 in

Moreira (2009) shows that the AR test is *uniformly most powerful* among the class of *unbiased tests* (UMPU), provided the IV model is just-identified and there is a single endogenous regressor. Result 1 applies to a larger class of tests (similar tests) and to a larger class of IV models (just-identified, arbitrary number of endogenous regressors).

REMARK 8 Remark 6, Result 1, and Theorem 1c imply that the AR test is admissible within the class of *all* tests in the context of a Gaussian just-identified model with an arbitrary number of endogenous regressors. Therefore, Result 1 complements Corollary 2 to Theorem 1 in Chernozhukov, Hansen, and Jansson (2009), which shows that the AR test is α-admissible in Gaussian over-identified models with a single endogenous regressor.[19]

5.1.3. ***Over-Identified Gaussian Linear IV model***

Chamberlain (2007) introduces a *canonical representation* of the Gaussian IV model with a single endogenous regressor ($n = 1$):

$$\begin{pmatrix} S \\ T \end{pmatrix} \equiv \begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \sim \mathcal{N}_{2k}\left(\rho(\phi \otimes \omega), \quad \mathbb{I}_{2k} \right).$$

The sample space is $\mathbb{R}^{2k}$ and the parameter space is as follows

$$\rho \in \mathbb{R}_+, \quad \phi \in \mathcal{S}^1(r(\beta_0)), \quad \omega \in \mathcal{S}^{k-1},$$

where $\mathcal{S}^m$ is the m unit sphere; that is, $\mathcal{S}^m = \{x \in \mathbb{R}^{m+1} | \, ||x|| = 1\}$, for any $m \in \mathbb{N}$ and

$$\mathcal{S}^1(r(\beta_0)) = \{(\phi_1, \phi_2) \in \mathcal{S}^1 \mid r(\beta_0)\phi_1 + \sqrt{1 - r^2(\beta_0)}\phi_2 \geq 0\},$$

with $r(\beta_0)$ equal to the correlation coefficient of the 2×2 matrix $b_0'\Omega b_0$.

The original parameters (β, Π) induce the following canonical parameters (ρ, ϕ, ω):

$$\rho = (A'\Omega^{-1}A)^{1/2}(\Pi'Z'Z\Pi)^{1/2}, \; \phi = C_0 A/(A'\Omega^{-1}A)^{1/2}, \; \omega = (Z'Z)^{1/2}\Pi/(\Pi'Z'Z\Pi)^{1/2},$$

where $A \equiv [\beta, 1]'$, and C_0 is the 2×2 matrix with first row equal to $(b_0'\Omega b_0)^{-1/2} b_0'$ and second row given by $(A_0'\Omega^{-1}A_0)^{-1/2} A_0'\Omega^{-1}$ defined in the previous section.

[19] See Lehmann and Romano (2005), p. 233 for the differences between α-admissibility and admissibility.

d) Boundary Sufficiency in the Canonical Model: The testing problem

$$\mathbf{H}_0 : \beta \leq (=)\, \beta_0 \quad vs. \quad \mathbf{H}_1 : \beta \nleq (\neq)\, \beta_0,$$

is equivalent to

$$\mathbf{H}_0 : \phi_1 \leq (=)\, 0 \quad vs. \quad \mathbf{H}_1 : \phi_1 \nleq (\neq)\, 0.$$

Therefore, on the boundary of the null hypothesis

$$\text{Bd}\Theta_0 = \{(\rho, \phi, \omega) \in \mathbb{R}_+ \times \mathcal{S}^1(r(\beta_0)) \times \mathcal{S}^{k-1} \mid \phi_1 = 0\},$$

the canonical statistical model becomes:

$$\begin{pmatrix} S \\ T \end{pmatrix} \sim \mathcal{N}_{2k}\left(\begin{pmatrix} 0 \\ \rho\omega \end{pmatrix}, \;\; \mathbb{I}_{2k} \right).$$

Hence, T is a boundary-sufficient statistic. Note that T is also boundedly complete.

e2) Priors for the over-identified IV model ($\mathbf{H}_0 : \phi_1 = 0$): Following Chamberlain (2007) the parameters (ρ, ϕ, ω) are treated as independent random variables. The distributions for ϕ and ω are uniform on their domain:

$$\phi \sim \mathcal{U}(\, \mathcal{S}^1(r(\beta_0))\,), \quad \omega \sim \mathcal{U}(\mathcal{S}^{k-1}).$$

Chamberlain (2007) shows that these choices arise from the solution of a minimax problem. I consider the following prior over the parameter ρ :

$$\rho \sim \sqrt{\lambda^2 \chi^2_k}.$$

Result 2 *The α-ECS test for the problem $\boldsymbol{H}_0 : \phi_1 = 0 \; vs. \; \boldsymbol{H}_1 : \phi_1 \neq 0$ in an over-identified IV model with a single endogenous regressor and the priors in* ***e2****) rejects the null hypothesis if the statistic:*

$$(S'S - T'T) + \frac{4(1+\lambda^2)}{\lambda^2} \ln \Big[I_0\Big(\frac{\lambda^2}{4(1+\lambda^2)} \big[(S'S - T'T)^2 + 4(S'T)^2 \big]^{1/2} \Big) \Big]$$

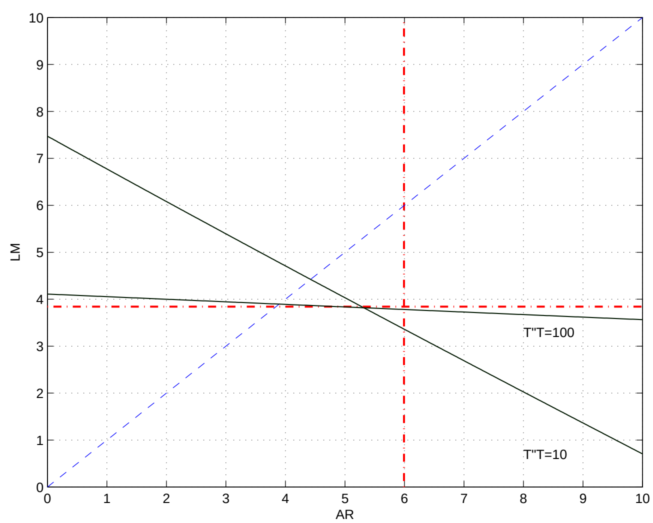
exceeds the critical value function $c^*(T; \lambda^2, \alpha)$, defined as the $(1-\alpha)$ quantile of the distribution of the statistic above with $S \sim \mathcal{N}_k(\boldsymbol{0}, \mathbb{I}_k)$ and T fixed. The function $I_0(\cdot)$

is the modified Bessel function of the first kind of order zero defined in Section 9.6, p. 375 of Abramowitz and Stegun (1964).

See Appendix A.6 for a proof of Result 2.

Remark 9 Let AR $\equiv S'S$ denote the Anderson and Rubin (1949) statistic for the over-identified IV model.[20] Let LM $\equiv (S'T)^2/T'T$ denote the Lagrange Multiplier statistic as defined in Andrews et al. (2006), p. 722. The ECS test in Result 4 is measurable with respect to the triplet (AR, LM, $T'T$). Hence, it is natural to ask whether the ECS test rejects the null hypothesis when both the AR and LM do.

Figure 1: 5% Conditional Critical Region
(AR,LM)
$k = 2$, λ^2=1



(Blue, dashed) Boundary of the sample space: AR $\geq$ LM. (Red, dot-dashed) 5% critical values for the AR and the LM statistics obtained as the upper 5% quantiles of the distributions χ^2_2 and χ^2_1, respectively.

Figure 1 reports "conditional" critical regions in the (AR, LM) space for two different values of $T'T$. The conditional critical region is the collection of (AR, LM) points at the right of the black (solid) lines (large AR and large LM). Each solid

[20] This is a slight abuse of notation as $AR = S'S/k$; see Andrews et al. (2006).

line traces the boundary of the rejection region of the ECS test for a given value of $T'T \in \{10, 100\}$.[21] The black solid line close to the LM critical value corresponds to the highest realization of $T'T$. The ECS test adjusts the χ^2_k threshold for the AR depending on the realizations of LM. Interestingly, the magnitude of the adjustment depends on the observed value of the boundary-sufficient statistic. For example, suppose LM is close to one and $T'T = 10$. The $\chi^2_{k,5\%}$ critical value for the AR is adjusted upwards and the null hypothesis is rejected only if AR$>9.7>\chi^2_{k,5\%}$. If, however, $T'T = 100$ the adjusment required is significantly larger. The "conditional" critical regions depicted in Figure 1 suggest that the ECS test in Result 2 rejects the null hypothesis whenever LM$>\chi^2_{1,5\%}$, provided $T'T$ is large.

Result 2* in Section A.7 of the Appendix presents a new test for the hypothesis $\beta \leq \beta_0$, or equivalently, for $\phi_1 \leq 0$ in the canonical model.

5.1.4. ***(Weakly) Just-Identified IV***

This section extends the results in Section 5.1.2 to weakly just-identified IV models outside the conditionally homoskedastic, serially uncorrelated framework. This generalization is relevant for practitioners working with heteroskedastic, time series, or panel data. The arguments in this section are analogous to those made in the just-identified Gaussian model. However, it is important to keep the examples in two different sections. The objective is to ilustrate the difference between a statistical model generated by a set of finite-sample distributional assumptions and one generated by a set of (asymptotic) weak convergence assumptions.

b) DISTRIBUTIONAL ASSUMPTIONS: I consider the following set of weak convergence assumptions

(1) *Weak Convergence*: $\text{vec}(\sqrt{T}(Z'Z)^{-1}Z'V) \xrightarrow{d} \mathcal{N}_{n+n^2}(\mathbf{0}, W)$, where W is a known nonsingular covariance matrix of dimension $(n+n^2) \times (n+n^2)$.
(2) *Local to zero:* $\Pi = C/\sqrt{T}$, where C is a $n \times n$ matrix.

c) STATISTICAL MODEL: The set of weak convergence assumptions induce the following limiting experiment [in the sense of Müller (2011)],

[21] The command `ezplot` in Matlab is used to graph the solution to the equation z(AR, LM, $T'T$; λ^2)-$c(T'T; \lambda^2) = 0$.

$$\sqrt{T}\widehat{\gamma}_{\text{OLS}} \equiv \text{vec}(\ \sqrt{T}(Z'Z)^{-1}Z'Y\ ) \xrightarrow{d} \mathcal{N}_{n+n^2}\left(\begin{pmatrix} C\beta \\ \text{vec}(C) \end{pmatrix},\ \ W \right).$$

Therefore, the main difference in the limiting statistical model is that the covariance matrix W need not have a kronecker structure.[22] Rotate the model by the matrix:

$$D_0 \equiv \begin{pmatrix} b_0' \\ A_0' \end{pmatrix} \otimes \mathbb{I}_n$$

so that

$$D_0\sqrt{T}\widehat{\gamma}_{\text{OLS}} \xrightarrow{d} \begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \sim \mathcal{N}_{n+n^2}\left(\begin{pmatrix} C(\beta-\beta_0) \\ \text{vec}(CA'A_0) \end{pmatrix},\ \ D_0WD_0' \right),$$

where

$$D_0\sqrt{T}\widehat{\gamma}_{\text{OLS}} = \begin{pmatrix} \sqrt{T}(\widehat{\gamma}_1 - \widehat{\gamma}_2\beta_0) \\ \text{vec}(\sqrt{T}[\widehat{\gamma}_1, \widehat{\gamma}_2]A_0) \end{pmatrix},$$

and $\widehat{\gamma}_1$, $\widehat{\gamma}_2$ are the OLS estimates of $\Pi\beta$ and Π respectively.

d) Boundary Sufficiency: The limiting experiment has two parameters: the structural parameter of interest, β, and the "drift" parameter C. The sample space is $\mathbb{R}^{n+n^2}$. In order to establish boundary sufficiency it is necessary to further transform the model. Let $[D_0WD_0']_n^{-1/2}$ represent the square root of the inverse of the first $n\times n$ block of the matrix D_0WD_0'. In section A.8 of the Appendix, I show that there is a $n+n^2$ square matrix of the form

$$D \equiv \begin{pmatrix} [D_0WD_0']_n^{-1/2} & \mathbf{0} \\ d_1 & d_2 \end{pmatrix},$$

such that $D(D_0WD_0')D' = \mathbb{I}_{n+n^2}$, where d_1 is a $n^2\times n$ matrix and d_2 is $n^2\times n^2$. Note that

$$D\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} = \begin{pmatrix} [D_0WD_0']_n^{-1/2}\gamma_1^* \\ d_1\gamma_1^* + d_2\gamma_2^* \end{pmatrix} \sim \mathcal{N}_{n+n^2}\left(D\begin{pmatrix} C(\beta-\beta_0) \\ \text{vec}(CA'A_0) \end{pmatrix},\ \ \mathbb{I}_{n+n^2} \right).$$

[22]Later, I will argue that this is an important statistical feature. For instance, the model loses invariance to certain groups of orthonormal rotations.

Therefore, whenever $\beta = \beta_0$, any change in the drift parameter C affects the distribution of $D(\gamma_1^{*\prime}, \gamma_2^{*\prime})'$ only through its effect on $d_1\gamma_1^* + d_2\gamma_2^*$, which is the boundary-sufficient statistic for the model.

e3) Priors for the (weakly) just-identified Model: There is a natural extension for the priors postulated for the Gaussian model in the previous section. Consider the following multivariate normal vector

$$\gamma \sim \mathcal{N}_{n+n^2}(\mathbf{0}, W).$$

Write $\gamma = (\gamma_1', \gamma_{21}', \gamma_{22}', \ldots \gamma_{2n}')'$ where γ_1 is $n \times 1$ and γ_{2i} is $n \times 1$ for all $i = 1, \ldots n$, and consider the following prior distribution over the parameters (β, C)

$$\Pi = [\gamma_{21}, \gamma_{22}, \ldots \gamma_{2n}],$$

$$C = [\gamma_{21}, \gamma_{22}, \ldots \gamma_{2n}]^{-1}\gamma_1.$$

Result 1* *The α-ECS test for the problem $\boldsymbol{H}_0 : \beta = \beta_0$ vs. $\boldsymbol{H}_1 : \beta \neq \beta_0$ in the limiting experiment of a weakly just-identified IV model with n endogenous regressors and the priors in **e3**) rejects the null hypothesis if*

$$\gamma_1^{*\prime}[D_0 W D_0']_n^{-1}\gamma_1^* > \chi^2_{n,1-\alpha}.$$

Hence, the test evaluated at sample analogues rejects if

$$T(\widehat{\gamma}_1 - \widehat{\gamma}_2\beta_0)'[D_0 W D_0']_n^{-1}(\widehat{\gamma}_1 - \widehat{\gamma}_2\beta_0) > \chi^2_{n,(1-\alpha)}$$

where $\widehat{\gamma}_1$ is the OLS estimate of $\Pi\beta$ and $\widehat{\gamma}_2$ is the OLS estimate of the matrix Π. The matrix W is the asymptotic variance of $\sqrt{T}\text{vec}[\widehat{\gamma}_1, \widehat{\gamma}_2]$, the matrix D_0 is defined in **c)**, and $[D_0 W D_0']_n$ is the first $n \times n$ block of the matrix $D_0 W D_0'$. The test in Result 1 is simply a robust version of the AR test as

$$\widehat{\gamma}_1 - \widehat{\gamma}_2\beta_0 = (Z'Z)^{-1}Z'(y - Y\beta_0).$$

5.2. ***Weakly Identified GMM Models***

In this section, I shall derive point ECS tests for weakly identified GMM models. The limiting experiment for this problem is based on the following observation: both the sample moment condition of a weakly identified GMM model and its derivative are asymptotically normal in large samples, provided both objects are evaluated at the boundary of the null hypothesis. The location parameter of the limiting normal distribution depends on the shape of the population moment function. Hence the limiting experiment of a weakly identified GMM model exhibits, in principle, an infinite dimensional nuisance parameter. I study problems in which the population moment function is known up to a finite-dimensional vector (as, for example, in an IV model with heteroskedastic and/or serially correlated errors).

There are two results in this section. First, I provide sufficient conditions under which the S-test of Stock and Wright (2000) is ECS. Second, I provide a general expression for ECS tests in a more general class of models. The concepts and main results in this section are ilustrated using a weakly identified IV model with non-homoskedastic and/or serially correlated errors

a) ECONOMETRIC MODEL: Let x_t be an $\mathbb{R}^d$-valued random variable. The econometrician observes the data set $\{x_t\}_{t=1}^T$, whose unknown distribution depends on a scalar parameter $\theta \in \mathbb{R}$. There is a known $\mathbb{R}^m$-valued function $h(x_t, \theta)$ that identifies the true parameter θ^* through the following moment condition:

$$\mathbb{E}_{\theta^*}[h(x_t, \theta)] = \mathbf{0} \quad \text{only at} \quad \theta = \theta^*, \, \forall \, t \quad \text{(Global Identification)}.$$

I assume the function $h(x_t, \theta)$ is almost-surely differentiable with respect to θ, with derivative $\dot{h}(x_t, \cdot) \equiv \partial h(x_t, \theta)/\partial \theta$ and that

$$\partial \mathbb{E}_{\theta^*}[h(x_t, \theta)]/\partial \theta = \mathbb{E}_{\theta^*}[\dot{h}(x_t, \theta)].$$

The testing problem of interest is

$$\mathbf{H}_0 : \theta^* = \theta_0 \quad vs. \quad \mathbf{H}_1 : \theta^* \neq \theta_0.$$

EXAMPLE (GMM-IV): Let $x_t \equiv (y_t, Y_t, Z_t)$, where y_t is the outcome variable; Y_t is a single endogenous regressor, and Z_t is a vector of $k \times 1$ instruments. Consider the function

$$h(y_t, Y_t, Z_t, \theta) = Z_t(y_t - \theta Y_t).$$

Note that

$$\mathbb{E}_{\theta^*}\left[h(y_t, Y_t, Z_t, \theta_0)\right] = (\theta^* - \theta_0)\mathbb{E}[Z_t Y_t] = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\Pi$$

and

$$\mathbb{E}_{\theta^*}[\dot{h}(x_t, \theta)] = -\mathbb{E}[Z_t Z_t']\Pi.$$

b) Distributional Assumptions: Stock and Wright (2000) developed nonstandard asymptotic theory for models defined by moment conditions when some or all of the parameters are weakly identified. I shall use their asymptotic framework to derive a limiting experiment as defined by Müller (2011). Consider the following set of (point-wise in θ^*) weak convergence assumptions for the sample moment condition and its derivative:

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(h(x_t, \theta_0) - \mathbb{E}_{\theta^*}[h(x_t, \theta_0)] \Big) \\ \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\dot{h}(x_t, \theta_0) - \mathbb{E}_{\theta^*}[\dot{h}(x_t, \theta_0)] \Big) \end{bmatrix} \xrightarrow{d} \mathcal{N}_{2m}\Big(\mathbf{0}, \Omega(\theta_0)\Big) \quad \forall\ \theta^*.$$

To model weak-identification, assume

$$\mathbb{E}_{\theta^*}[h(x_t, \theta_0)] = C_t(\theta^*, \theta_0, \delta)/\sqrt{T},$$

where C_t is known up to the finite-dimensional nuisance parameter $\delta \in \mathbb{R}^n$. The global identification assumption implies $C_t(\theta_0, \theta_0, \delta) = 0$ for all t, regardeless of the value of the nuisance parameter δ. Consider the following regularity conditions for C_t and its derivative $\dot{C}_t(\theta^*, \theta, \delta) \equiv \partial C_t(\theta^*, \theta, \delta)/\partial\theta$:

$$C(\theta^*, \theta_0, \delta) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} C_t(\theta^*, \theta_0, \delta) < \infty,$$

$$\dot{C}(\theta^*, \theta_0, \delta) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \dot{C}_t(\theta^*, \theta_0, \delta) < \infty.$$

Example (GMM-IV): Using the local-to-zero assumption of Staiger and Stock (1997),

$$\mathbb{E}_{\theta^*}\left[h(y_t, Y_t, Z_t, \theta_0)\right] = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\Pi = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\delta/\sqrt{T}$$

and

$$\mathbb{E}_{\theta^*}\left[\dot{h}(y_t, Y_t, Z_t, \theta_0)\right] = -\mathbb{E}[Z_t Z_t']\Pi = -\mathbb{E}[Z_t Z_t']\delta/\sqrt{T}.$$

Therefore,

$$C_t(\theta^*, \theta_0, \delta) = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\delta$$

and

$$\dot{C}_t(\theta^*, \theta_0, \delta) = -\mathbb{E}[Z_t Z_t']\delta.$$

Under standard regularity conditions for the second moments of Z_t,

$$C(\theta^*, \theta_0, \delta) = (\theta^* - \theta_0)\delta \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[Z_t Z_t'] < \infty,$$

$$\dot{C}(\theta^*, \theta_0, \delta) = -\delta \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[Z_t Z_t'] < \infty.$$

c) Statistical Model: The set of weak convergence assumptions and the weak identification condition yield the following limiting statistical model for the GMM problem:

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} h(x_t, \theta_0) \\ \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \dot{h}(x_t, \theta_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}_{2m}\left(\begin{pmatrix} C(\theta^*, \theta_0, \delta) \\ C'(\theta^*, \theta_0, \delta) \end{pmatrix}, \quad \Omega(\theta_0) \right) \quad \forall\ \theta^*.$$

Example (GMM-IV): The limiting experiment for the GMM-IV model with a single endogenous regressor is given by

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t(y_t - \theta_0 Y_t) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t Y_t \end{pmatrix} \xrightarrow{d} \mathcal{N}_{2k}\left(\begin{pmatrix} (\theta^* - \theta_0)Q\delta \\ Q\delta \end{pmatrix}, \quad \Omega(\theta_0) \right),$$

where

$$Q \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[Z_t Z_t'].$$

Since Q is assumed known and nonsingular in the limiting experiment it is possible to redefine $\widetilde{\delta}$ as $Q\delta$. In a slight abuse of notation $\widetilde{\delta}$ is relabeled as δ. The specific form of the matrix $\Omega(\theta_0)$ depends on primitive assumptions about the data. Suppose for simplicity that $\{y_t, Y_t, Z_t\}$ is obtained from an independent sample with heteroskedasticity. In that case

$$\Omega(\theta_0) = \begin{pmatrix} \lim_{t\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[Z_t Z_t'(y_t - \theta_0 Y_t)^2\Big] & \lim_{t\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[Z_t Z_t'(y_t - \theta_0 Y_t)v_{2t}\Big] \\ \lim_{t\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[Z_t Z_t'(y_t - \theta_0 Y_t)v_{2t}\Big] & \lim_{t\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[Z_t Z_t' v_{2t}^2\Big] \end{pmatrix}.$$

Therefore, the limiting distribution of the sample moment condition for a linear IV model is Gaussian centered at $(\theta^* - \theta_0)\delta$. The limiting distribution of the derivative of the sample moment function is also Gaussian, but centered at $-\delta$. The components are jointly normal and their dependence structure changes depending on whether the data is heteroskedastic, autocorrelated, or clustered.

d) Boundary Sufficiency: In order to establish the existence of a boundary-sufficient statistic, I will rotate and standarize the limiting experiment described above. Let $[\Omega(\theta_0)]_m$ denote the upper left $m \times m$ block of the matrix $\Omega(\theta_0)$; that is, the asymptotic variance of the sample moment condition. In section A.8 of the Appendix, I show there is a $2m$ square matrix of the form:

$$D(\theta_0) \equiv \begin{pmatrix} [\Omega(\theta_0)]_m^{-1/2} & \mathbf{0} \\ d_1 & d_2 \end{pmatrix},$$

such that $D(\theta_0)\Omega(\theta_0)D(\theta_0)' = \mathbb{I}_{2m}$, where d_1 and d_2 are $m \times m$ matrices. Therefore

$$\begin{pmatrix} m_T(\theta_0) \\ d_T(\theta_0) \end{pmatrix} \equiv D(\theta_0) \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h(x_t, \theta_0) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \dot{h}(x_t, \theta_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} m(\theta_0) \\ d(\theta_0) \end{pmatrix}$$

$$\sim \mathcal{N}_{2m} \left(\begin{pmatrix} [\Omega(\theta_0)]_m^{-1/2} C(\theta^*, \theta_0, \delta) \\ d_1 C(\theta^*, \theta_0, \delta) + d_2 \dot{C}(\theta^*, \theta_0, \delta) \end{pmatrix}, \ \mathbb{I}_{2m} \right).$$

Thus, the limiting experiment of a weakly identified GMM model has the following features. The sample space is $\mathbb{R}^{2m}$: the set of possible values for the vector $(m(\theta_0)', d(\theta_0)')$. The parameter space is $\mathbb{R}^{n+1}$: the set of possible values for the parameter of interest θ^* and the nuisance vector δ. The statistical model is a Gaussian Location problem with independent components.

The global identification assumption implies that whenever $\theta^* = \theta_0$

$$\begin{pmatrix} m(\theta_0) \\ d(\theta_0) \end{pmatrix} \sim \mathcal{N}_{2m} \left(\begin{pmatrix} \mathbf{0} \\ d_2 \dot{C}(\theta^*, \theta_0, \delta) \end{pmatrix}, \ \mathbb{I}_{2m} \right).$$

Thus, $d(\theta_0)$ is a boundary sufficient statistic in the limiting experiment of the weakly identified GMM model.

e) PRIORS FOR THE WEAKLY IDENTIFIED GMM MODEL: First, I will provide sufficient conditions under which the S-test of Stock and Wright (2000)—based on the *continuously updated* GMM objective function—is ECS. For a fixed θ_0, consider the mapping $C^* : \mathbb{R}^{n+1} \to \mathbb{R}^{2m}$ given by

$$C^*(\theta^*, \delta) = \begin{pmatrix} C(\theta^*, \theta_0, \delta) \\ \dot{C}(\theta^*, \theta_0, \delta) \end{pmatrix}.$$

ASSUMPTION RGMM1: C^* is a continuous function.

EXAMPLE (GMM-IV): In the GMM-IV model, the dimension of the nuisance parameter (n) equals the number of instruments (k). Likewise, the dimension of the moment conditions (m) equals k. The mapping $C^* : \mathbb{R}^{k+1} \to \mathbb{R}^{2k}$ is given by

$$\begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}.$$

The mapping C^* is continuous in (θ^*, δ). Note that C^* is crucial for comparing the power performance of different testing procedures. The following result states sufficient conditions under which the S-test of Stock and Wright (2000) is ECS.

Result 3 *Let $n+1 \geq 2m$ and let assumption* RGMM1 *hold. Suppose that there is a full-support prior p_1 over $\mathbb{R}^{n+1}$ such that:*

$$C^*(\theta^*, \delta) \sim \mathcal{N}_{2m}(\boldsymbol{0}, \Omega(\theta_0)).$$

*Then the α-*ECS *test for the problem $\boldsymbol{H}_0 : \theta^* = \theta_0$ vs. $\boldsymbol{H}_1 : \theta^* \neq \theta_0$ in the limiting experiment of a weakly identified GMM model rejects the null hypothesis if*

$$m(\theta_0)' m(\theta_0) > \chi^2_{m,1-\alpha}.$$

See Appendix A.9 for a proof of Result 3.

The test, evaluated at sample analogues, coincides with the S-test of Stock and Wright (2000):

$$\Big(\frac{1}{T}\sum_{t=1}^{T} h(x_t, \theta_0)\Big)' [\Omega(\theta_0)]_m^{-1} \Big(\frac{1}{T}\sum_{t=1}^{T} h(x_t, \theta_0)\Big) > \chi^2_{m,1-\alpha},$$

where $[\Omega(\theta_0)]_m$ is the asymptotic variance of

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h(x_t, \theta_0),$$

and, to simplify notation, I have assumed that such a covariance matrix is known.[23]

Example (GMM-IV): The condition of Result 3 is simple to verify in the GMM-IV model. Note that $k+1 \geq 2k$ if and only if $k=1$. In this case, the S-test is ECS and rejects if:

---

[23] If this were not the case, one could replace $\Omega(\theta_0)$ with an estimator $\widehat{\Omega}_T(\theta_0)$. The (Gaussian) weak convergence assumption combined with $\widehat{\Omega}_T(\theta_0) \xrightarrow{p} \Omega(\theta_0)$ yield the same limiting experiment.

$$\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t(y_t-\theta_0 Y_t)\Big)'\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'(y_t-\theta_0 Y_t)^2\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t(y_t-\theta_0 Y_t)\Big) > \chi^2_{1,1-\alpha},$$

which coincides with the robust version of the AR test derived in section 5.1. When $k > 1$, there are no priors over (θ^*, δ) for which the function $C^*(\theta^*, \delta)$ behaves as a Gaussian distribution on $\mathbb{R}^{2k}$. This is simply because C^* takes values on a strict subset of $\mathbb{R}^{2k}$.

Now, I will provide a general expression for ECS tests in weakly identified GMM models. Let

$$\gamma(\theta_0)' = [m(\theta_0)', d(\theta_0)'].$$

Result 4 *Let p_1 be a full-support prior over $\mathbb{R}^{n+1}$ and suppose that assumptions* RGMM1 *and R2 hold. Then the α-*ECS *test statistic for the problem $\boldsymbol{H}_0 : \theta^* = \theta_0$ vs. $\boldsymbol{H}_1 : \theta^* \neq \theta_0$ rejects the null hypothesis if*

$$\int_{\mathbb{R}^{n+1}} \exp\Big(\gamma(\theta_0)' D(\theta_0) C^*(\theta^*, \delta)\Big) \exp\Big(-\frac{1}{2} C^*(\theta^*, \delta)' \Omega(\theta_0)^{-1} C^*(\theta^*, \delta)\Big) p_1(\theta^*, \delta) d\theta^* d\delta,$$

is larger than its $1-\alpha$ quantile, conditional on $d(\theta_0)$.

See Appendix A.10 for a proof of Result 4.

Example (GMM-IV): The sample analogue of the test statistic in Result 4 is given by

$$\begin{aligned} z(\widehat{m}(\theta_0), \widehat{d}(\theta_0)) \quad = \quad & \int_{\mathbb{R}^{k+1}} \Big[\exp\Big(\begin{pmatrix} \widehat{m}(\theta_0) \\ \widehat{d}(\theta_0) \end{pmatrix}' \widehat{D}(\theta_0) \begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}\Big) \\ & \exp\Big(-\frac{1}{2}\begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}' \widehat{\Omega}(\theta_0)^{-1} \begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}\Big) p_1(\theta^*, \delta)\Big] d\theta^* d\delta, \end{aligned}$$

where

$$\widehat{D}(\theta_0) = \begin{pmatrix} \Big[\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'(y_t - \theta_0 Y_t)^2\Big]^{-1/2} & \mathbf{0} \\ \widehat{d}_1 & \widehat{d}_2 \end{pmatrix}, \quad \widehat{\Omega}(\theta_0)^{-1} = \widehat{D}(\theta_0)' \widehat{D}(\theta_0),$$

and $\widehat{d}_1, \widehat{d}_2$ are the sample analogues of the matrix defined in Lemma 3 in Appendix A applied to $\Omega(\theta_0)$. The boundary sufficient statistic is given by:

$$\widehat{d}(\theta_0) = \widehat{d}_1 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t(y_t - \theta_0 Y_t) + \widehat{d}_2 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t Y_t,$$

and

$$\widehat{m}(\theta_0) = \Big(\frac{1}{T} \sum_{t=1}^{T} Z_t Z_t' (y_t - \theta_0 Y_t)^2 \Big)^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t (y_t - \theta_0 Y_t).$$

The critical value $c(\widehat{d}(\theta_0))$ is obtained by fixing $\widehat{d}(\theta_0)$ and computing the 1-α quantile of the random variable:

$$z(m, \widehat{d}(\theta_0)), \quad m \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k).$$

The ECS test rejects the null hypothesis if $z(\widehat{m}(\theta_0), \widehat{d}(\theta_0)) > c(\widehat{d}(\theta_0))$.

5.3. ***Dynamic Effects in Structural VARs***

Montiel, Stock, and Watson (2012) develop methods for inference in structural vector autoregressions (SVARs) in which the structural shocks are identified using external instruments. They focus on the possibility of potentially weak instruments. This section shows that the test used by Montiel, Stock, and Watson (2012) to build confidence intervals for dynamic effects in "just-identified" SVARs is ECS. The test rejects for large values of the sample covariance between the instrument that identifies the target structural shock and a linear combination of all the reduced-form shocks in the vector autoregression. This section also presents an ECS test for the over-identified SVAR model.

a) ECONOMETRIC MODEL: Let the $r \times 1$ time series Y_t follow the reduced-form stationary VAR with p lags:

$$Y_t = A(L) Y_{t-1} + \eta_t,$$

where $A(L)$ is a known lag-polynomial that is assumed invertible. The $r \times 1$ vector

η_t represents the reduced form innovations.[24] There is a $r \times 1$ vector of structural shocks ε_t that satisfy:

$$\eta_t = H\varepsilon_t = [H_1, H_2, \dots H_t]\epsilon_t.$$

The unknown $r \times r$ matrix H is assumed invertible, H_i denotes the i-th column of H and h_{im} denotes the m-th element of H_i. The structural moving average representation of the reduced-form VAR is given by:

$$Y_t = A(L)^{-1}H\varepsilon_t.$$

Let $C'_{hj} = (c_{hj1}, c_{hj2}, \dots c_{hjr})$ denote the j-th row of the h-lag matrix of $A(L)^{-1}$. The object of interest is the dynamic effect of a shock ε_{1t} over variable j at horizon h. The null hypothesis states that an impulse in the structural shock ϵ_{1t} will have an effect of κ_0 over the j-th component of Y_{t+h}; that is

$$\mathbf{H}_0 : C'_{hj}H_1 = \kappa_0 \quad vs. \quad \mathbf{H}_1 : C'_{hj}H_1 \neq \kappa_0.$$

b) Distributional Assumptions: The $k \times 1$ vector of external instruments Z_t is used to identify the dynamic effect with respect to the structural shock of interest through the moment condition

$$\mathbb{E}[\varepsilon_t \otimes Z_t] = e_1 \otimes \alpha, \quad e_1 \in \mathbb{R}^r,\ e_1 = (1, 0, 0, \dots 0)';\ \alpha \in \mathbb{R}^k,$$

which implies that

$$\mathbb{E}[\eta_t \otimes Z_t] = \mathbb{E}[H\varepsilon_t \otimes Z_t] = (H \otimes I_k)\mathbb{E}[\varepsilon_t \otimes Z_t] = H_1 \otimes \alpha.^{25}$$

In order to "normalize" the effect of interest, it is assumed that $h_{11} = 1$.[26] Consequently, the moment condition above identifies H_1 and the parameter of interest, $C'_{hj}H_1$.

To model the potential weak correlation between the instruments (Z_t) and the structural shock (ε_{1t}), assume that $\alpha = a/\sqrt{T}$.

[24]Assuming $A(L)$ is known entails no loss of generality; see Montiel, Stock and Watson (2012) for details.

[25]Note that $\{\eta_t\}$ is observed as $A(L)$ is assumed known.

[26]Therefore, the structural shock h_{11} is measured in units of the variable y_{1t} in the reduced form VAR.

c) STATISTICAL MODEL: The data generating process for $\{\eta_t, Z_t\}_{t=1}^{T}$ is restricted by imposing the following weak convergence assumptions

$$(1/\sqrt{T})\sum_{t=1}^{T}\eta_t \otimes Z_t = \begin{pmatrix} (1/\sqrt{T})\sum_{t=1}^{T}\eta_{1t}Z_t \\ (1/\sqrt{T})\sum_{t=1}^{T}\eta_{\bullet t}\otimes Z_t \end{pmatrix} \xrightarrow{d} \gamma \sim \mathcal{N}_{rk}\Big(H_1 \otimes a, \Omega\Big)$$

where

$$\eta_{\bullet t} = \begin{pmatrix} \eta_{2t} \\ \vdots \\ \eta_{rt} \end{pmatrix}.$$

d) BOUNDARY SUFFICIENCY: Consider first the just-indentified case ($k = 1$). Let $\kappa \equiv C'_{hj}H_1$ and let κ_0 denote the null hypothesis of interest. Define

$$C'_0 = (c_{hj1} - \kappa_0, c_{hj2}, \ldots c_{hjr}),$$

and let $C_0^{\perp\prime}$ denote the *orthonormal part* of C'_0; that is, $C_0^{\perp\prime}$ is the $(r-1)\times r$ matrix such that

(i) The $r \times r$ matrix $\begin{pmatrix} C'_0 \\ C_0^{\perp\prime} \end{pmatrix}$ has full rank.

(ii) $C_0{}'C_0^{\perp} = \mathbf{0}$.

(iii) For each $m = 1 \ldots r-1$, $C_{0m}^{\perp}{}'C_{0m}^{\perp} = 1$; where $C_{0m}^{\perp}{}'$ denotes the m-th row of $C_0^{\perp\prime}$.

The standarized limiting experiment for the SVARs testing problem is given by

$$\begin{pmatrix} S \\ T_1 \end{pmatrix} \equiv \begin{pmatrix} C'_0\gamma/(C'_0\Omega C_0)^{1/2} \\ \Big(C_0^{\perp\prime}\Omega^{-1}C_0^{\perp}\Big)^{-1/2} C_0^{\perp\prime}\Omega^{-1}\gamma \end{pmatrix}$$

$$\sim \mathcal{N}_r\left(\begin{pmatrix} (\kappa-\kappa_0)a \\ \Big(C_0^{\perp\prime}\Omega^{-1}C_0^{\perp}\Big)^{-1/2} C_0^{\perp\prime}\Omega^{-1}H_1 a \end{pmatrix}, \mathbb{I}_r \right).$$

Under the null hypothesis $\kappa - \kappa_0 = 0$. Therefore, T_1 is a boundedly complete boundary-sufficient statistic, whenever $k = 1$.

e1) PRIORS FOR JUST-IDENTIFIED SVARS: Let

$$\begin{pmatrix} m_1 \\ \vdots \\ m_r \end{pmatrix} \sim \mathcal{N}_r(\mathbf{0}, \Omega),$$

and consider the following distribution over the parameters (H_1, a) of the statistical model above

$$a = m_1, \quad h_{1n} = m_n/m_1,\ n = 2, \dots, r.$$

RESULT 5 *The α-ECS test for the problem $\boldsymbol{H}_0 : \kappa = \kappa_0$ vs. $\boldsymbol{H}_1 : \kappa \neq \kappa_0$ in a just-identified SVAR model with priors in **e1)** rejects the null hypothesis if:*

$$S'S > \chi^2_{1,1-\alpha}.$$

See Appendix A.11 for a proof of Result 5.

The sample analogue of the ECS test statistic above is

$$\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (C_0'\eta_t) Z_t\Big)^2 / C_0'\Omega C_0,$$

which is equivalent to the "robust" AR statistic (Result 1*) from the following just-identified IV model:

a) Outcome variable: $C_{hj}'\eta_t$
b) Endogenous variable: η_{1t} (with coefficient κ)
c) Instrument: Z_t
d) Null hypothesis: $\kappa = \kappa_0$

REMARK 10 C_0 and η_t have been assumed to be known sample objects.

e2) PRIORS FOR THE OVER-IDENTIFIED MODEL WITH A "KRONECKER" COVARIANCE MATRIX: Suppose $\Omega = \Sigma \otimes Q = \mathbb{E}[\eta_t \eta_t'] \otimes \mathbb{E}[Z_t Z_t']$. In Appendix A.11 it is shown that the limiting experiment of the SVARs testing problem with a kronecker covariance matrix admits the following representation

$$\begin{pmatrix} S \\ T_1 \\ \vdots \\ T_{r-1} \end{pmatrix} \sim \mathcal{N}_{rk}\Big(\phi \otimes \rho\omega \ , \ \mathbb{I}_r \otimes \mathbb{I}_k\Big).$$

The sample space is R^{rk} with a typical element denoted by $(S', T_1', T_2' \dots T_{r-1}')'$. The parameter space is given by $\mathbb{R}_+ \times \mathcal{S}_R^{r-1} \times \mathcal{S}^{k-1}$, with typical element (ρ, ϕ, ω).[27] The hypothesis $\kappa = \kappa_0$ is equivalent to:

$$\mathbf{H}_0 : \phi_1 = 0 \quad vs. \quad \mathbf{H}_1 : \phi_1 \neq 0.$$

Hence, $T \equiv (T_1', T_2', \dots T_{r-1}')'$ is a boundary-sufficient statistic. Consider the independent priors

$$\rho \sim \sqrt{\lambda^2 \chi_k^2}, \quad \omega \sim \mathcal{U}(\mathcal{S}^k), \quad \phi \sim \mathcal{U}(\mathcal{S}_R^{r-1})$$

Let $M \equiv [S, T_1, \dots T_{r-1}]'[S, T_1, \dots T_{r-1}]$.

Result 6 *The* ECS *test for the problem* $\boldsymbol{H}_0 : \phi_1 = 0$ *vs.* $\boldsymbol{H}_1 : \phi_1 \neq 0$ *in the over-identified SVAR model with priors in* ***e2)*** *rejects the null hypothesis if the statistic*

$$\int_{\mathcal{S}_R^{r-1}} \exp\Big(\frac{\lambda^2}{2(1+\lambda^2)}\phi' M \phi\Big) d\lambda_{\mathcal{S}^{r-1}}(d\phi)$$

is larger than its $1-\alpha$ *quantile, conditional on* T*.*

See Appendix A.12 for a proof of Result 6.

The sample analogues of S and T are given by

$$\widehat{S} \equiv (1/\sqrt{T}) \sum_{t=1}^{T} \Big(C_0' \eta_t / (C_0' \widehat{\Sigma} C_0')^{1/2} \otimes \widehat{Q}^{-1/2} Z_t\Big),$$

and

$$\widehat{T} = (1/\sqrt{T}) \sum_{t=1}^{T} \Big(\Big(C_0^{\perp\prime} \widehat{\Sigma}^{-1} C_0^{\perp}\Big)^{-1/2} C_0^{\perp\prime} \widehat{\Sigma}^{-1} \eta_t \otimes \widehat{Q}^{-1/2} Z_t\Big).$$

[27] $\mathcal{S}_R^{r-1} = \{\phi \in \mathcal{S}^{r-1} \mid e_1 \Sigma A' \phi \geq 0\}$.

6. SUMMARY OF THE MAIN RESULTS AND CONCLUSION

Boundary sufficiency arises naturally in three widely used models in econometrics: Linear Instrumental Variables Regression (IV), the Generalized Method of Moments (GMM), and Structural Vector Autoregressions (SVARs). This property is common to other hypothesis testing problems with nuisance parameters; for example, the Linear Regression Model with a sign restriction in Elliott et al. (2012); the predictive regression model with nearly integrated regressors studied in Stock and Watson (1996), Jansson and Moreira (2006), and Elliott et al. (2012); and testing problems in exponential family models. Boundary sufficiency is an attractive feature, for it allows the econometrician to control the rate of Type I error—which can dramatically vary in standard Wald tests—regardless of the values of nuisance parameters.

This paper used statistical decision theory as a guiding principle to derive a new class of tests for hypothesis testing problems with a boundary-sufficient statistic. The tests are *efficient*, for they minimize a weighted sum of the average rates of Type I and Type II error (average risk). The tests are *conditionally similar on the boundary*, for they control the rate of Type I error on the set of null parameter values that are the closest to the alternative set by conditioning on the realizations of the boundary-sufficient statistic.

This paper showed that *Efficient Conditionally Similar-on-the-boundary* (ECS) tests are *admissible* within the class of conditionally similar-on-the-boundary procedures. Moreover, ECS tests were shown to verify an important finite-sample optimality property: admissibility within the class of *all* tests, provided the boundary-sufficient statistic is boundedly complete.

Theorem 1 in this paper provided a systematic method to derive admissible tests within the class of conditionally and unconditionally similar-on-the-boundary tests. The idea is conceptually simple: it suffices to trade off the average rates of Type I and Type II error using a monotone continuous function $\mathbf{W} : \mathbb{R}^2 \to \mathbb{R}$. When $\mathbf{W}$ is linear, the exercise is equivalent to average risk minimization (using full-support priors) subject to a conditional or unconditional similarity-on-the-boundary constraint. The solution to this problem is well-defined, for the domain of the optimization problem is *weak* compact* and the objective function is *weak* continuous*. ECS tests are thus defined as the solution to a minimization problem over a space of functions.

Theorem 2 in this paper showed that the minimization problem defining ECS tests has a convenient closed form solution: the ECS test statistic can be expressed as a linear combination of the null and alternative integrated likelihoods. The critical value *function* is given by the conditional quantiles of the ECS test statistic. Theorem 1 and 2 complement recent findings by Moreira and Moreira (2012), which develop methods to approximate the solution of risk minimization problems over the space of similar tests, without requiring a boundary-sufficient statistic.

This paper applied the theory of ECS tests to hypothesis testing problems in IV, GMM, and SVARs. The emphasis was on "point" problems, in which case the ECS tests maximize weighted average power (WAP) subject to a conditional similarity constraint on the null set.

Result 1 showed that the Anderson and Rubin (1949) (AR) test is ECS in just-identified IV models with Gaussian reduced-form errors, independent observations, fixed instruments, and an arbitrary number of endogenous regressors. Result 1* extended this result to models with heteroskedastic, autocorrelated, and/or clustered data. The priors over the structural parameters of the IV model (β and Π) for which the AR test maximizes WAP were shown to have an interesting feature: there are no other priors for which the implied distribution over the reduced-form parameters ($\Pi\beta$ and Π) is Gaussian, centered at zero, and with the same covariance matrix as the distribution of their sample counterparts. Since the boundary-sufficient statistic in the just-identified IV model is boundedly complete (regardless of the number of endogenous regressors), the AR is admissible in the class of all tests.

Result 2 derived a novel test for point-null hypotheses in the over-identified IV model studied by Andrews et al. (2006) and Chamberlain (2007). The new ECS test enjoys basic optimality properties that neither CLR nor the TSLS (LIML) Wald tests have been shown to satisfy: namely, admissibility in the class of all tests and efficiency in the class of similar tests. The "conditional" critical region of the new test can be expressed in terms of the AR and the Lagrange Multiplier (LM) statistics. If the LM is below (above) its conventional χ^2_1 critical value, the ECS test automatically adjusts upwards (downwards) the χ^2_k threshold for the AR. The magnitude of the adjustment depends on the value of the boundary-sufficient statistic and the ECS test rejects the null hypothesis whenever the AR exceeds the adjusted critical value. Result 2* derived a new test for one-sided problems.

Result 3 showed that the S-test of Stock and Wright (2000) is ECS in some weakly identified GMM models in which the parameter of interest is scalar and the population moment function is known up to a finite-dimensional vector. To the best of my knowledge, this is the first optimality result derived for weakly identified GMM models. A key component of Result 3 is the theory of Müller (2011), which motivates the study of a statistical model derived from a set of weak convergence assumptions—as opposed to assumptions about the finite distribution of the data.

Result 4 provided a general expression for ECS tests in GMM models. An important observation is that the boundary-sufficient statistic has the same dimension as the derivative of the sample moment condition. In general, it is not clear whether a *dimension reduction* for the boundary-sufficient statistic is available—as it is assumed by Kleibergen (2007), whose tests condition on a scalar statistic containing information about the rank of the matrix of derivatives.

Finally, Result 5 showed that the test used by Montiel, Stock, and Watson (2012) to build confidence intervals for dynamic effects in "just-identified" SVARs is ECS. The test rejects for large values of the sample covariance between the instrument that identifies the target structural shock in the model and a linear combination of all the reduced-form shocks in the vector autoregression. Result 6 derived ECS tests for the over-identified SVAR model.

There are two observations that highlight the importance of *Efficient Conditionally Similar-on-the-boundary* (ECS) tests.

First, efficient tests and, more generally, admissible tests need not be (conditionally) similar. For instance, standard Bayes tests—that is, tests that reject for large values of a ratio of integrated likelihoods—are, by construction, efficient and thus admissible. Despite their admissibility, Bayes tests face an important limitation: their rate of Type I error can vary over the null set and, in some cases, such rate can be arbitrarily close to one regardless of the critical value used in their implementation.

Second, there are already tests in the literature that use a boundary-sufficient statistic to control the rate of Type I error. For example, Moreira (2003) proposed the Conditional Likelihood Ratio test (CLR) for IV and Kleibergen (2007) extended the CLR to GMM problems. These procedures are, by construction, conditionally similar on the boundary. However, as far as I know, they are neither admissible nor efficient, even in some restricted sense. In fact, it is not clear whether the CLR and its extensions are admissible within the class of conditionally similar tests. Without an analytical claim for admissibility, there is no guarantee that these procedures cannot be improved.

Concluding Remark: A continued focus of the econometrics literature in the past two decades has been the *finite-sample* analysis of widely used estimation and testing procedures. Studying the performance of any statistical decision rule (e.g., estimator, test, or confidence interval) in a finite sample requires—in one way or another—that *there be a statistical model*, which inevitably connects the properties of the decision rules under consideration with *statistical decision theory.* It is then possible to use classical concepts—for example, that of a risk function—to think about *optimal* selection of estimators, tests, or confidence intervals. This paper followed this approach. The decision problem of interest was hypothesis testing in IV, GMM, and SVARs. The main optimality concept was that of finite-sample admissibility. There was an additional constraint motivated by applied work: invariance of the rates of Type I error with respect to nuisance parameters—i.e., similarity—in some region of the null set. This paper identified a common statistical property in the three problems under consideration: boundary sufficiency. This property was used to derive a new class of tests and to establish a new sense of efficiency for some existing procedures.

APPENDIX A

A.1. *Proof of Lemma 1*

This lemma is used to show that the class of α-conditionally similar-on-the-boundary tests and the class of α-similar-on-the-boundary-tests is compact, relative to the space of essentially bounded measurable functions endowed with the weak* topology (Lemma 1, below).

Preliminaries 1 (L^1 and L^∞): Since the sample space $\mathbf{X} \in \mathcal{B}(\mathbb{R}^s)$, the triplet $(\mathbf{X}, \mathcal{B}(\mathbb{R}^s)_{\mathbf{X}}, \lambda^s)$ is a well-defined σ-finite measure space. Note that $\mathcal{B}(\mathbb{R}^s)_{\mathbf{X}} = \mathcal{B}(\mathbf{X})$ whenever $\mathbf{X}$ is endowed with the sub-space topology relative to $\mathbb{R}^s$. Following Rudin (2006), p. 65, let $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ denote the space of all real-valued measurable functions f that satisfy $||f||_1 \equiv \int_{\mathbf{X}} |f(x)|dx < \infty$. Let $L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ denote the class of all essentially bounded real-valued measurable functions (Rudin (2006) p. 66).

Remark 11 Identify the class of all tests $\mathcal{C}$ as a subset of $L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$

$$\mathcal{C} \equiv \{\phi \in L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s) \mid \phi(x) \in [0,1] \text{ for } \lambda^s\text{-a.e. } x \in \mathbf{X}\}.$$

And note that the elements of any statistical model $\{f(x,\theta)\}_{\theta \in \mathbf{\Theta}}$ are elements of $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$, by the definition of probability density function $\int_{\mathbf{X}} f(x,\theta)dx = 1 < \infty$ for all $\theta \in \mathbf{\Theta}$.

Preliminaries 2 (The dual space of L^1): Let $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$ denote the dual space of $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$, i.e., the space of all continuous (w.r.t. $||f||_1$) linear functionals on $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$; see Rudin (2005), p. 56. Let Λ denote an element of the dual space $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$. By Theorem 6.16 in Rudin (2006), p. 127 and Theorem 1.18 in Rudin (2005), p. 15; the space $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$ is isometrically isomorphic to $L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$. Therefore, one can identify each functional Λ with a unique element (up to equivalence) $g \in L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$, and vice versa: for $f \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)^*$, $\Lambda \in [L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$ is of the form

$$\Lambda(f) \equiv \int_{\mathbf{X}} g(x)f(x)dx \quad \text{for some} \quad g \in L^{\infty}(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s).$$

Preliminaries 3 (weak* topology on L^∞): Endow the space $L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ with the topology induced by the weak*-topology on the space $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$; see Rudin (2005), p. 67, 68. The new topological space is denoted by $(L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s), \mathcal{T}^*)$. Denote convergence in such topology by $\rightarrow^*$. Note that, by definition, $\{g_n\}_{n\in\mathbb{N}} \rightarrow^* g$ if and only if

$$\int_{\mathbf{X}} f(x)g_n(x)dx \rightarrow \int_{\mathbf{X}} f(x)g(x)dx \quad \forall \quad f \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s).$$

Let $(\mathbf{X}, \boldsymbol{\Theta}, f, \Theta_0)$ be a hypothesis testing problem. Let $\mathcal{G} \subset L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ be an arbitrary collection of bounded functions. Define

$$\mathcal{C}(\alpha\text{-}\mathcal{G}) \equiv \{\phi \in \mathcal{C} \mid \quad \mathbb{E}_\theta[(\phi(X) - \alpha)g(X)] = 0 \quad \forall \theta \in \mathrm{Bd}\Theta \quad \forall \, \mathrm{g} \in \mathcal{G}\}$$

Let $(L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s), \mathcal{T}^*)$ be the space of essentially bounded functions topologized with the weak* topology. For any $A \subset L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X})$, let $\mathcal{T}^*_A$ denote the subset topology induced by $\mathcal{T}^*$

Lemma 1: The set $\mathcal{C}(\alpha\text{-}\mathcal{G})$ is compact relative to $(\mathcal{C}, \mathcal{T}^*_{\mathcal{C}})$.

Proof: The outline of the proof is the following. I show that the set $\mathcal{C}(\alpha\text{-}\mathcal{G})$ is a sequentially closed subset of $\mathcal{C}$ with the relative weak* topology. Then I use the Banach-Alaoglu theorem and the topological separability of $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ to establish the compacteness of $\mathcal{C}(\alpha\text{-}\mathcal{G})$.

(*Sequential Closedness*) Take any convergent sequence of tests $\phi_n \rightarrow^* \phi$ with $\{\phi_n\}_{n\in\mathbb{N}} \subseteq \mathcal{C}(\alpha\text{-}\mathcal{G})$. I want to show that $\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})$. First, I show that $\phi(x) \in \mathcal{C}$; i.e., $\phi \in [0, 1]$ for almost every $x \in \mathbf{X}$. Suppose not. Then there exists a measurable set $A \in \mathcal{B}(\mathbf{X})$ with $\lambda^s(A) > 0$ such that $\phi(x) > 1$ or $\phi(x) < 0$ for all $x \in A$. Without loss of generality assume $\phi(x) > 1$. Since λ^s is σ-finite, there exists a countable collection $\{E_n\}_{n\in\mathbb{N}}$ such that $\cup_{n\in\mathbb{N}} E_n = \mathbf{X}$ and $\lambda^s(E_n) < \infty$ for every n. Consider the sequence of sets $\{A \cap E_n\}_{n\in\mathbb{N}}$. Note that $0 \leq \lambda^s(A \cap E_n) < \infty$ for all $n \in \mathbb{N}$. In addition, there exists $N \in \mathbb{N}$ for which $0 < \lambda^s(A \cap E_N)$, otherwise $\lambda^s(A) = \lambda^s(\cup_{n=1}^\infty (A \cap E_n)) \leq \sum_{n=1}^\infty \lambda^s(A \cap E_n) = 0$. Consider the indicator function $\mathbb{1}_{A\cap E_N}$. Since $0 < \lambda^s(A \cap E_N) < \infty$, the indicator function $\mathbb{1}_{A\cap E_N} \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$. Note that

$$\lambda^s(A \cap E_N) < \int_{\mathbf{X}} \mathbb{1}_{A\cap E_N}(x)\phi(x)dx = \lim_{n\rightarrow\infty} \int_{\mathbf{X}} \mathbb{1}_{A\cap E_N}(x)\phi_n(x)dx \leq \lambda^s(A \cap E_N).$$

A contradiction. Therefore $\phi(x) \leq 1$ λ^s-almost everywhere in $\mathbf{X}$. An analogous argument yields $\phi(x) \geq 0$ λ^s-almost everywhere. Therefore $\phi \in \mathcal{C}$. Now, I need to show that $\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})$. By assumption, for every $\theta \in \mathrm{Bd}\Theta_0$ $f(\cdot;\theta)$ is an element of $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$. In addition, $g \in \mathcal{G}$ is bounded. Consequently, $f(\cdot, \theta)g(\cdot) \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$. Since $\phi_n \in \mathcal{C}(\alpha\text{-}\mathcal{G})$ for every $n \in \mathbb{N}$ weak*

convergence yields

$$
\begin{aligned}
0=\lim_{n\to\infty}\int_{\mathbf{X}} f(x;\theta)g(x)(\phi_n(x)-\alpha)dx &= \Big(\lim_{n\to\infty}\int_{\mathbf{X}} f(x;\theta)g(x)\phi_n(x)dx\Big)-\int_{\mathbf{X}} f(x;\theta)g(x)\alpha dx\\
&= \int_{\mathbf{X}} f(x;\theta)g(x)\phi(x)dx-\int_{\mathbf{X}} f(x;\theta)g(x)\alpha dx\\
&= \int_{\mathbf{X}} f(x;\theta)g(x)(\phi(x)-\alpha)dx.
\end{aligned}
$$

So $\phi\in\mathcal{C}(\alpha\text{-}\mathcal{G})$. This implies $\mathcal{C}(\alpha\text{-}\mathcal{G})$ is sequentially closed in $\mathcal{C}$ endowed with the weak* topology.

(*Compactness*) Let

$$V\equiv\Big\{f\in L^1(\mathbf{X},\mathcal{B}(\mathbf{X}),\lambda^s)\;:\;\int_{\mathbf{X}}|f(x)|dx\leq 1\Big\}$$

Note that V is a neighborhood of the function $\mathbf{0}$ in the space $L^1(\mathbf{X},\mathcal{B}(\mathbf{X}),\lambda^s)$. Let

$$K\equiv\Big\{g\in L^\infty(\mathbf{X},\mathcal{B}(\mathbf{X}),\lambda^s)\;:\;\Big|\int_{\mathbf{X}} f(x)g(x)\Big|dx\leq 1\quad\forall\quad f\in V\Big\}. \tag{A.1.1}$$

Note that $C(\alpha\text{-}\mathcal{G})\subseteq\mathcal{C}\subseteq K$, as for any test $\Big|\int_{\mathbf{X}} f(x)\phi(x)dx\Big|\leq\int_{\mathbf{X}}|f(x)|\phi(x)dx\leq\int_{\mathbf{X}}|f(x)|dx\leq 1$. By the Banach-Alaouglu Theorem the set K is compact in the weak* topology; see Rudin (2005), p. 68, Theorem 3.15. Furthermore, the space $L^1(\mathbf{X},\mathcal{B}(\mathbf{X}),\lambda^s)$ is topologically separable as $(\mathbf{X},\mathcal{B}(\mathbf{X}),\lambda^s)$ is a separable measure space; see exercise 10, Chapter 1 of Stein (2011). Therefore, Theorem 3.16 in Rudin (2005) p. 70 implies that the topological space $(K,\mathcal{T}^*_K)$ is compact and metrizable. Since every metrizable space is first-countable—consequently, Frechet-Urysohn— the sequential closure of $C(\alpha)$ coincides with its closure. Therefore, the set $D^*(\alpha)$ is a closed subset of the compact topological space $(K,\mathcal{T}^*_K)$. I conclude that $(\mathcal{C}(\alpha\text{-}\mathcal{G}),\mathcal{T}^*_{\mathcal{C}(\alpha\text{-}\mathcal{G})})$ is compact and metrizable. *Q.E.D.*

Corollary 1: The space of α-similar-on-the-boundary tests, $\mathcal{C}(\alpha\text{-sb})$ is weak* compact.

Proof: Set $\mathcal{G}=\{g(x)=1\;\forall\,x\in\mathbf{X}\}$. *Q.E.D.*

Corollary 2: The space of α-conditionally similar-on-the boundary tests, $\mathcal{C}(\alpha\text{-csb})$ is weak* compact.

Proof: Set $\mathcal{G}=\{g\;\big|\;g(x_1,x_2)=1\text{ if }x_2\in\mathcal{F};g(x_1,x_2)=0\text{ if }x\in\mathcal{F}\text{ for some }\mathcal{F}\in\mathcal{B}(\mathbf{X}_2)\}$. Fix $\theta\in\text{Bd}\Theta_0$ and consider the random variables $\phi(X_1,X_2)$ and $Y\equiv X_2$ defined on the probability space $(\mathbf{X},\mathcal{B}(\mathbf{X}),P_\theta)$, where P_θ is the measure induced by $f(x;\theta)$. Note that

$$\mathbb{E}_\theta[(\phi(X_1,X_2)-\alpha)g(X_1,X_2)]=\alpha\quad\forall g\in\mathcal{G}$$

implies that

$$\int_{\mathbf{X}_1\times\mathcal{F}}(\phi(X_1,X_2)-\alpha)dP_\theta=0\quad\forall\mathcal{F}\in\mathcal{B}(\mathbf{X}_2).$$

By definition of conditional expectation (Billingsley (1995) p. 445), it follows that

$$E[\phi(X_1, X_2) - \alpha | X_2] = 0,$$

except, perhaps, in a set of measure zero under P_θ. And this holds for every $\theta \in \mathrm{Bd}\Theta_0$. *Q.E.D.*

A.2. *Proof of Theorem 1*

T1a (Outline): I have shown that the class of tests $\mathcal{C}(\alpha\text{-}\mathcal{G})$ is weak* compact. This class is non-empty, as it contains the randomized test $\phi(x) = \alpha$. To establish Theorem 1 it will be sufficient to show that the objective function

$$\mathcal{W}^*(\phi) \equiv \mathbf{W}\Big(\int_{\text{int } \Theta_1} R(\phi, \theta) p_1(\theta) d\theta \, , \int_{\text{Int } \Theta_0} R(\phi, \theta) p_0(\theta) d\theta \Big).$$

is continuous in the weak* topology.

T1a-Step 1 (Fubini's Theorem:) Since the image of any test $\phi \in \mathcal{C}$ is contained in the interval $[0, 1]$ λ^s-a.e. and $f(x;\theta) \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ for all θ, it follows that $\Big(\int_{\mathbf{X}} \phi(x) f(x;\theta) dx \Big) \leq 1$ for every $\theta \in \Theta$. Furthermore, since $p_1(x)$ and $p_0(x)$ are also probability density functions on $\mathrm{Int}\Theta_1$ and $\mathrm{Int}\Theta_0$ the following holds

$$\int_{\mathrm{Int}\Theta_1} \Big(\int_{\mathbf{X}} \phi(x) f(x;\theta) dx \Big) p_1(\theta) d\theta \leq 1 < \infty$$

and

$$\int_{\mathrm{Int}\Theta_0} \Big(\int_{\mathbf{X}} \phi(x) f(x;\theta) dx \Big) p_0(\theta) d\theta \leq 1 < \infty.$$

Therefore, an application of Fubini's theorem in Billingsley (1995), p. 234 yields

$$\int_{\text{int } \Theta_1} R(\phi, \theta) p_1(\theta) d\theta \equiv \int_{\text{int } \Theta_1} \Big(\int_{\mathbf{X}} (1 - \phi(x)) f(x;\theta) dx \Big) p_1(\theta) d\theta = \int_{\mathbf{X}} (1 - \phi(x)) f_1^*(x) dx$$

and

$$\int_{\text{Int } \Theta_0} R(\phi, \theta) p_0(\theta) d\theta \equiv \int_{\text{int } \Theta_0} \Big(\int_{\mathbf{X}} \phi(x) f(x;\theta) dx \Big) p_0(\theta) d\theta = \int_{\mathbf{X}} \phi(x) f_0^*(x) dx.$$

where f_1^* and f_0^* are the "integrated" likelihoods given by

$$\text{(A.2.1)} \qquad f_1^*(x) \equiv \int_{\text{int } \Theta_1} f(x;\theta) p_1(x) d\theta, \quad f_0^*(x) \equiv \int_{\text{int } \Theta_0} f(x;\theta) p_0(x).$$

Note that both f_1^* and f_0^* are elements of $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$. Note that the mapping $\mathbf{W} : \mathbb{R}^2 \to \mathbb{R}$ induces a functional $\mathcal{W}^*$ over $\mathcal{D}$:

$$\text{(A.2.2)} \qquad \mathcal{W}^*(\phi) \equiv \mathbf{W}\Big(\int_{\mathbf{X}} (1 - \phi(x)) f_1^*(x) dx, \int_{\mathbf{X}} \phi(x) f_0^*(x) dx \Big)$$

T1a-Step 2 (Sequential Continuity of $\mathcal{W}^*$:) I now show that $\mathcal{W}^*$ is continuous on the compact metrizable space $(\mathcal{C}(\alpha\text{-}\mathcal{G}), \mathcal{T}^*_{\mathcal{C}(\alpha\text{-}\mathcal{G})})$. It suffices to establish sequential continuity. Take any sequence of tests $\phi_n \to^* \phi$. Since both f_1^* and f_0^* are elements of $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda_{\mathbf{X}})$, convergence in the weak* topology yields

$$\int_{\mathbf{X}} \phi_n(x) f_1^*(x)dx \to \int_{\mathbf{X}} \phi(x) f_1^*(x)dx \quad \text{and} \quad \int_{\mathbf{X}} \phi_n(x) f_0^*(x)dx \to \int_{\mathbf{X}} \phi(x) f_0^*(x)dx.$$

Therefore, the continuity of $\mathbf{W}$ implies

$$\begin{aligned} \mathcal{W}^*(\phi_n) &\equiv \mathbf{W}\Big(1 - \int_{\mathbf{X}} \phi_n(x) f_1^*(x)dx, \int_{\mathbf{X}} \phi_n(x) f_0^*(x)dx\Big) \to \mathbf{W}\Big(1 - \int_{\mathbf{X}} \phi(x) f_1^*(x)dx, \int_{\mathbf{X}} \phi(x) f_0^*(x)dx\Big) \\ &= \mathcal{W}^*(\phi). \end{aligned}$$

Therefore, $\mathcal{W}^*$ is a continuous functional defined on the compact space $(\mathcal{C}(\alpha\text{-}\mathcal{G}), \mathcal{T}^*_{\mathcal{C}(\alpha\text{-}\mathcal{G})})$, and $\mathcal{C}(\alpha\text{-}\mathcal{G}) \neq \emptyset$, as it contains the test $\phi(x) = \alpha$. This implies $M(\mathbf{W}, p_1, p_0, \mathcal{G}) \neq \emptyset$.

T1b : Let $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G})$. I show that if $\phi' \in \mathcal{C}(\alpha\text{-}\mathcal{G})$ satisfies

$$\mathbb{E}_\theta[\phi'(X)] \leq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_0 \tag{A.2.3}$$

and

$$\mathbb{E}_\theta[\phi'(X)] \geq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_1 \tag{A.2.4}$$

then

$$\mathbb{E}_\theta[\phi'(x)] = \mathbb{E}_\theta[\phi^*(x)] \quad \forall \quad \theta \in \Theta = \Theta_0 \cup \Theta_1. \tag{A.2.5}$$

Consequently, there is no test $\phi' \in \mathcal{C}(\alpha\text{-}\mathcal{G})$ that "weakly dominates" ϕ^* ; i.e, $R(\phi', \theta) \leq R(\phi^*, \theta)$ with strict inequality for some θ.

Suppose (A.2.3) and (A.2.4) hold, but (A.2.5) does not. Then, one of the following claims is true:

C1 There exists $\tilde{\theta} \in \Theta_1$ such that $\Delta_{\phi',\phi^*}(\tilde{\theta}) \equiv \mathbb{E}_{\tilde{\theta}}[\phi'(X)] - \mathbb{E}_{\tilde{\theta}}[\phi^*(X)] > 0$

C2 There exists $\tilde{\theta} \in \Theta_0$ such that $\Delta_{\phi',\phi^*}(\tilde{\theta}) \equiv < \mathbb{E}_{\tilde{\theta}}[\phi'(X)] - \mathbb{E}_{\tilde{\theta}}[\phi^*(X)] < 0$.

Assume first that C1 holds. The continuity of $\Delta_{\phi',\phi}(\cdot)$ at $\tilde{\theta}$ implies the existence of an open neighborhood $\tau_{\tilde{\theta}}$ for which $\Delta_{\phi',\phi^*}(\theta) < 0$ for all $\theta \in \tau_{\tilde{\theta}}$. Note that $\Theta_1 \neq \emptyset$ is an open set. It follows that the set $\mathcal{S}_{\tilde{\theta}}$ defined by $\mathcal{S}_{\tilde{\theta}} \equiv \tau_{\tilde{\theta}} \cap \Theta_1$ satisfies three properties: it is non-empty, it is open, and it is contained in Θ_1. Since $p_1(\theta)$ has full support on $\text{Int}\Theta_1$, $\int_A p_1(\theta)d\theta > 0$ for any open set A contained in Θ_1. Note that $\Delta_{\phi',\phi^*}(\theta) > 0$ for all $\theta \in S_{\tilde{\theta}}$ and equations (A.2.3)-(A.2.4) imply

$$\int_{\text{Int}\Theta_0} \Big(\int_{\mathbf{X}} \phi'(x) f(x;\theta)dx\Big) p_0(\theta)d\theta \leq \int_{\text{Int}\Theta_0} \Big(\int_{\mathbf{X}} \phi^*(x) f(x;\theta)dx\Big) p_0(\theta)d\theta.$$

and

$$\int_{\Theta_1}\Big(\int_{\mathbf{X}}(1-\phi'(x))f(x;\theta)dx\Big)p_1(\theta)d\theta < \int_{\Theta_1}\Big(\int_{\mathbf{X}}(1-\phi^*(x))f(x;\theta)dx\Big)p_1(\theta)d\theta$$

The monotonicity of $\mathbf{W}$ implies that $\mathcal{W}^*(\phi') < \mathcal{W}(\phi^*)$. This contradicts the fact that $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G})$. I conclude C1 cannot hold.

Now, suppose C2 holds. Since the function $g^*(x) = 1$ belongs to $\mathcal{G}$, then $\tilde{\theta}$ must belong to $\text{Int}\Theta_0$. If $\text{Int}\Theta_0 = \emptyset$ the proof is over. If $\text{Int}\Theta_0 \neq \emptyset$ then —by analogy with the previous paragraph— there exists an open set $S_{\tilde{\theta}}$ contained in $\text{Int}\Theta_0$ such that $\Delta_{\phi',\phi^*}(\theta) < 0$ for all $\theta \in S_{\tilde{\theta}}$. Since this set has positive probability under p_0, this implies

$$\int_{\text{Int}\Theta_0}\Big(\int_{\mathbf{X}}\phi'(x)f(x;\theta)dx\Big)p_0(\theta)d\theta < \int_{\text{Int}\Theta_0}\Big(\int_{\mathbf{X}}\phi^*(x)f(x;\theta)dx\Big)p_0(\theta)d\theta$$

and

$$\int_{\Theta_1}\Big(\int_{\mathbf{X}}(1-\phi'(x))f(x;\theta)dx\Big)p_1(\theta)d\theta \leq \int_{\Theta_1}\Big(\int_{\mathbf{X}}(1-\phi^*(x))f(x;\theta)dx\Big)p_1(\theta)d\theta.$$

Which, once again, contradicts the fact that $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G})$.

Therefore, (A.2.3) and (A.2.4) imply (A.2.5). I conclude that ϕ^* is admissible in $\mathcal{C}(\alpha\text{-}\mathcal{G})$.

T1c (Outline): Let $\mathcal{G}^* \equiv \{g : \mathbf{X} \to \mathbb{R} \mid g(x) = 1 \quad \forall\, x \in \mathbf{X}\}$, so that the class $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$ coincides with $\mathcal{C}(\alpha\text{-sb})$. I show that a test $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}^*)$ is admissible in the class of all tests. The proof is divided into two steps.

Step 1: First I show that if $\phi' \in \mathcal{C}$ satisfies

(A.2.6) $\quad \mathbb{E}_\theta[\phi'(X)] \leq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_0$

and

(A.2.7) $\quad \mathbb{E}_\theta[\phi'(X)] \geq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_1$

then ϕ' is α-similar on $\text{Bd}\Theta_0$. Consequently, any test ϕ' that "weakly dominates" ϕ^* (i.e, $R(\phi', \theta) \leq R(\phi^*, \theta)$ with strict inequality for some θ) must be α-similar on the boundary of Θ_0.

Let $\mathcal{C}_{ns} \subset \mathcal{C}$ be the class of tests that are not similar on the boundary of Θ_0. This is, $\phi \in \mathcal{C}_{ns}$ if and only if there exists $\theta, \theta' \in \text{Bd}\Theta_0$ such that $\mathbb{E}_\theta[\phi(x)] \neq \mathbb{E}_{\theta'}[\phi(x)]$. Partition $\mathcal{C}$ according to $\mathcal{C}_{ns}$ so that $\mathcal{C} \equiv \mathcal{C}_{ns} \cup (\mathcal{C} \backslash \mathcal{C}_{ns})$. Take any test $\phi' \in \mathcal{C}_{ns}$ that satisfies (A.2.3). Since ϕ' is an element of $\mathcal{C}_{ns}$ and Θ_0 contains its boundary (as it is closed), there exists $\theta \in \text{Bd}\Theta_0$ such that

$\Delta_{\phi',\phi^*}(\theta) \equiv \mathbb{E}_\theta[\phi'(X)] - \mathbb{E}_\theta[\phi^*(X)] < 0$. Because $\Delta_{\phi',\phi^*}(\theta) < 0$ and the function $\Delta_{\phi',\phi^*}(\cdot)$ is continuous at θ, there exists an open neighborhood $\tau_\theta \in \mathcal{T}$ such that $\Delta_{\phi',\phi^*}(\theta) < 0$ for all $\theta \in \tau_\theta$. Since θ is an element of Bd Θ_0, then $\tau_\theta \cap \Theta_1 \neq \emptyset$. The latter implies there exists $\theta_1 \in \Theta_1$ such that $\Delta_{\phi',\phi^*_\alpha}(\theta_1) = \mathbb{E}_{\theta_1}[\phi'(X)] - \mathbb{E}_{\theta_1}[\phi^*_\alpha(X)] < 0$. Therefore, equation (A.2.3) and (A.2.4) cannot hold. We conclude there is no test $\phi' \in \mathcal{C}_{ns}$ that satisfies (A.2.3) and (A.2.4).
Since $\mathcal{C}_{ns}$ partitions $\mathcal{C}$, a test $\phi' \in \mathcal{C}$ that satisfies (A.2.3) and (A.2.4) must be an element of $\mathcal{C}\backslash\mathcal{C}_{ns}$ (as $\phi' \notin \mathcal{C}_{ns}$). Equation (A.2.3) implies ϕ' is α'-similar on the boundary with $\alpha' \leq \alpha$. Two cases follow: $\alpha' < \alpha$ or $\alpha' = \alpha$. In the first case, the argument in the previous paragraph implies that ϕ' will violate (A.2.4). We conclude that any test that satisfies (A.2.3) and (A.2.4) must be α-similar on $\text{Bd}\Theta_0$.

Step 2: Since $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}^*)$, ϕ^* is admissible in $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$. Therefore, there is no α-similar-on-the-boundary test such that $R(\phi', \theta) \leq R(\phi^*, \theta)$ with strict inequality for some $\theta \in \Theta$. Since —by Step 1— any test $\phi' \in C$ that that satisfies (A.2.3) and (A.2.4) must be α-similar on $\text{Bd}\Theta_0$, I conclude ϕ^* is admissible in $\mathcal{C}$

A.3. *Proof of Theorem 2*

Step 1 T2: (ECS-tests objective function). Let

$$\mathbf{X}_1(x_2) \equiv \{x_1 \in \mathbf{X}_1 \mid (x_1, x_2) \in \mathbf{X}\}.$$

Fubini's theorem (T1a-Step 1) implies that ϕ^* is an ECS tests if and only if ϕ^* solves the problem:

$$\min_{\phi \in \mathcal{C}} \quad \tau \int_X (1 - \phi(x)) f_1^*(x) dx + (1 - \tau) \int_X \phi(x) f_0^*(x) dx$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) f_{\text{Bd}}(x_1 | x_2) dx_1 = \alpha$$

except, perhaps, for x_2 that belong to a set of measure zero under every $h(x_2, \theta)$, $\theta \in \text{Bd}\Theta_0$. Re-write the objective function as

$$\max_{\phi \in \mathcal{C}} \quad \tau \int_X \phi(x) f_1^*(x) dx - (1 - \tau) \int_X \phi(x) f_0^*(x) dx.$$

The product structure of $\mathbf{X}$ and the linearity of the integral allows a further expansion of the previous equation:

$$\max_{\phi \in \mathcal{C}} \quad \int_{X_2} \Big(\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) [\, \tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2) \,] dx_1 \Big) dx_2$$

Step 2 T2: (Necessity) First I show that the test $\phi^*(x_1, x_2)$ that rejects the null hypothesis whenever

$$[\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] / f_{\text{Bd}}(x_1 | x_2) > c(x_2; \alpha)$$

is an element of the set $M(\tau, p_1, p_0)$—provided $c(x_2, \alpha)$ is defined as the (1-α) quantile of the random variable $z_{ecs}(X_1, x_2; p_1, p_0, \tau)$ for every $x_2 \in \mathbf{X}_2$. That is to say

$$c(x_2;\alpha) \equiv \underset{q \in \mathbf{X}_1(x_2)}{\arg\min}\ \mathbb{E}_{f_{\mathrm{Bd}}(x_1|x_2)}\Big[\rho_{(1-\alpha)}\Big(z_{\mathrm{ECS}}(x_1,x_2;p_1,p_0,\tau)-q\Big)\Big].$$

Note first that the Generalized Neyman Pearson Lemma in Ferguson (1967) p. 204 implies that for a fixed x_2 the test $\phi^*(x_1, x_2)$ solves the problem

$$\max_{\phi \in \mathcal{C}} \quad \int_{\{x_1 \in \mathbf{X}_1 \ | \ (x_1,x_2)\in \mathbf{X}\}} \phi(x_1,x_2)[\,\tau f_1^*(x_1,x_2)-(1-\tau)f_0^*(x_1,x_2)\,]dx_1$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1,x_2) f_{\mathrm{Bd}}(x_1|x_2)dx_1 = \alpha.$$

Hence, to show that $\phi^*(x_1, x_2) \in M(\tau, p_1, p_0)$ it only remains to prove that $\phi^*(x_1, x_2)$ is measurable. That is, $\phi^*(x_1, x_2) \in \mathcal{C}_{X_2}(\alpha\text{-csb})$. Assumption R1 imply that $\phi^*(x_1, x_2)$ is continuous in x_1, for every x_2. Assumption R2 imply that the test is measurable in x_2, for every x_1. Therefore, $\phi^*(x_1, x_2)$ is a Carathéodory function, as defined in Aliprantis and Border (2006), p. 153. Since the sample space $\mathbf{X}$ is separable (by assumption) and metrizable (for it is a subset of a euclidean space), Lemma 4.5.1 in Aliprantis and Border (2006) p. 153 implies $\phi^* : \mathbf{X} \to [0, 1]$ is measurable.

Step 3 T2 (Sufficiency) Now I show that ECS tests are equal to ϕ^* almost everywhere in $\mathbf{X}$. Let $\phi' \in M(\tau, p_1, p_0)$. I claim there is no set of lebesgue $\lambda_{\mathbf{X}_2}$-positive measure in $A \in \mathcal{B}(\mathbf{X}_2)$ such that for each $x_2 \in A$, $\phi^*(x_1, x_2) \neq \phi'(x_1, x_2)$ in a set of lebesgue $\lambda_{\mathbf{X}_1}$-positive measure in $\mathbf{X}_1$. Suppose this is not the case. The maximizer in Step 2 is unique almost surely in $X_1(x_2)$. Hence, for all $x_2 \in A$

$$\int_{\mathbf{X}_1(x_2)} \phi^*(x_1,x_2)[\tau f_1^*(x_1,x_2)-(1-\tau)f_0^*(x_1,x_2)]dx_1 > \int_{\mathbf{X}_1(x_2)} \phi'(x_1,x_2)[\tau f_1^*(x_1,x_2)-(1-\tau)f_0^*(x_1,x_2)]dx_1$$

Since A has positive measure, Integrating over x_2 yields

$$\int_{\mathbf{X}} \phi^*(x)[\tau f_1^*(x)-(1-\tau)f_0^*(x)]dx > \int_{\mathbf{X}} \phi'(x)[\tau f_1^*(x)-(1-\tau)f_0^*(x)]dx$$

which contradicts the fact that ϕ' is an ECS-test.

A.4. *Example 1: Boundary Sufficiency for the IV model*

Let $\phi_n(\cdot, \mu, W)$ denote the probability density function of a multivariate normal random vector of dimension n with mean parameter μ and covariance matrix W. Consider the standarized OLS coefficients:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2})\widehat{\gamma}_{\mathrm{OLS}} = \begin{pmatrix} (Z'Z)^{-1/2}Z'(y_1 - Y_2\beta_0)(b_0'\Omega b_0)^{-1/2} \\ \\ \mathrm{vec}[(Z'Z)^{1/2}Z'Y\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}] \end{pmatrix}$$

which yield the statistical model for IV:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \sim \mathcal{N}_{k+nk} \begin{pmatrix} \begin{matrix} (b_0'\Omega b_0)^{-1/2}(Z'Z)^{1/2}\Pi(\beta-\beta_0) \\ \text{vec}[(Z'Z)^{1/2}\Pi\, A'\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}] \end{matrix} & , & \mathbb{I}_{k+nk} \end{pmatrix}$$

Note that

$$f(\gamma_1^*, \gamma_2^*; \beta=\beta_0, \Pi) = \phi_k(\gamma_1^*, \mathbf{0}, \mathbb{I}_k)\phi_{nk}(\gamma_2^*, \text{vec}[(Z'Z)^{1/2}\Pi\, A'\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}], \mathbb{I}_{nk})$$

Therefore, γ_2^* is a boundary-sufficient statistic. Furthermore, for parameters (β, Π) on the boundary of the null hypothesis; that is, whenever $\beta = \beta_0$ and $\Pi \in \mathbb{R}^{nk\times n}$:

$$\gamma_2^* \sim \mathcal{N}_{nk}\left(\text{vec}[(Z'Z)^{1/2}\Pi\,(A_0'\Omega^{-1}A_0)^{1/2}], \mathbb{I}_{nk}\right)$$

Using the properties of the kronecker product, the mean vector can be written as

$$((A_0'\Omega^{-1}A_0)^{1/2} \otimes (Z'Z)^{1/2})\text{vec}(\Pi)$$

Since both matrices $A_0'\Omega^{-1}A_0$ and $(Z'Z)^{1/2}$ are of full rank the distributions for γ_2^* can be re-parameterized as a function of an unrestrictied element of $\mathbb{R}^{k\times n}$. That γ_2^* is boundedly complete follows from Theorem 4.3.1 in Lehmann and Romano (2005) implies.

A.5. *Proof of Result 1*

The rotated and standarized just-indentified IV model has the following distribution:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2})\widehat{\gamma}_{\text{OLS}} \sim \mathcal{N}_{n+n^2} \begin{pmatrix} (C_0 \otimes (Z'Z)^{1/2}) \begin{matrix} \Pi\beta \\ \text{vec}(\Pi) \end{matrix} & , & \mathbb{I}_{n+n^2} \end{pmatrix}$$

Consider the following prior distribution over the parameters (β, Π) :

$$\Pi = [\gamma_{21}, \gamma_{22}, \ldots \gamma_{2n}]$$

$$\beta = [\gamma_{21}, \gamma_{22}, \ldots \gamma_{2n}]^{-1}\gamma_1$$

where

$$\gamma \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \lambda^2\Omega \otimes (Z'Z)^{-1})$$

λ^2 is introduced as an extra parameter controlling the precision of the prior. I will show that regardless of the value of λ^2 the ECS test rejects for large values of the Anderson and Rubin (1949) test.

Write $\gamma = (\gamma_1', \gamma_{21}', \gamma_{22}', \ldots \gamma_{2n}')'$ with γ_1 is $n \times 1$ and γ_{2i} is $n \times 1$ for all $i = 1, \ldots n$ and re-write the density of the vector $({\gamma_1^*}', {\gamma_2^*}')$ as $f(\gamma_1^*, \gamma_2^*; \beta(\gamma), \Pi(\gamma))$. The integrated likelihood in Theorem 2 is given by:

$$\begin{aligned}
f_1^*(\gamma_1^*,\gamma_2^*) &= \int_{\mathbb{R}^{n+n^2}} f(\gamma_1^*,\gamma_2^*;\beta(\gamma),\Pi(\gamma))\phi_{n+n^2}(\gamma,\mathbf{0},\lambda^2\Omega\otimes(Z'Z)^{-1})d\gamma \\
&= a_1\int_{\mathbb{R}^{n+n^2}} \exp\Big(-\frac{1}{2}\Big[\gamma^*-(C_0\otimes(Z'Z)^{1/2})\gamma\Big]'\Big[\gamma^*-(C_0\otimes(Z'Z)^{1/2})\gamma\Big]\Big) \\
&\quad \phi_{n+n^2}(\gamma,\mathbf{0},\lambda^2\Omega\otimes(Z'Z)^{-1})d\gamma \\
&\quad \Big(\text{(where I have used the fact } \begin{pmatrix}\Pi(\gamma)\beta(\gamma)\\ \text{vec}(\Pi(\gamma))\end{pmatrix} = \begin{pmatrix}\gamma_1\\ \text{vec}(\gamma_{21},\gamma_{22},\ldots\gamma_{2n})\end{pmatrix}\Big)\Big) \\
&= a_2\int_{\mathbb{R}^{n+n^2}} \exp\Big(-\frac{1}{2}\Big[\gamma^*-(C_0\otimes(Z'Z)^{1/2})\gamma\Big]'\Big[\gamma^*-(C_0\otimes(Z'Z)^{1/2})\gamma\Big]\Big) \\
&\quad \exp\Big(-\frac{1}{2\lambda^2}\gamma'(\Omega^{-1}\otimes Z'Z)\gamma\Big)d\gamma \\
&\quad \text{(By definition of } \phi_{n+n^2}) \\
&= a_2\int_{\mathbb{R}^{n+n^2}} \exp\Big(-\frac{1}{2}\Big[\gamma^*-(C_0\otimes(Z'Z)^{1/2})\gamma\Big]'\Big[\gamma^*-(C_0\otimes(Z'Z)^{1/2})\gamma\Big]\Big) \\
&\quad \exp\Big(-\frac{1}{2\lambda^2}\gamma'(C_0'\otimes Z'Z^{1/2})(C_0\otimes(Z'Z)^{1/2})\gamma\Big)d\gamma \\
&\quad (\text{since } C_0'C_0=\Omega^{-1}) \\
&= a_3\int_{\mathbb{R}^{n+n^2}} \exp\Big(-\frac{1}{2}\Big[\gamma^*-\mu^*\Big]'\Big[\gamma^*-\mu^*\Big]\Big)\exp\Big(-\frac{1}{2\lambda^2}\mu^{*\prime}\mu^*\Big)d\gamma \\
&= (\text{ using the (linear) change of variables } \mu^*=(C_0\otimes(Z'Z)^{1/2})\gamma) \\
&\quad a_3\Big[\int_{R^{m_1}} \exp\Big(-\frac{1}{2}(\gamma_1^*-\mu_1^*)'(\gamma_1-\mu_1^*)\Big)\exp\Big(-\frac{1}{2\lambda^2}\mu_1^{*\prime}\mu_1^*\Big)d\mu_1^*\Big] \\
&\quad \Big[\int_{R^{m_2}} \exp\Big(-\frac{1}{2}(\gamma_2-\mu_2^*)'(\gamma_2-\mu_2^*)\Big)\exp\Big(-\frac{1}{2\lambda^2}\mu_2^{*\prime}\mu_2^*\Big)d\mu_2^*\Big] \\
&= a_3\exp\Big(-\frac{1}{2}\gamma_1^{*\prime}\gamma_1^*\Big)\int_{\mathbb{R}^{m_1}} \exp\Big(\gamma_1^{*\prime}\mu_1^*\Big)\exp(-\frac{1}{2b^2}\mu_1^{*\prime}\mu_1^*)d\mu_1^* \\
&\quad \exp\Big(-\frac{1}{2}\gamma_2^{*\prime}\gamma_2\Big)\int_{\mathbb{R}^{m_2}} \exp\Big(x_2'\mu_2^*\Big)\exp(-\frac{1}{2b^2}\mu_2^{*\prime}\mu_2^*)d\mu_2^* \\
&\quad (\text{where } b^2=\lambda^2/(1+\lambda^2)) \\
&= c_1\exp\Big(-\frac{1}{2}\gamma_1^{*\prime}\gamma_1\Big)\exp\Big(\frac{b^2}{2}\gamma_1^{*\prime}\gamma_1\Big)\exp\Big(-\frac{1}{2}\gamma_2^{*\prime}\gamma_2\Big)\exp\Big(\frac{b^2}{2}\gamma_2^{*\prime}\gamma_2^*\Big) \\
&\quad \text{(where I have used the definition of the Moment Generating Function} \\
&\quad \text{of a multivariate normal; and } c_1 \text{ is a non-negative constant.)}
\end{aligned}$$

Note that the boundary conditional likelihood for the IV model is given by:

$$f_{\text{Bd}}(\gamma_1^*|\gamma_2^*) = \phi_n(\gamma_1^*,\mathbf{0},\mathbb{I}_n)$$

The ECS test statistic of Theorem 2 is defined as:

$$\begin{aligned}
z(\gamma_1^*,\gamma_2^*;p_1) &= f_1^*(\gamma_1^*,\gamma_2^*)/f_{\text{Bd}}(\gamma_1^*|\gamma_2^*) \\
&= c_2\exp\Big(\frac{b^2}{2}\gamma_1^{*\prime}\gamma_1^*\Big)\exp\Big(-\frac{1}{2}\gamma_2^{*\prime}\gamma_2^*\Big)\exp\Big(\frac{b^2}{2}\gamma_2^{*\prime}\gamma_2^*\Big)
\end{aligned}$$

where c_2 is a non-negative constant. Consider the critical value function

$$c(x_2;\alpha) = c_2 \exp\Big(\frac{b^2}{2}\chi^2_{n,1-\alpha}\Big)\exp\Big(-\frac{1}{2}\gamma_2^{*\prime}\gamma_2^*\Big)\exp\Big(\frac{b^2}{2}\gamma_2^{*\prime}\gamma_2^*\Big)$$

where $\chi^2_{m_1,1-\alpha}$ is the $1-\alpha$ quantile of a central $\chi^2_{m_1}$ distribution. Note that $c(\gamma_2^*;\alpha)$ is measurable and for each fixed γ_2^* and

$$\mathbb{P}_{f_{\mathrm{Bd}}(\gamma_1^*|\gamma_2^*)}(z(\gamma_1^*,\gamma_2^*;p_1) > c(\gamma_2^*,\alpha)) = \alpha$$

Hence, the assumptions of Theorem 2 are satisfied and the ECS rejects if:

$$c_2 \exp\Big(\frac{b^2}{2}\gamma_1^{*\prime}\gamma_1^*\Big)\exp\Big(-\frac{1}{2}\gamma_2^{*\prime}\gamma_2^*\Big)\exp\Big(\frac{b^2}{2}\gamma_2^{*\prime}\gamma_2^*\Big)$$

exceeds the critical value function

$$c_2 \exp\Big(\frac{b^2}{2}\chi^2_{n,1-\alpha}\Big)\exp\Big(-\frac{1}{2}\gamma_2^{*\prime}\gamma_2^*\Big)\exp\Big(\frac{b^2}{2}\gamma_2^{*\prime}\gamma_2^*\Big)$$

which happens if and only if $\gamma_1^{*\prime}\gamma_1^* > \chi^2_{n,1-\alpha}$. Since

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2})\widehat{\gamma}_{\mathrm{OLS}} = \begin{pmatrix} (Z'Z)^{-1/2}Z'(y_1 - Y_2\beta_0)(b_0'\Omega b_0)^{-1/2} \\ \\ (Z'Z)^{1/2}Z'Y\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2} \end{pmatrix}$$

The ECS test rejects for large values of the Anderson and Rubin (1949) statistic.

A.6. *Proof of Result 2*

Chamberlain's (2007) re-parameterization is given by:

$$\rho = (A'\Omega^{-1}A)^{1/2}(\Pi'Z'Z\Pi)^{1/2}, \quad \phi = C_0A/(A'\Omega^{-1}A)^{1/2}, \quad \omega = (Z'Z)^{1/2}\Pi/(\Pi'Z'Z\Pi)^{1/2}$$

where $A \equiv [\beta, 1]'$, and

$$C_0 \equiv \begin{pmatrix} (b_0'\Omega b_0)^{-1/2}b_0' \\ \\ (A_0'\Omega^{-1}A_0)^{-1/2}A_0'\Omega^{-1} \end{pmatrix}$$

$$b_0 = [1, -\beta_0']' \quad A_0 = [\beta_0, 1]'$$

Remark: The value β_0 and the reduced-form covariance matrix Ω impose a restriction on the possible values for the parameter ϕ. Note first that $\Omega C_0'C_0 = \mathbb{I}_2$. Therefore:

$$\begin{aligned} (0,1)\Omega C_0'\phi &= (0,1)\Omega C_0'C_0A/A\Omega^{-1}A')^{1/2} \\ &= (0,1)\mathbb{I}_2(\beta,1)'\,/(A\Omega^{-1}A')^{1/2} \\ &= 1/(A\Omega^{-1}A')^{1/2} \geq 0 \end{aligned}$$

Hence, the parameter ϕ belongs to the intersection of the unit sphere $\mathcal{S}^1$ and the half space $\{x \in \mathbb{R}^2 \mid (0,1)\Omega C_0'x \geq 0\}$. In fact,

$$
\begin{aligned}
(0,1)\Omega C_0' &= \Big((0,1)\Omega b_0/(b_0'\Omega b_0)^{1/2}\ ,\ (0,1)A_0/(A_0'\Omega^{-1}A_0)^{1/2}\Big) \\
&= \Big((0,1)\Omega b_0/(b_0'\Omega b_0)^{1/2}\ , (\omega_1^2\omega_2^2-\omega_{12}^2)^{1/2}/(b_0'\Omega b_0)^{1/2}\Big) \\
&= \Big((0,1)\Omega b_0/(b_0'\Omega b_0)^{1/2}\ , (\omega_1^2\omega_2^2-(\omega_{12}-\beta_0\omega^2+\beta_0\omega_s^2)^2)^{1/2}/(b_0'\Omega b_0)^{1/2}\Big) \\
&= \Big(\omega_2 r(\beta_0), \omega_2\sqrt{1-r^2(\beta_0)}\Big)
\end{aligned}
$$

where $r(\beta_0)$ corresponds to the structural correlation implied by β_0:

$$
r(\beta_0) = (0,1)\Omega b_0/\Big[(b_0'\Omega b_0)^{1/2}\omega_2\Big]
$$

Thus, the domain for the parameter ϕ in the canonical model is given by:

$$
\boldsymbol{\Theta} = \Big\{(\rho^2,\phi) \in \mathbb{R}_+ \times \mathcal{S}^1\ :\ r(\beta_0)\phi_1 + \sqrt{1-r^2(\beta_0)}\,\phi_2 \geq 0\Big\}
$$

Derivation of the Integrated Likelihoods: Let:

$$
f(S,T;\rho,\phi,\omega) = c_1 \exp\Big(-\frac{1}{2}([S',T']'-\rho(\phi\otimes\omega))'([S',T']'-\rho(\phi\otimes\omega))\Big)
$$

where c_1 is a non-negative constant. Let $Q \equiv [S,T]'[S,T]$.

Step 1: (Integrate ω) Note that:

$$
\begin{aligned}
\widetilde{f}(S,T;\rho,\phi) &\equiv c_2\int_{\mathcal{S}^{k-1}} f(S,T;\rho,\phi,\omega)d\lambda_{\mathcal{S}^{k-1}}(\omega) \\
&= a_2(Q)\exp\Big(-\rho^2/2\Big)\int_{\mathcal{S}^{k-1}}\exp\Big(([S,T]\phi)'\rho\omega\Big)d\lambda_{\mathcal{S}^{k-1}}(\omega)
\end{aligned}
$$

where $\lambda_{\mathcal{S}^{k-1}}(\cdot)$ is the uniform measure over the $k-1$ dimensional sphere $\mathcal{S}^{k-1}$ defined in Chamberlain (2007) and Stroock (1999). In addition,

$$
a_2(Q) \equiv c_2\exp\Big(-\frac{1}{2}[S'S+T'T]\Big)
$$

c_2 is a non-negative constant.

Step 2: (Integrate ρ) By assumption $\rho \sim \sqrt{\lambda^2\chi_k^2}$ independently of ϕ and ω. The latter implies that the density of ρ (with parameter λ^2) is given by:

$$
m_1(\rho;\lambda) \equiv \frac{1}{\lambda^2}\frac{1}{2^{k/2}\Gamma(k/2)}(\rho^2/\lambda^2)^{(k/2)-1}e^{-(\rho^2/2\lambda^2)}2\rho
$$

Note that using Fubini's Theorem and the change of variables formula:

$$
\int_{\mathbb{R}+}\widetilde{f}(S,T;\rho,\phi)m_1(\rho;\lambda^2)d\rho
$$

$$
\begin{aligned}
&= a_2(Q)\int_{\mathbb{R}^+}\Big(\exp(-\rho^2/2)\int_{\mathcal{S}^{k-1}}\exp\Big(\Big[(S,T)\rho\phi\Big]'\boldsymbol{\omega}\Big)d\lambda_{\mathcal{S}^{k-1}}(\boldsymbol{\omega})\Big)m_1(\rho;\lambda)d\rho\\
&= a_2(Q)\int_{\mathcal{S}^{k-1}}\Big(\int_{\mathbb{R}_+}\exp\Big(\Big[(S,T)\rho\phi\Big]'\boldsymbol{\omega}\Big)m_1(\rho;\lambda)\exp(-\rho^2/2)d\rho\Big)d\lambda_{\mathcal{S}^{k-1}}(\boldsymbol{\omega})\\
&= a_3(Q)\int_{\mathcal{S}^{k-1}}\Big(\int_{\mathbb{R}_+}\exp\Big(\Big[(S,T)\rho\phi\Big]'\boldsymbol{\omega}\Big)\exp(-\rho^2/2b^2)\rho^{k-1}dp\Big)d\lambda_{\mathcal{S}^{k-1}}(\boldsymbol{\omega})\\
&\qquad\qquad\qquad\qquad(\text{by definition of } m_1,\ b^2\equiv[\,\lambda^2/(1+\lambda^2)\,]\quad)\\
&= a_3(Q)\int_{\mathcal{S}^{k-1}}\Big(\int_{\mathbb{R}_+}\exp\Big(\Big[(S,T)\phi\Big]'\rho\omega\Big)\exp(-(\rho\boldsymbol{\omega})'(\rho\boldsymbol{\omega})/2b^2)\rho^{k-1}dp\Big)d\lambda_{\mathcal{S}^{k-1}}(\boldsymbol{\omega})
\end{aligned}
$$

where the last line follows from $\boldsymbol{\omega}'\boldsymbol{\omega}=1$ and $a_3(Q)=a_2(Q)2/(\lambda^k 2^{k/2}\Gamma(k/2))$. Finally, consider the non-negative measurable function $f:\mathbb{R}^k\to\mathbb{R}$

$$
f(x)=\exp\Big(\Big[(S,T)\phi\Big]'x\Big)\exp(-x'x/2b^2)
$$

Theorem 5.2.2, p. 86 in Stroock (1999) implies:

$$
\begin{aligned}
\int_{\mathbb{R}^+}\widetilde{f}(S,T;\rho,\phi)m_1(\rho;\lambda^2)d\rho &= a_3(Q)\int_{\mathbb{R}^K}\exp\Big(\Big[(S,T)\phi\Big]'x\Big)\exp(-x'x/2b^2)dx\\
&= a_4(Q)\exp\Big(\frac{b^2}{2}\phi'Q\phi\Big)
\end{aligned}
$$

where the last inequality follows by definition of the moment generating function of a k-dimensional multivariate normal evaluated at $(S,T)\phi$. Note that $a_4(Q)\equiv(2\pi\lambda^2)^{k/2}a_3(Q)$.

Step 3: (Integrate ϕ) For simplicity, I will assume throughout the remaining part of this section that $r(\beta_0)\geq 0$. Consider the mapping $m:[-\pi,\pi]\to\mathcal{S}^1$ given by $m(\theta)=[-\sin(\theta)\ ,\ \cos(\theta)\ ]$. Note that $m(\cdot)$ evaluated at $-\pi$ gives the point $(0,1)$ in the unit circle. As θ increases, the mapping $m(\cdot)$ traces $\mathcal{S}^1$ counter-clock wise. Therefore,

$$
\mathcal{S}^1(r(\beta_0))=\{\phi\in\mathcal{S}^1\mid r(\beta_0)\phi_1+\sqrt{1-r^2(\beta_0)}\phi_2\geq 0\}
$$

can be expressed as:

$$
\begin{aligned}
&\Big\{\theta\in[-\pi,\pi]\ :\ r(\beta_0)(-\sin(\theta))+\sqrt{1-r^2(\beta_0)}\cos(\theta)\geq 0\Big\}\\
&= \Big[\tan^{-1}\Big(\sqrt{1-r(\beta_0)^2}/r(\beta_0)\Big)-\pi\ ,\ \tan^{-1}\Big(\sqrt{1-r(\beta_0)^2}/r(\beta_0)\Big)\Big]\\
&= [\pi_0,\pi_0+\pi],\quad \text{where}\quad \pi_0\equiv\tan^{-1}\Big(\sqrt{1-r(\beta_0)^2}/r(\beta_0)\Big)-\pi
\end{aligned}
$$

and $\pi_0<0$ when $r(\beta_0)>0$. The parameter $\phi\sim\mathcal{U}(\mathcal{S}^1(r(\beta_0)))$ if and only if $\theta\sim\mathcal{U}[\pi_0,\pi_0+\pi]$. Define:

$$
f^*_{[\pi_l,\pi_u]}(S,T)\equiv a_4(Q)\frac{1}{\pi_l-\pi_u}\int_{\pi_l}^{\pi_u}\exp\Big(\frac{b^2}{2}\phi(\theta)'Q\phi(\theta)d\theta\Big)d\theta;\quad \phi(\theta)'=[-\sin(\theta),\cos(\theta)]
$$

The following Lemma is crucial for the derivation of the point and one-sided ECS tests in the IV model. Let

$$\zeta_{max} = \frac{1}{2}\left[(S'S + T'T) + \sqrt{(S'S - T'T)^2 + 4(S'T)^2}\right]$$

$$\zeta_{min} = \frac{1}{2}\left[(S'S + T'T) - \sqrt{(S'S - T'T)^2 + 4(S'T)^2}\right]$$

denote the maximum and minimum eigenvalues of the matrix $Q \equiv [S,T]'[S,T]$ and let

Lemma 2: . Let π_l, π_u belong to the interval $[\pi_0, \pi + \pi_0]$. Then

$$\begin{aligned} f^*_{[\pi_l,\pi_u]}(S,T) &= a_4(Q)\exp\left(\frac{b^2}{4}\left(\zeta_{\max} + \zeta_{\min}\right)\right)\frac{\pi}{\pi_u - \pi_l} I_0(\kappa(Q))\left[\Phi^{VM}_{[0,2\pi]}\left(2(\pi_u - \pi_0)|\kappa(Q), \mu(Q)\right)\right. \\ &\quad \left. -\Phi^{VM}_{[0,2\pi]}\left(2(\pi_l - \pi_0)|\kappa(Q), \mu(Q)\right)\right] \end{aligned}$$

where $\Phi^{VM}_{[0,2\pi]}$ is the Von-Mises distribution in Mardia and Jupp (2000), p. 36 with *mean direction parameter:*

$$\mu(Q) = 2(\widehat{\theta}_{max} - \pi_0) \geq 0$$

and *concentration parameter*

$$\kappa(Q) = \frac{b^2}{4}(\zeta_{max} - \zeta_{min}) \in [0, 2\pi]$$

$I_0(\cdot)$ is the modified Bessel function of the first kind, defined in Abramowitz and Stegun (1964), Section 9.6, p. 375.

Proof: Let $L \equiv S'S - \zeta_{min}$. Note that L is the Likelihood Ratio Statistic as defined in Andrews et al. (2006) p. 722. Define:

$$e_{max} \equiv \begin{cases} (L, S'T)'/\sqrt{L^2 + (S'T)^2} & \text{if} \quad r(\beta_0)L + \sqrt{1 - r(\beta_0)}S'T > 0 \\ -(L, S'T)'/\sqrt{L^2 + (S'T)^2} & \text{if} \quad r(\beta_0)L + \sqrt{1 - r(\beta_0)}S'T \leq 0 \end{cases}$$

Note that e_{max} is the maximum eigenvalue of the matrix Q adjusted to belong to the domain $\mathcal{S}^1(r(\beta_0))$. Define $\widehat{\theta} \in [\pi_0, \pi]$ implicitily by the following equation:

$$[-\sin(\widehat{\theta}), \cos(\widehat{\theta})]' = e_{max}$$

Therefore,

$$P \equiv \begin{pmatrix} -\sin(\widehat{\theta}) & \cos(\widehat{\theta}) \\ cos(\widehat{\theta}) & \sin(\widehat{\theta}) \end{pmatrix}$$

yields the spectral decomposition of the matrix Q; that is:

$$P \begin{pmatrix} \zeta_{max} & 0 \\ 0 & \zeta_{min} \end{pmatrix} P' = Q$$

. Note that for any $\theta \in [\pi_l, \pi_u]$:

$$P' \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} = \begin{pmatrix} \sin(\widehat{\theta})\sin(\theta) + \cos(\widehat{\theta})\cos(\theta) \\ -\cos(\widehat{\theta})\sin(\theta) + \sin(\widehat{\theta})\cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos(\widehat{\theta}_{max} - \theta) \\ \sin(\widehat{\theta}_{max} - \theta) \end{pmatrix}$$

Therefore:

$$\begin{aligned}
f^*_{[\pi_l,\pi_u]}(S,T) &= a_4(Q)\frac{1}{\pi_u - \pi_l}\int_{\pi_l}^{\pi_u} \exp\left(\frac{b^2}{2}\left[\zeta_{max}\cos^2(\widehat{\theta}_{\max} - \theta) + \zeta_{min}\sin^2(\widehat{\theta}_{\max} - \theta)\right]\right)d\theta \\
&= a_4(Q)\frac{1}{\pi_u - \pi_l}\int_{\widehat{\theta}_{\max}-\pi_u}^{\widehat{\theta}_{\max}-\pi_l} \exp\left(\frac{b^2}{2}\left[\zeta_{max}\cos^2(\theta) + \zeta_{min}\sin^2(\theta)\right]\right)d\theta \\
&\qquad \text{(where he have changed the integration variable)} \\
&= \exp\left(\frac{b^2}{2}\zeta_{min}\right)a_4(Q)\frac{1}{\pi_u - \pi_l}\int_{\widehat{\theta}_{\max}-\pi_u}^{\widehat{\theta}_{\max}-\pi_l} \exp\left(\frac{b^2}{2}\left[(\zeta_{max} - \zeta_{min})\cos^2(\theta)\right]\right)d\theta \\
&\qquad (\text{as} \sin^2(\theta) + \cos^2(\theta) = 1) \\
&= a_4(Q)\exp\left(\frac{b^2}{2}\zeta_{min}\right)\exp\left(\frac{b^2}{4}(\zeta_{max} - \zeta_{min})\right)\frac{1}{\pi_u - \pi_l} \\
&\quad \int_{\widehat{\theta}_{\max}-\pi_u}^{\widehat{\theta}_{\max}-\pi_l} \exp\left(\frac{b^2}{4}(\zeta_{max} - \zeta_{min})\cos(2\theta)\right)d\theta \\
&\qquad (\text{as } 2\cos^2(\theta) - 1 = \cos(2\theta)) \\
&= a_4(Q)\exp\left(\frac{b^2}{2}\zeta_{min}\right)\exp\left(\frac{b^2}{4}(\zeta_{max} - \zeta_{min})\right)\frac{1}{2(\pi_u - \pi_l)} \\
&= \int_{2(\widehat{\theta}_{\max}-\pi_u)}^{2(\widehat{\theta}_{\max}-\pi_l)} \exp\Big(\kappa(Q)\cos(\theta)\Big)d\theta \\
&\qquad \text{(where we have used the change of variable } \tilde{\theta} = 2\theta) \\
&\qquad (\kappa(Q) \equiv \frac{b^2}{4}(\zeta_{max} - \zeta_{min})) \\
&= a_4(Q)\exp\left(\frac{b^2}{4}(\zeta_{max} + \zeta_{\min})\right)\frac{1}{2(\pi_u - \pi_l)} \\
&\quad \int_{2(\pi_l-\pi_0)}^{2(\pi_u-\pi_0)} \exp\Big(\kappa(Q)\cos(\theta - 2(\widehat{\theta}_{max} - \pi_0)\Big)du \\
&\qquad \text{(where we have used the change of variable } u = 2(\widehat{\theta}_{max} - \pi_0) - \theta
\end{aligned}$$

Note that $\widehat{\theta}_{max} - \pi_0 \in [0, \pi]$. Therefore, $\mu(Q) \equiv 2(\widehat{\theta}_{max} - \pi_0) \in [0, 2\pi]$. Using the definition of the Von-Mises distribution (supported on $[0, 2\pi]$) in Mardia and Jupp (2000) it follows that:

$$\begin{aligned}
f^*_{[\pi_l,\pi_u]}(S,T) &= a_4(Q)\exp\left(\frac{b^2}{4}\Big(\zeta_{\max} + \zeta_{\min}\Big)\right)\frac{\pi}{\pi_u - \pi_l}I_0(\kappa(Q))\Big[\Phi^{VM}_{[0,2\pi]}\Big(2(\pi_u - \pi_0)|\kappa(Q), \mu(Q)\Big) \\
&\quad -\Phi^{VM}_{[0,2\pi]}\Big(2(\pi_l - \pi_0)|\kappa(Q), \mu(Q)\Big)\Big]
\end{aligned}$$

$\mu(Q)$ is the mean direction parameter, and $\kappa(Q)$ is the concentration parameter. $I_0(\cdot)$ is the modified

Bessel function of the first kind, defined in Abramowitz and Stegun (1964), Section 9.6, p. 375. *Q.E.D.*

Proof of Result 2: (ECS test for $\phi_1 = 0$) From Lemma 2 above it follows that the integrated likelihood for independent priors:

$$\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))) \quad \omega \sim \mathcal{U}(\mathcal{S}^k) \quad \rho \sim \sqrt{\lambda^2 \chi^2_k}$$

is given by:

$$f_1^*(S,T) = f^*_{[\pi_0, \pi_0+\pi]} = a_1 \exp(-\frac{1}{2}[S'S + T'T]) \exp\left(\frac{b^2}{4}\left(S'S + T'T\right)\right) I_0(\kappa(Q))$$

where a_1 is a non-negative constant, $b^2 = \lambda^2/(1+\lambda^2)$, and

$$\kappa(Q) = \frac{b^2}{4}\left((S'S - T'T)^2 + 4(S'T)^2\right)^{1/2}$$

Note that the boundary conditional likelihood for the model is given by:

$$f_{\mathrm{Bd}}(S,T) = a_2 \exp\left(-\frac{1}{2}S'S\right)$$

Both $f_1^*(S,T)$ and $f^*_{\mathrm{Bd}}(S,T)$ are separately continuous, so that Assumption R1 is verified. Note that:

$$z(S,T,p_1) = \frac{a_1}{a_2}\exp\left(-\frac{1}{2}T'T\right)\exp\left(\frac{b^2}{4}[S'S + T'T]\right) I_0(\kappa(Q))$$

The quantile function $c(T,\alpha)$ is continuous in T and, therefore, measurable. So that the ecs test rejects if

$$\frac{a_1}{a_2}\exp\left(-\frac{1}{2}T'T\right)\exp\left(\frac{b^2}{4}[S'S + T'T]\right) I_0(\kappa(Q)) \geq c(T,\alpha)$$

Which holds if and only if:

$$S'S - T'T + \frac{4(1+\lambda^2)}{\lambda^2}\ln\left[I_0\Big(\frac{\lambda^2}{4(1+\lambda^2)}\Big((S'S - T'T)^2 + 4(S'T)^2\Big)^{1/2}\Big)\right]$$

is larger than the critical value function $c^*(T,\alpha)$, defined as the $1-\alpha$ quantiles (conditional on T) of the expression above under the distribution $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$.

A.7. *Proof of Result* 2^*

Let:

$$\mathcal{S}^1(r(\beta_0))^- \equiv \{\phi \in \mathcal{S}^1(r(\beta_0)) \mid \phi_1 \leq 0\}$$

Define $\mathcal{S}^1(r(\beta_0))^+$ analogously. Under the assumption $r(\beta_0) \geq 0$:

$$[-\sin(\theta), \cos(\theta)] \in \mathcal{S}^1(r(\beta_0))^- \iff \theta \in [0, \pi_0 + \pi]$$

Consider the independent priors under the alternative:

$$\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))^+) \quad \omega \sim \mathcal{U}(\mathcal{S}^k) \quad \rho \sim \sqrt{\lambda^2 \chi^2_k}$$

and the independent priors under the null:

$$\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))^-) \quad \omega \sim \mathcal{U}(\mathcal{S}^k) \quad \rho \sim \sqrt{\lambda^2 \chi^2_k}$$

Result 2*: *The α-ECS test for the problem $\boldsymbol{H}_0 : \phi_1 \leq 0$ vs. $\boldsymbol{H}_1 : \phi_1 > 0$ in an over-identified IV model with a single endogenous regressors and the priors above rejects if the statistic*:

$$\begin{aligned} z(S,T;\tau) \quad &\equiv \quad \exp\left(\frac{b^2}{4}\Big(S'S - T'T\Big)\right) I_0(\kappa(Q)) \\ &\Big[\frac{\tau}{\pi_0}\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big) + \frac{1-\tau}{\pi_0+\pi}\Big(1-\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big)\Big)\Big] \end{aligned}$$

is smaller than the critical value function $c^*_{os}(T;\lambda^2,\alpha)$, defined as the $(1-\alpha)$ quantiles of $z(S,T;\tau)$ with $S \sim \mathcal{N}_k(\boldsymbol{0}, \mathbb{I}_k)$ and T fixed. The function $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero defined in Section 9.6, p. 375 of Abramowitz and Stegun (1964).

Proof of Result 2*: From Lemma 2 it follows that the integrated likelihood under the alternative is given by:

$$\begin{aligned} f^*_1(S,T) = f^*_{[\pi_0,0]} \quad &= \quad a_1 \exp(-\frac{1}{2}[S'S+T'T]) \exp\left(\frac{b^2}{4}\Big(S'S+T'T\Big)\right)\frac{\pi}{-\pi_0} I_0(\kappa(Q)) \\ &\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big) \end{aligned}$$

And the integrated likelihood under the null:

$$\begin{aligned} f^*_0(S,T) = f^*_{[0,\pi_0+\pi]} \quad &= \quad a_1 \exp(-\frac{1}{2}[S'S+T'T]) \exp\left(\frac{b^2}{4}\Big(S'S+T'T\Big)\right)\frac{\pi}{\pi_0+\pi} I_0(\kappa(Q)) \\ &\Big[1-\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big)\Big] \end{aligned}$$

Therefore:

$$[\tau f^*_1(S,T) - (1-\tau) f^*_0(S,T)]/f_{\text{Bd}}(S,T)$$

:

$$\begin{aligned} = \quad &\frac{a_1}{a_2}\exp(-\frac{1}{2}T'T)\exp\left(\frac{b^2}{4}\Big(S'S+T'T\Big)\right)\pi I_0(\kappa(Q)) \\ &\Big[-\frac{\tau}{\pi_0}\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big) - \frac{1-\tau}{\pi_0+\pi}\Big(1-\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big)\Big)\Big] \end{aligned}$$

Therefore, the ECS one-sided test rejects if:

$$\begin{aligned} z(S,T;\tau) \quad &\equiv \quad \exp\left(\frac{b^2}{4}\Big(S'S - T'T\Big)\right) I_0(\kappa(Q)) \\ &\Big[\frac{\tau}{\pi_0}\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big) + \frac{1-\tau}{\pi_0+\pi}\Big(1-\Phi^{VM}_{[0,2\pi]}\Big(-2\pi_0 \mid \kappa(Q), \mu(Q)\Big)\Big)\Big] \end{aligned}$$

is smaller than the $(1-\alpha)$ quantile of the distribution of the statistic $z(S,T;\tau)$ computed with T fixed and $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$.

A.8. *Lemma 3*

Let W be a $s \times s$, $s \in \mathbb{N}\backslash\{1\}$, symmetric matrix partitioned in the following blocks

$$W = \begin{pmatrix} W_1 & W_{12} \\ W_{21} & W_2 \end{pmatrix},$$

where W_1 is $s_1 \times s_1$ and W_2 is $s_2 \times s_2$. Let

$$R_{12} = W_1^{-1/2} W_{12} W_2^{-1/2} \quad (s_1 \times s_2 \text{ matrix}),$$

where $W_i^{-1/2}$ denotes the symmetric square root of W_i. Define $B \equiv \mathbb{I}_{s_2} - R_{12}' R_{12}$. Note that B is the Schur complement of $\mathbb{I}_{s_1}$ in the positive definite matrix:

$$\begin{pmatrix} \mathbb{I}_{s_2} & R_{12}' \\ R_{12} & \mathbb{I}_{s_1} \end{pmatrix}.$$

Consequently, there is a positive definite and symmetric matrix, $B^{-1/2}$, such that $B^{-1/2} B B^{-1/2} = \mathbb{I}_{s_2}$. Let

$$D = \begin{pmatrix} W_1^{-1/2} & \mathbf{0} \\ -B^{-1/2} R_{12}' W_1^{-1/2} & B^{-1/2} W_2^{-1/2} \end{pmatrix}.$$

Lemma 3: $DWD' = \mathbb{I}_s$

Proof: Note that

$$\begin{aligned} DW &= \begin{pmatrix} W_1^{-1/2} & \mathbf{0} \\ -B^{-1/2} R_{12}' W_1^{-1/2} & B^{-1/2} W_2^{-1/2} \end{pmatrix} \begin{pmatrix} W_1 & W_{12} \\ W_{21} & W_2 \end{pmatrix} \\ &= \begin{pmatrix} W_1^{1/2} & W_1^{-1/2} W_{12} \\ -B^{-1/2} R_{12}' W_1^{1/2} + B^{-1/2} W_2^{-1/2} W_{21} & -B^{-1/2} R_{12}' W_1^{-1/2} W_{12} + B^{-1/2} W_2^{1/2} \end{pmatrix} \end{aligned}$$

Therefore, DWD' equals

$$\begin{aligned} & \begin{pmatrix} W_1^{1/2} & W_1^{-1/2} W_{12} \\ -B^{-1/2} R_{12}' W_1^{1/2} + B^{-1/2} W_2^{-1/2} W_{21} & -B^{-1/2} R_{12}' W_1^{-1/2} W_{12} + B^{-1/2} W_2^{1/2} \end{pmatrix} \\ & \begin{pmatrix} W_1^{-1/2} & -W_1^{-1/2} R_{12} B^{-1/2} \\ \mathbf{0} & W_2^{-1/2} B^{-1/2} \end{pmatrix} \\ = & \begin{pmatrix} \mathbb{I}_K & -R_{12} B^{-1/2} + R_{12} B^{-1/2} \\ -B^{-1/2} R_{12}' + B^{-1/2} R_{12}' & \mathbf{0} + -B^{-1/2} R_{12}' R_{12} B^{-1/2} + B^{-1/2} B^{-1/2} \end{pmatrix} \\ = & \begin{pmatrix} \mathbb{I}_{s_1} & \mathbf{0} \\ \mathbf{0} & \mathbb{I}_{s_2} \end{pmatrix} \end{aligned}$$

Q.E.D.

A.9. *Proof of Result 3*

The result is established in two steps. First, I use a simple change of variables formula that simplifies the derivation of the integrated likelihood. Second, I derive the ECS test.

Step 1: Following the notation in Billingsley (1995) let

$$\Omega \equiv \mathbb{R}^{n+1},\ \mathcal{F} \equiv \mathcal{B}(\mathbb{R}^{n+1}),\ \Omega' \equiv \mathbb{R}^{2m},\ \mathcal{F}' \equiv \mathcal{B}(\mathbb{R}^{2m})$$

By assumption, the function $T \equiv C^* : \mathbb{R}^{n+1} \to \mathbb{R}^{2m}$ is measurable. The prior p_1 on $\mathbb{R}^{n+1}$ and the function C^* induce a probability measure over the measurable space $(\Omega', \mathcal{F}')$ in the usual way:

$$\mu T^{-1}(A') \equiv \mathbb{P}^*(A') \equiv \int_{\{x \in \mathbb{R}^{n+1} \mid C^*(x) \in A'\}} p_1(x) dx$$

Also, the measure $P^*(A')$ is $\mathcal{N}_{2m}(\mathbf{0}, \Omega(\theta_0))$, by assumption. Define $f : \mathbb{R}^{2m} \to \mathbb{R}$ by:

$$f(x) = c \exp\Big(- \frac{1}{2}[\gamma(\theta_0) - D(\theta_0)x]'[\gamma(\theta_0) - D(\theta_0)x]\Big).$$

Let μ denote the probability measure associated with p_1 in $\mathbb{R}^{n+1}$. Theorem 16.13 in Billingsley (1995) imply

$$\int_{\mathbb{R}^{n+1}} f(C^*(\theta^*, \delta))\mu(d(\theta^*, \delta)) = \int_{\mathbb{R}^{2m}} f(x)\mu T^{-1}(dx),$$

which by Theorem 16.11 in Billingsley (1995) and the definition of a density (with respect to lebesgue measure) yield:

$$\int_{\mathbb{R}^{n+1}} f^*(C^*(\theta^*, \delta))p_1(\theta^*, \delta)d\theta^* d\delta = \int_{\mathbb{R}^{2m}} f(x)\phi_{2m}(x, \mathbf{0}, \Omega(\theta_0))dx.$$

Note that the integrated likelihood

$$\begin{aligned} f_1^*(\gamma(\theta_0)) &= c \exp\Big(- \frac{1}{2}\gamma(\theta_0)'\gamma(\theta_0)\Big) \int_{\mathbb{R}^{2m}} \exp\Big(\gamma(\theta_0)'D(\theta_0)x\Big) \exp\Big(x'\Omega(\Theta_0)^{-1}x\Big)dx \\ &\quad \text{(where c is a non-negative constant)} \\ &= c_1 \exp\Big(- \frac{1}{2}\gamma(\theta_0)'\gamma(\theta_0)\Big) \exp\Big(\frac{1}{4}\gamma(\theta_0)'\gamma(\theta_0)\Big) \\ &\quad \text{(by definition of the moment generating function of a multivariate normal).} \end{aligned}$$

Since the boundary conditional likelihood for the GMM limiting experiment is given by

$$f_{\text{Bd}}(m(\theta_0) \mid d(\theta_0)) = c_2 \exp\Big(- \frac{1}{2}m(\theta_0)'m(\theta_0)\Big),$$

the ECS test rejects the null hypothesis if $m(\theta_0)'m(\theta_0) > c$

A.10. *Proof of Result 4*

The integrated likelihood

$$\begin{aligned} f_1^*(\gamma(\theta_0)) &= c\exp\Big(-\frac{1}{2}\gamma(\theta_0)'\gamma(\theta_0)\Big) \\ &\quad \int_{\mathbb{R}^{n+1}} \exp\Big(\gamma(\theta_0)'D(\theta_0)C^*(\theta^*,\delta)\Big)\exp\Big(-\frac{1}{2}C^*(\theta^*,\delta)'\Omega(\theta_0)^{-1}C^*(\theta^*,\delta)\Big) \\ &\quad p_1(\theta^*,\delta)d\theta^* d\delta \\ &\quad \text{(where c is a non-negative constant)} \end{aligned}$$

Since the boundary conditional lkelihood for the weakly identified GMM model is given by

$$c_2\exp\Big(-\frac{1}{2}m(\theta_0)'m(\theta_0)\Big).$$

Therefore, the ECS test statistic equals $\frac{c}{c_2}\exp\Big(-\frac{1}{2}d(\theta_0)'d(\theta_0)\Big)$ times

$$\int_{\mathbb{R}^{n+1}} \exp\Big(\gamma(\theta_0)'D(\theta_0)C^*(\theta^*,\delta)\Big)\exp\Big(-\frac{1}{2}C^*(\theta^*,\delta)'\Omega(\theta_0)^{-1}C^*(\theta^*,\delta)\Big)p_1(\theta^*,\delta)d\theta^* d\delta$$

which is well defined by assumption GMM1 and also separaterly continuous in $m(\theta_0)$ and $d(\theta_0)$. Since R2 holds, then Result 4 follows.

A.11. *Proof of Result 5*

Let $\xi = (S', T_1')'$ and

$$A \equiv \begin{pmatrix} C_0'/(C_0'\Omega C_0')^{1/2} \\ \Big(C_0^{\perp\prime}\Omega^{-1}C_0^{\perp}\Big)^{-1/2} C_0^{\perp\prime}\Omega^{-1} \end{pmatrix}.$$

By construction

$$H_1 a = \begin{pmatrix} m_1 \\ \vdots \\ m_r \end{pmatrix}.$$

The integrated likelihood

$$\begin{aligned} f_1^*(\xi) &= c\exp\Big(-\frac{1}{2}\xi'\xi\Big) \\ &\quad \int_{\mathbb{R}^r} \exp\Big(\xi' A m\Big)\exp\Big(-\frac{1}{2}m'A'Am\Big) \\ &\quad p_1(m)dm \\ &\quad \text{(where c is a non-negative constant)} \\ &= c\exp\Big(-\frac{1}{2}\xi'\xi\Big)\int_{\mathbb{R}^r}\exp\Big(\xi'Am\Big)\exp\Big(-m'\Omega^{-1}m\Big)dm \\ &= c\exp\Big(-\frac{1}{2}(S'S+T_1'T_1\Big)\exp\Big(\frac{1}{4}(S'S+T_1'T_1)\Big) \end{aligned}$$

Since the boundary conditional lkelihood for SVAR model is given by

$$c_2 \exp\Big(-\frac{1}{2}S'S\Big).$$

Therefore, the ECS test statistic equals

$$\frac{c}{c_2}\exp\Big(-\frac{1}{2}T_1'T_1\Big)\exp\Big(\frac{1}{4}(S'S+T_1'T_1)\Big)$$

Since R2 holds (see the argument in Result 1), then Result 5 follows.

A.12. *Proof of Result 6*

The distributional assumption SVAR yields a limiting experiment as defined by Müller (2011):

$$\text{(A.12.1)} \quad \Big(\widehat{A}\otimes\widehat{Q}^{-1/2}\Big)(1/\sqrt{T})\sum_{t=1}^{T}\eta_t\otimes Z_t \xrightarrow{d} \mathcal{N}_{rk}\Big(AH_1\otimes Q^{-1/2}a\ ,\ \mathbb{I}_r\otimes\mathbb{I}_k\Big)$$

where

$$A\equiv\begin{pmatrix} C_0'/(C_0'\Sigma C_0')^{1/2} \\ \Big(C_0^{\perp\prime}\Sigma^{-1}C_0^{\perp}\Big)^{-1/2}C_0^{\perp\prime}\Sigma^{-1}\end{pmatrix},\quad AH_1\in\mathbb{R}^r,\quad Q^{-1/2}a\in\mathbb{R}^k$$

Note that:

$$\text{(A.12.2)}\quad AH_1=\begin{pmatrix} C_0'H_1/(C_0'\Sigma C_0')^{1/2} \\ \Big(C_0^{\perp\prime}\Sigma^{-1}C_0^{\perp}\Big)^{-1/2}C_0^{\perp\prime}\Sigma^{-1}H_1\end{pmatrix}=\begin{pmatrix} (\kappa-\kappa_0)/(C_0'\Sigma C_0')^{1/2} \\ \Big(C_0^{\perp\prime}\Sigma^{-1}C_0^{\perp}\Big)^{-1/2}C_0^{\perp\prime}\Sigma^{-1}H_1\end{pmatrix}$$

Therefore, the null hypothesis $\kappa-\kappa_0$ holds if and only if the first element of the column vector AH_1 equals zero. The limiting experiment (A.12.1) admits the following re-parameterization. Let

$$\text{(A.12.3)}\quad \phi\equiv AH_1/||AH_1||,\quad \omega\equiv Q^{-1/2}a/||Q^{-1/2}a||,\quad \rho\equiv||AH_1||\,||Q^{-1/2}a||,$$

ϕ is an element of the $r-1$ sphere, $\mathcal{S}^{r-1}\equiv\{x\in\mathbb{R}^r\ :\ ||x||=1\}$; ω is an element of S^{k-1} and r is a non-negative scalar.

REMARK 12 The normalization $h_{11}=1$ imposes a restriction on ϕ. By construction, the full rank matrix A satisfies $A\Sigma A'=\mathbb{I}_r$, which implies $\Sigma A'A=\mathbb{I}_r$. Consequently, $e_1\Sigma A'AH_1=1$ where $e_1=(1,0,0,\ldots 0)\in\mathbb{R}^r$. Since $e_1\Sigma A'$ defines a hyperplane in $\mathbb{R}^r$, ϕ is restricted to be an element of the positive half-space associated to that hyperplane, as $e_1\Sigma A'\phi=1/||AH_1||\geq 0$. We use $\mathcal{S}_R^{r-1}$ to denote the intersection of the half-space and $\mathcal{S}^{r-1}$.

A canonical description of the SVAR testing problem is given by the following model. The sample space is R^r, with a typical element denoted by $(S',T_1',T_2'\ldots T_{r-1}')'$. The parameter space is given by $\mathbb{R}_+\times\mathcal{S}_R^{r-1}\times\mathcal{S}^{k-1}$, with typical element (ρ,ϕ,ω). The statistical model is given by:

$$(S',T_1',T_2'\ldots T_{r-1}')'\sim\mathcal{N}_{rk}\Big(\phi\otimes\rho\omega\ ,\ \mathbb{I}_r\otimes\mathbb{I}_k\Big)$$

and the hypothesis of interest is

$$\mathbf{H}_0 : \phi_1 = 0 \quad vs. \quad \mathbf{H}_1 : \phi_1 \neq 0$$

.

Note that $\phi_1 = 0$ implies

$$f(S', T_1', \ldots T_{r-1}'; \rho, \phi_0, \omega) = (2\pi)^{-k/2} \exp\Big(-\frac{1}{2} S'S\Big) (2\pi)^{-(r-1)k/2} \Pi_{s=1}^{r-1} \exp\Big(-\frac{1}{2}(T_s - \phi_{s+1}\rho\omega)'(T_s - \phi_{s+1}\rho\omega)\Big)$$

Hence, boundary sufficiency is verified with $g(S', T_1', \ldots T_{r-1}') = (2\pi)^{-k/2} \exp\Big(-\frac{1}{2} S'S\Big)$.

We now derive the ECS test associated with the following independent priors. Let $\lambda_{\mathcal{S}^{r-1}}$ denote the surface measure of the $(r-1)$ sphere $\mathcal{S}^{r-1}$ (see Stroock (1999), p. 83)

$$\text{(A.12.4)} \quad \rho\omega \sim \mathcal{N}_k\Big(\mathbf{0}, \lambda^2 \mathbb{I}_k\Big)$$

and

$$\text{(A.12.5)} \quad \phi \sim \mathcal{U}(\mathcal{S}_R^{r-1})$$

where $\mathcal{U}(\mathcal{S}_R^{r-1})$ denotes the uniform measure on the restricted $(r-1)$ sphere; this is, for any measurable subset $\mathcal{A}$ of $\mathcal{S}^{r-1}$, $\mathbb{P}(\mathcal{A}) = \lambda_{\mathcal{S}^{r-1}}(\mathcal{S}^{k-1})^{-1} \int_A \lambda_{\mathcal{S}^{r-1}}(d\phi)$. We derive the integrated likelihood in three parts. In Part I we separate the likelihood $f(S', T_1', \ldots T_{r-1}'; \rho, \phi, \omega)$ into three different components. In Part II we compute the partially integrated likelihood with respect to prior on $\rho\omega$. In Part III we present the ECS Test.

Part I: Let $T = (T_1', T_2', \ldots T_{r-1}')'$

$$\begin{aligned} -2\ln f(S', T_1', \ldots T_{r-1}'; \rho, \phi, \omega) - \text{cons} &= \left(\begin{pmatrix} S \\ T \end{pmatrix} - \phi \otimes \rho\omega\right)' \left(\begin{pmatrix} S \\ T \end{pmatrix} - \phi \otimes \rho\omega\right) \\ &= \begin{pmatrix} S \\ T \end{pmatrix}' \begin{pmatrix} S \\ T \end{pmatrix} - 2 \begin{pmatrix} S \\ T \end{pmatrix}' \phi \otimes \rho\omega + (\rho\omega)'(\rho\omega) \\ &= S'S + T'T - 2\phi'(S, T_1, \ldots T_{r-1})'\rho\omega + (\rho\omega)'(\rho\omega) \end{aligned}$$

Part II: Let $M = [S, T_1, \ldots T_{r-1}]'[S, T_1, \ldots T_{r-1}]$. Note that:

$$\begin{aligned} & \int_{\mathbb{R}^k} f(S', T_1', \ldots T_{r-1}'; \rho, x) \exp\Big(-\frac{1}{2\lambda^2} x'x\Big) dx \\ = \;& \exp\Big(-\frac{1}{2}\text{tr}(Q)\Big) \int_{\mathbb{R}^k} \exp\Big(\phi'(S, T_1, \ldots T_{r-1})'\rho\omega\Big) \exp\Big(-\frac{1}{2b^2}\Big) dx \\ & \qquad (\text{where} \quad b^2 \equiv \lambda^2/(1+\lambda^2)) \\ = \;& a_1(\pi, k, b^2) \exp\Big(-\frac{1}{2}\text{tr}(M)\Big) \exp\Big(\frac{b^2}{2}\phi' M \phi\Big) \end{aligned}$$

where the last inequality follows from the definition of the moment generating function for a multivariate normal random variable.

Part III: Therefore, the ECS rejects if:

$$\int_{\mathcal{S}_R^{r-1}} \exp\left(\frac{\lambda^2}{2(1+\lambda^2)}\phi' M \phi\right) d\lambda_{\mathcal{S}^{r-1}}(d\phi) > c(T;\lambda^2) \quad \text{(A.12.6)}$$

REFERENCES

Abramowitz, M. and I. Stegun (1964): *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55, Dover publications.

Aliprantis, C. and K. Border (2006): *Infinite dimensional analysis: a hitchhiker's guide*, Springer Verlag.

Anderson, T. and H. Rubin (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D., M. Moreira, and J. Stock (2006): "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression."," *Econometrica*, 74, 715–52.

Berger, J. (1985): *Statistical decision theory and Bayesian analysis*, Springer.

Billingsley, P. (1995): "Probability and measure," *Jown Wiley & Sons.*

Bound, J., D. Jaeger, and R. Baker (1995): "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American statistical association*, 90.

Chamberlain, G. (2007): "Decision theory applied to an instrumental variables model," *Econometrica*, 75, 609–652.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Admissible invariant similar tests for instrumental variables regression," *Econometric Theory*, 25, 806–818.

Elliott, G., U. Müller, and M. Watson (2012): "Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis," .

Ferguson, T. (1967): *Mathematical statistics: A decision theoretic approach*, vol. 7, Academic Press New York.

Hansen, L., J. Heaton, and A. Yaron (1996): "Finite-sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

Jansson, M. and M. Moreira (2006): "Optimal inference in regression models with nearly integrated regressors," *Econometrica: Journal of the Econometric Society*, 74, 681–714.

Kleibergen, F. (2007): "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics," *Journal of Econometrics*, 139, 181–216.

Le Cam, L. (1986): *Asymptotic methods in statistical decision theory*, Springer.

Lehmann, E. and J. Romano (2005): *Testing statistical hypotheses*, Springer Verlag.

Linnik, J. (1968): *Statistical problems with nuisance parameters*, vol. 20, American Mathematical Society.

Mardia, K. and P. Jupp (2000): *Directional statistics*, vol. 28, Wiley.

Mavroeidis, S., M. Plagborg-Moller, and J. a. Stock (2012): "Empirical evidence on inflation expectations in the new Keynesian Phillips curve," Working Paper.

Montiel Olea, J. and C. Pflueger (2012): "A Robust Test for Weak Instrumets," Working Paper.

Montiel Olea, J., J. Stock, and M. Watson (2012): "Inference in Structural VARs with External Instruments," *Working Paper.*

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

——— (2009): "Tests with correct size when instruments can be arbitrarily weak," *Journal of Econometrics*, 152, 131–140.

Moreira, M. and H. Moreira (2012): "Contributions to the Theory of Similar tests," Working Paper.

Müller, U. K. (2011): "Efficient tests under a weak convergence assumption," *Econometrica: Journal of the Econometric Society*, 79, 395–435.

Munkres, J. (2000): *Topology (Second Edition)*, Prentice Hall Inc.

Nelson, C. and R. Startz (1990): "The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one," *Journal of Business*, 63, 125–140.

Neyman, J. (1935): "Sur la vérification des hypothèses statistiques composées," *Bull. Soc. Math. France*, 346–366.

Rudin, W. (2005): "Functional analysis. International series in pure and applied mathematics," .

——— (2006): *Real and complex analysis*, Tata McGraw-Hill Education.

Staiger, D. and J. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557–586.

Stein, E. (2011): *Functional Analysis: Introduction to Further Topics in Analysis*, Princeton Univ Pr.

Stock, J. and M. Watson (1996): "Confidence sets in regressions with highly serially correlated regressors," *manuscript, Harvard University*.

Stock, J. and J. Wright (2000): "GMM with weak identification," *Econometrica*, 68, 1055–1096.

Stroock, D. (1999): *A concise introduction to the theory of integration*, Birkhauser.

Wald, A. (1950): "Statistical decision functions," *New York*, 660.