

# EFFICIENT CONDITIONALLY SIMILAR TESTS: FINITE-SAMPLE THEORY AND LARGE-SAMPLE APPLICATIONS

JOSÉ LUIS MONTIEL OLEA<sup>1</sup>

This paper presents a new class of tests for hypothesis testing problems with a *boundary-sufficient statistic*: the Efficient Conditionally Similar tests (ECS). The paper focuses on two-sided testing problems with nuisance parameters, but the theory here developed can be applied more generally.

The first part of this paper shows that the new testing procedures characterize two important finite-sample properties: *admissibility* and *similarity on the boundary*. Thus, this paper proves that ECS tests are admissible and similar; but more important, that every admissible, similar procedure is *essentially* an ECS test. The characterization result requires the boundary-sufficient statistic to be boundedly complete. An immediate finite-sample application is a new test for the coefficient of a single-endogenous regressor in a gaussian over-identified Instrumental Variables model.

The second part of this paper shows that certain weakly-identified *extremum problems* can be mapped into a *limiting statistical model* with a boundedly-complete, boundary-sufficient statistic. This observation broadens the scope of the finite-sample theory presented in the first part of the paper. The ECS tests for the limiting model evaluated at sample analogues are shown to be, in a certain sense, *asymptotically efficient similar*. An immediate large-sample application is a new test for a just-identified probit model with endogeneity.

KEYWORDS: Admissibility, Efficiency, Finite-Sample, Similarity, Test, Weak Identification.

## 1. INTRODUCTION

THIS PAPER STUDIES a two-sided hypothesis testing problem:

$$(1.1) \quad \mathbf{H}_0 : \pi = \pi_0 \quad \text{vs.} \quad \mathbf{H}_1 : \pi \neq \pi_0, \quad \pi \in \Pi \subseteq \mathbb{R}^{d_\pi},$$

with a *nuisance parameter*  $\beta \in B \subseteq \mathbb{R}^{d_\beta}$ . The main assumption of this paper is the existence of a parametric statistical model,  $f(x; \beta, \pi)$ , with a *boundary-sufficient* statistic; this is, there is a partition of the data  $x = (x_1, x_2) \in \mathbf{X} \subseteq \mathbb{R}^{d_s}$  such that:

$$(1.2) \quad f_{\text{Bd}}(x_1|x_2) \equiv f(x_1|x_2; \beta, \pi_0) = f(x_1|x_2; \beta', \pi_0), \quad \forall \beta, \beta' \in B.$$

Testing problems with a boundary-sufficient statistic include: testing for the coefficient of a single-endogenous regressor in an Instrumental Variables model with gaussian innovations [Andrews, Moreira, and Stock (2006, 2007, 2008)]; testing in a Linear Regression Model with a sign restriction on the coefficient of a control variable [Elliott, Müller, and Watson (2013)]; testing in a Linear Regression Model with a persistent regressor [Jansson and Moreira

---

<sup>1</sup>A previous version of this paper was circulated under the title “Efficient Conditionally Similar-on-the-Boundary Tests”. I am deeply indebted to my main advisors, Matías Cattaneo, Gary Chamberlain, Tomasz Strzalecki, and James Stock, for their continuous guidance, support, patience, and encouragement. I owe special thanks to Isaiah Andrews, Denis Chetverikov, Yuhta Ishii, Michal Kolesar, Timothy McQuade, Mikkel Plagborg-Møller, Daniel Pollmann, and Guillaume Pouliot for helpful comments and suggestions. I would also like to thank seminar audiences at Banco de México, Brown, Chicago Booth (Econometrics), Cowles Summer Conference, Davis ARE, Duke, Imperial College London-Business School, ITAM, University of Michigan, NYU Economics, and TSE. All errors are my own. E-mail: montiel.olea@nyu.edu. This version: September 8, 2013.

(2006)]; testing for the ratio of means in a bivariate gaussian model [Fieller (1954)]. In addition, this paper shows that the weakly-identified *extremum problems* of Andrews and Cheng (2012) and the weakly-identified Generalized Method of Moments models introduced by Stock and Wright (2000) can be mapped to a *limiting statistical model* satisfying (1.2).

The first contribution of this paper is a new class of tests: the Efficient Conditionally Similar Tests (ECS). ECS tests are *constrained* maximizers of Weighted Average Power (WAP) with user-specified weights  $w(\beta, \pi)$ . An  $\alpha$ -level ECS test rejects the null hypothesis whenever:

$$(1.3) \quad z(x_1, x_2) \equiv \left( f_w^*(x_1, x_2) / f_{\text{Bd}}(x_1 | x_2) \right) > c(x_2; \alpha).$$

The function  $f_w^*(x_1, x_2)$  denotes the *integrated likelihood* based on the statistical model  $f(x_1, x_2; \beta, \pi)$  and weights  $w(\beta, \pi)$ . The *critical value function*  $c(x_2; \alpha)$  is defined as the (conditional)  $1-\alpha$ -quantile of  $z(X_1, x_2)$ , with  $X_1 \sim f_{\text{Bd}}(x_1 | x_2)$ .

This paper shows that the ECS tests essentially characterize two classical finite-sample properties: admissibility and similarity. *Admissibility*, on the one hand, is a minimum optimality requirement for testing procedures: there is no other test with smaller rates of Type I and Type II error. *Similarity*, on the other hand, is a stringent size-control condition: the rate of Type I error is constrained to be invariant with respect to the value of nuisance parameters. The characterization result states that ECS tests are admissible and similar; but more important, that every admissible, similar procedure is *risk* equivalent to an *extended* ECS test. The characterization result assumes that the boundary-sufficient statistic is *boundedly-complete*, as defined by Lehmann and Romano (2005). An immediate application of the finite-sample theory in this paper is a new test for the over-identified gaussian Instrumental Variables (IV) model with homoskedastic errors, non-stochastic instruments, and known reduced-form covariance matrix. The details are presented in Appendix C. Appendix D derives an ECS test for the Linear Regression Model with a sign restriction studied by Elliott et al. (2013).

The second contribution of this paper is an application of ECS tests to the class of extremum problems recently studied by Andrews and Cheng (2012)—henceforth AC12. This large-sample application broadens the scope of the finite-sample theory presented in the paper. Theorem 1 in Müller (2011) is used to show that *sample analogues* of the ECS tests defined for a *limiting statistical model* are, in a certain sense, *asymptotically efficient similar* under local-to-unidentification asymptotics.<sup>1</sup> Thus, the finite-sample optimality properties of the ECS tests have interesting large-sample consequences. An immediate application of the framework developed in this section is a new test for a just-identified probit model with endogeneity. The details are presented in Appendix E. Appendix F derives the limiting experiment for the nonlinear regression model estimated by Nonlinear Least Squares studied in AC12. Appendix G applies the ECS tests to the weakly identified Generalized Method of Moments (GMM) models of Stock and Wright (2000).

The remainder of this paper is organized as follows. Section 2 presents the finite-sample theory behind the ECS tests. The section concludes with a new test for the gaussian IV model. Section 3 develops a framework that allows the application of ECS tests to the extremum problems considered by AC12. The section introduces a new statistic, the  $\hat{\beta}$ -score-star, and presents a new test for the probit model with endogeneity. Section 4 concludes. Appendix A

---

<sup>1</sup>I use the phrase *limiting statistical model* or *limiting experiment* in the modern sense of Müller (2011) and not in the classical sense of Le Cam (1986). Thus, a limiting statistical model will refer to a statistical model derived from a set of weak convergence assumptions. Section 3.5 illustrates this idea.

collects the proofs concerning the finite-sample theory developed in this paper. Appendix B collects the proofs concerning the general application of the ECS tests to extremum problems in AC12. Appendices C to H are Supplementary Materials.

## 2. FINITE-SAMPLE THEORY

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

### 2.1. Notation Preliminaries

Let  $\mathcal{B}(\mathbb{R}^n)$  denote the Borel  $\sigma$ -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  $\sigma$ -algebra. *Measurability* of the function  $f : \mathcal{S} \rightarrow \mathbb{R}$  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  $\mu$  is denoted  $\int_{\mathcal{S}} f(s) d\mu(s)$  or  $\int_{\mathcal{S}} f d\mu$  if no ambiguity arises. All vectors are column vectors. For notational convenience,  $(a, b)$  will sometimes replace  $(a', b)'$ . The dimension of the column vector “a” is denoted  $d_a$ .

### 2.2. Basic Elements of a Parametric Testing Problem

**SAMPLE SPACE, PARAMETER SPACE, AND STATISTICAL MODEL:** The finite-sample parametric testing problems studied in this paper have the following components. There is a random vector  $X$  that takes values in the *sample space*  $\mathbf{X} \subseteq \mathbb{R}^s$ . There is a *parameter space*  $\Theta = B \times \Pi \subseteq \mathbb{R}^{d_{\beta} \times d_{\pi}} = \mathbb{R}^p$  whose elements  $\theta = (\beta, \pi) \in \Theta$  are used to index a set of probability density functions (w.r.t. to Lebesgue measure in  $\mathbb{R}^s$ ) over the sample space,  $X \sim f(x, \theta)$ . The collection  $\{f(\cdot, \beta, \pi)\}_{(\beta, \pi) \in \Theta}$  is called a *statistical model*. The mapping  $f : \mathbf{X} \times \Theta \rightarrow \mathbb{R}_+$  is called the *likelihood function*.

**NULL HYPOTHESIS:** The null hypothesis  $\mathbf{H}_0$  states  $X \sim f(x, \beta, \pi_0)$  for some  $\beta \in B$ . The alternative hypothesis  $\mathbf{H}_1$  states  $X \sim f(x, \beta, \pi)$  for  $\pi \neq \pi_0$ ,  $\pi \in \Pi$ . The null set  $\Theta_0$  is the set of parameters  $(\beta, \pi)$  that satisfy  $\mathbf{H}_0$ . The alternative set  $\Theta_1$  is defined as  $\Theta \setminus \Theta_0$ . The testing problem is abbreviated  $\mathbf{H}_0 : \theta \in \Theta_0$  vs.  $\mathbf{H}_1 : \theta \in \Theta_1$ . The set  $\text{Bd}\Theta_0$  will denote the (topological) boundary of the set  $\Theta_0$ .<sup>2</sup> In the two-sided testing problem (1.1), the set  $\text{Bd}\Theta_0$  coincides with  $\Theta_0$ .

**REMARK 1:** Appendix A shows that the main results in this paper are also applicable to general testing problems with a closed composite null and an open composite alternative, for example one-sided problems with a nuisance parameter. Thus, the results in the paper allow for models in which  $\text{Bd}\Theta_0 \subset \Theta_0$ . Appendix D presents an ECS test for a one-sided hypothesis in the linear regression model with a sign restriction studied in Elliott et al. (2013).

<sup>2</sup>Let  $\mathcal{T}$  be the subspace topology on  $\Theta \subseteq \mathbb{R}^p$  and let  $\tau_{\theta}$  denote an open neighborhood of  $\theta \in \Theta$ ; i.e.,  $\theta \in \tau_{\theta}$  and  $\tau_{\theta} \in \mathcal{T}$ . Define

$$\text{Bd}\Theta_0 \equiv \{\theta \in \Theta \mid \tau_{\theta} \cap \Theta_0 \neq \emptyset \text{ and } \tau_{\theta} \cap \Theta_1 \neq \emptyset, \forall \tau_{\theta} \in \mathcal{T}\}.$$

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

### 2.3. Main Finite-Sample Assumptions

Assumption F0, F1, F2 below restrict the class of statistical models under consideration. The first assumption imposes a minimum regularity condition. The second assumption is the key requirement of this paper: the existence of a boundary-sufficient statistic. The third assumption imposes a constraint on the boundary-sufficient statistic: bounded completeness.

**ASSUMPTION F0** (CONTINUITY): The statistical model  $f(x, \theta)$  is:

- i) continuous in  $\theta$  for almost every  $x \in \mathbf{X}$ ,
- ii) continuous in  $x$  for almost every  $\theta \in \Theta$ .<sup>3</sup>

**ASSUMPTION F1** (BOUNDARY SUFFICIENCY): There is a partition of the data  $X = (x_1, x_2)$  such that:

$$(2.1) \quad f_{\text{Bd}}(x_1|x_2) \equiv f(x_1|x_2; \theta) = f(x_1|x_2; \theta') \quad \forall \theta \in \text{Bd}\Theta_0.$$

REMARK 2: In the two sided testing problem (1.1) with a nuisance parameter, the set  $\text{Bd}\Theta_0$  coincides with  $\Theta_0$ . Therefore the notions of *boundary* sufficiency (2.1) and *null* sufficiency (1.2) are equivalent.

MAIN OBJECTIVE OF THIS PAPER: It is well known that a boundary-sufficient statistic can be used to control the *null* rejection probability of a statistical test in a two-sided problem with a nuisance parameter [Ferguson (1967), Moreira (2003), Andrews et al. (2006), Lehmann and Romano (2005)]. This paper is concerned with the question of how to generate “good” statistical tests in the presence of Assumption F1.

The following property plays a central role in the characterization result presented in this paper:

**ASSUMPTION F2:** (BOUNDED COMPLETENESS) [Lehmann and Romano (2005) p. 115] Let  $m : \mathbf{X}_2 \rightarrow \mathbb{R}$  be an arbitrary bounded measurable function. Let  $h(x_2, \theta)$  denote the marginal density of  $x_2$  based on  $f(x_1, x_2, \beta, \pi)$ . The boundary-sufficient statistic is boundedly-complete if

$$\int m(X_2)h(x_2, \theta)dx_2 = 0, \quad \forall \theta \in \text{Bd}\Theta_0 \implies m(X_2) = 0.$$

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

Theorem 4.3.1 in Lehmann and Romano (2005) provides a sufficient condition to guarantee that a family of distributions is complete, and thus, boundedly complete. In the examples studied in this paper, it will be sufficient to show that the set  $\text{Bd}\Theta_0$  contains a rectangle of the same dimension as the boundary-sufficient statistic.

Bounded completeness will be used to show that all similar-on-the-boundary tests must be “conditionally” (on the boundary-sufficient statistic) similar. This is a well-known result

---

<sup>3</sup>“Almost-every” is relative to the restriction of the Lebesgue measure in  $\mathbb{R}^s$  (or  $\mathbb{R}^p$ ) to  $\mathbf{X}$  (or  $\Theta$ )

in the theory of statistical hypothesis testing. See Theorem 4.3.2 in [Lehmann and Romano \(2005\)](#).

#### 2.4. Tests, Type I, and Type II error

TESTS: A *test* is a measurable mapping  $\phi : \mathbf{X} \rightarrow [0, 1]$ . The scalar  $\phi(x)$  is interpreted as the probability of rejecting  $\mathbf{H}_0$  after a realization  $x$  of  $X$ . Let  $\mathcal{C}$  denote the class of all tests.

In a finite-sample set-up, tests are usually compared on the basis of the Type I and Type II error functions:

TYPE I ERROR: The *rate of Type I error* of test  $\phi$  at  $\theta \in \Theta_0$  is defined as

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

Type I error refers to the probability of rejecting the null hypothesis when the true parameter belongs to the null set.

TYPE II ERROR The *rate of Type II error* of  $\phi$  at  $\theta \in \Theta_1$  is defined as

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

Type II error refers to the probability of not rejecting the null hypothesis when the null is not true.

RISK FUNCTION: Type I and Type II error are summarized by the risk function, defined as

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

Two tests  $\phi, \phi'$  are *risk equivalent* if  $R(\phi, \theta) = R(\phi', \theta)$  for all  $\theta \in \Theta$ .

#### 2.5. Admissibility, Similarity, and ECS Tests

This paper proposes a new class of tests for the hypothesis testing problem (1.1) in statistical models with a boundary-sufficient statistic: ECS tests. Broadly speaking, the appeal of the ECS tests is that they are (essentially) the only way to generate tests that satisfy the following properties:

**ADMISSIBILITY:** ([Ferguson \(1967\)](#), p. 54) The test  $\phi$  is *admissible* within the class  $\mathcal{C}^* \subseteq \mathcal{C}$  if there is no  $\phi' \in \mathcal{C}^*$  such that  $R(\phi', \theta) \leq R(\phi, \theta)$  for all  $\theta \in \Theta$ , with strict inequality for at least one  $\theta \in \Theta$ .<sup>4</sup>

---

<sup>4</sup>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  $\phi'$  *weakly dominates*  $\phi$  if  $R(\phi', \theta) \leq R(\phi, \theta)$  with strict inequality for at least one  $\theta \in \Theta$ .

**SIMILARITY ON  $\text{Bd}\Theta_0$ :** A test  $\phi$  is  $\alpha$ -similar on the boundary ( $\alpha$ -sb) if:

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

**ON ADMISSIBILITY:** *Admissibility* was first introduced by Wald (1950) and it is a well-known concept in mathematical statistics. Tests that violate admissibility 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 seems a reasonable minimal requirement that a test must satisfy.

**ON SIMILARITY:** *Similarity* was first introduced by Neyman (1935) and it has been extensively studied by Linnik (1968).<sup>5</sup> Note that in the two-sided problem (1.1)  $\alpha$ -sb is equivalent to invariance of the rate of Type I error with respect to the nuisance parameter  $\beta$ :

$$\mathbb{E}_{(\beta, \pi_0)}[\phi(X)] = \alpha, \quad \forall \beta \in B.$$

Therefore, in two-sided problems with a nuisance parameter similarity on the boundary provides a stringent size-control condition. In more general testing problems, similarity-on-the-boundary is usually motivated as a necessary condition for *unbiasedness*; see, for example, Andrews (2012).

**ADMISSIBLE SIMILAR-ON-THE-BOUNDARY TESTS:** Admissibility and similarity are classical concepts in statistical decision theory. However, to the best of my knowledge, there are no general results concerning the construction of admissible similar-on-the-boundary tests in the presence of a boundary-sufficient statistic. This paper tries to fill this gap in the literature.

**ECS TESTS:** Let  $w(\beta, \pi)$  be probability measures over  $B \times \Pi$ . The  $\alpha$ -level ( $w$ )-Efficient Conditionally Similar Test for the two-sided problem (1.1) with a boundary-sufficient statistic  $x_2$  is given by:

$$(2.2) \quad \phi_{\text{ECS}}(x_1, x_2) = \begin{cases} 1 & \text{if } z(x_1, x_2) \equiv f_w^*(x_1, x_2)/f_{\text{Bd}}(x_1|x_2) > c(x_2; \alpha, w) \\ 0 & \text{i.o.c.} \end{cases},$$

with

$$(2.3) \quad f_w^*(x_1, x_2) \equiv \int_{B \times \Pi} f(x_1, x_2; \beta, \pi) dw(\beta, \pi).$$

and

$$(2.4) \quad c(x_2; \alpha) \equiv \arg \min_{q \in \mathbb{R}} \int \rho_{(1-\alpha)}(z(x_1, x_2) - q) f_{\text{Bd}}(x_1|x_2) dx_1,$$

where  $\rho_{1-\alpha}(u) = u[(1-\alpha) - \mathbb{1}\{u < 0\}]$  denotes the check function.

---

<sup>5</sup>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 (*ensemble critique d'aire ‘ $\alpha$ ’ bien déterminée par l’hypothèse à vérifier*). Linnik (1968) refers to such regions as  $\alpha$ -similar regions.

REMARK 3: For each  $x_2$ ,  $c(x_2; \alpha)$  corresponds to the conditional  $(1-\alpha)$  quantile of the random variable  $z(X_1, x_2)$ , with  $X_1 \sim f_{\text{Bd}}(x_1|x_2)$ .

ON THE USER-SELECTED WEIGHT: ECS tests are indexed by a user-specified weight function  $w(\beta, \pi)$ .<sup>6</sup> Lemma 3 in Appendix A.4 shows that an  $\alpha$  level,  $(w)$ -ECS test for the two-sided problem (1.1) maximizes Weighted Average Power (WAP) inside the class of  $\alpha$ -sb tests, provided Assumption F2 holds. Thus, the weight function  $w(\beta, \pi)$  represents, in a way, the part of the parameter space for which an ECS test has good power properties. This is particularly relevant in problems that do not admit a Uniformly Most Powerful (UMP) test.

CHARACTERIZATION RESULT: The main contribution of this paper to the theory of similar-on-the-boundary tests is the characterization result in Theorem 1. Part i) shows that ECS test are admissible and similar-on-the-boundary. Part ii) shows that every admissible similar-on-the boundary test is *essentially* an ECS test. To formalize the notion of “essentially” this paper introduces the following definition:

**EXTENDED ECS TESTS:** A similar-on-the-boundary test  $\phi$  is an extended ECS if  $\forall \epsilon > 0$  there exists a borel probability measure  $w_\epsilon(\beta, \pi)$  supported on a non-empty subset of  $\Theta \setminus \text{Bd}\Theta_0$  such that:

$$\int_{\Theta \setminus \text{Bd}\Theta_0} R(\phi, \theta) dw_\epsilon(\theta) \leq \int_{\Theta \setminus \text{Bd}\Theta_0} R(\phi_{\text{ECS}}^{w_\epsilon}, \theta) dw_\epsilon(\theta) + \epsilon$$

REMARK 4: Note that a similar test  $\phi$  for the two-sided problem (1.1) is an extended ECS if and only if  $\forall \epsilon > 0$  there exists a borel probability measure  $w_\epsilon(\beta, \pi)$  supported on a non-empty subset of  $\Theta_1$  such that:

$$\text{WAP}(\phi, w_\epsilon) \equiv \int_{\Theta_1} \left( \int_{\mathbf{X}} \phi(x) f(x, \theta) dx \right) dw_\epsilon(\theta) \geq \text{WAP}(\phi_{\text{ECS}}^{w_\epsilon}, w_\epsilon) - \epsilon$$

Extended ECS tests are *essentially* ECS tests: for any  $\epsilon > 0$  there is an ECS test for which the WAP of an extended ECS is at most  $\epsilon$  below the WAP achieved by an ECS test. Appendix A shows that extended ECS tests are in fact extended *constrained* Bayes procedures, as defined in [Ferguson \(1967\)](#) pg. 50, Definition 3.

**THEOREM 1:** *Suppose Assumption F0, F1, F2 hold and suppose that the critical value function  $c(\cdot; \alpha) : \mathbf{X}_2 \rightarrow \mathbb{R}$  is measurable. Then,*

- i)  $\phi_{\text{ECS}}$  is admissible and similar on the boundary, provided  $w(\beta, \pi)$  has full-support on  $\Theta_1$ .
- ii) Every admissible and similar-on-the-boundary test is risk equivalent to an extended ECS test.

PROOF: See Appendix A.4

*Q.E.D.*

COMMENT ON PART i): The proof of part i) uses the Neyman Pearson Lemma and Assumption F2 to show that ECS tests maximize WAP subject to a similarity constraint. The

---

<sup>6</sup>For notational convenience, this paper uses the notation  $\phi_{\text{ECS}}(x_1, x_2)$  instead of  $\phi_{\text{ECS}}(x_1, x_2; w)$

admissibility of ECS tests inside the class of similar-on-the-boundary procedures follows directly from the WAP maximization property, provided the weight function has full-support. Assumption F0 is then used to show that the continuity of the risk function implies that a constrained WAP maximizer cannot be dominated by a non-similar procedure.

COMMENT ON PART ii): The proof of ii) is the most interesting part of the characterization result. The proof is based on an *essentially complete class theorem* [see Theorem 2.9.2 and 2.10.3 in Ferguson (1967) and also Le Cam (1986), Chapter 2, Theorem 1]. An important insight of this paper is that the set of  $\alpha$ -similar tests is compact in the weak\* topology (see Lemma 1 in Appendix A). Since the risk function of the testing problem (Type I and Type II error) is continuous with respect to the same topology, the essentially complete class theorem applies: this is, the set of extended Bayes tests in  $C(\alpha\text{-sb})$  is an essentially complete class (see Ferguson (1967), pg. 55, Definition 3). The final step of the proof is to show that extended Bayes tests in  $C(\alpha\text{-sb})$  are extended ECS tests.

GENERALIZATIONS OF THEOREM 1: Appendix A.4 presents a proof of Theorem 1 in the context of a general hypothesis testing problem  $(\mathbf{X}, \Theta, f, \Theta_0)$ , with a closed null set and a boundary-sufficient statistic  $x_2$ . In this general set-up the ECS tests are generalized as follows. Given probability measures  $w_i$  over  $\text{int}\Theta_i$  ( $i = 0, 1$ ) define  $\phi_{\text{ECS}}(x_1, x_2)$  as:

$$\begin{cases} 1 & \text{if } z(x_1, x_2) \equiv \left[ \tau f_{w_1}^*(x_1, x_2) - (1 - \tau) f_{w_0}^*(x_1, x_2) \right] / f_{\text{Bd}}(x_1 | x_2) > c(x_2; \alpha, w) \\ 0 & \text{i.o.c.} \end{cases}$$

with

$$f_{w_i}^*(x_1, x_2) \equiv \int_{\text{int}(\Theta_i)} f(x_1, x_2; \theta) dw_i(\theta), \quad i = 0, 1$$

and

$$c(x_2; \alpha) \equiv \arg \min_{q \in \mathbb{R}} \int \rho_{(1-\alpha)}(z(x_1, x_2) - q) f_{\text{Bd}}(x_1 | x_2) dx_1,$$

where  $\rho_{1-\alpha}(u) = u[(1-\alpha) - \mathbb{1}\{u < 0\}]$  denotes the check function. Appendix A shows that the generalized ECS tests minimize *average risk* subject to a similarity-on-the-boundary constraint.

## 2.6. Example: Gaussian IV

The over-identified linear IV model with a single right-hand endogenous regressor in Andrews et al. (2006) is used to illustrate the finite-sample theory of this paper.<sup>7</sup> This section proposes a new ECS test for the over-identified Gaussian IV model and presents a power comparison against the Conditional Likelihood Ratio (CLR) test, the Anderson-Rubin (AR) test, and the Lagrange Multiplier (LM) test. By Theorem 1 none of these competitor tests can have greater or equal power than an ECS test all over the parameter space.

---

<sup>7</sup>The model has independent observations; homoskedastic, gaussian reduced-form errors; non-stochastic instruments; and known reduced-form covariance matrix, denoted  $\Omega$ .



GAUSSIAN IV MODEL: Let  $\pi \in \mathbb{R}$  denote the coefficient of the single endogenous regressor and let  $\beta \in \mathbb{R}^k$  denote the unknown vector of first-stage coefficients.<sup>8</sup> Let  $Z \in \mathbb{R}^{n \times k}$  denote the non-stochastic matrix of instruments and let  $\Omega \in \mathbb{R}^{2 \times 2}$  denote the known reduced-form covariance matrix. Consider the statistical model:

$$(2.5) \quad \begin{pmatrix} S \\ T \end{pmatrix} \sim \mathcal{N}_{2k} \left( \begin{pmatrix} (b_0' \Omega b_0)^{-1/2} (\pi - \pi_0) \\ (a_0' \Omega^{-1} a_0)^{-1/2} a_0' \Omega^{-1} a \end{pmatrix} \otimes (Z' Z)^{1/2} \beta, \mathbb{I}_{2k} \right)$$

where  $a = [\pi, 1]'$ ,  $a_0 = [\pi_0, 1]'$ ,  $b_0 = [1, -\pi_0]'$  and  $(S, T)$  is a one-to-one transformation of the OLS estimators of the reduced-form parameters. This statistical model is derived in [Andrews et al. \(2006\)](#) pg. 719 and Lemma 2 pg. 720.

SAMPLE SPACE, PARAMETER SPACE, AND ASSUMPTIONS F IN THE IV MODEL: The sample space is  $\mathbb{R}^{2k}$  and the parameter space is  $\mathbb{R}^k \times \mathbb{R}$ , with elements  $(\beta, \pi)$ . The testing problem of interest is (1.1). The first-stage coefficient,  $\beta$ , is the nuisance parameter in the model. The statistical model (2.5) satisfies Assumption F0. Assumption F1 and F2 are easily verified: [Andrews et al. \(2006\)](#) have shown that  $T$  is a boundedly-complete, boundary-sufficient statistic.

CANONICAL WEIGHTS FOR THE IV MODEL: The *canonical* parameterization of the over-identified Gaussian IV model proposed in [Chamberlain \(2007\)](#) is used to motivate weights for the ECS test. The original parameters  $(\beta, \pi)$  induce the following canonical parameters  $(\rho, \phi, \omega)$ :

$$\rho = (a' \Omega^{-1} a)^{1/2} (\beta' Z' Z \beta)^{1/2}, \quad \phi = C_0 a / (a' \Omega^{-1} a)^{1/2}, \quad \omega = (Z' Z)^{1/2} \beta / (\beta' Z' Z \beta)^{1/2},$$

where  $C_0$  is the  $2 \times 2$  matrix defined in [Andrews et al. \(2006\)](#) with first row equal to  $(b_0' \Omega b_0)^{-1/2} b_0'$  and second row given by  $(a_0' \Omega^{-1} a_0)^{-1/2} a_0' \Omega^{-1}$ .<sup>9</sup> The *canonical parameter space* is:

$$\rho \in \mathbb{R}_+, \quad \phi \in \mathcal{S}^1(r(\pi_0)), \quad \omega \in \mathcal{S}^{k-1},$$

where  $\mathcal{S}^m$  is the  $m$  unit sphere.<sup>10</sup> The set  $\mathcal{S}^1(r(\pi_0))$  is a subset of  $\mathcal{S}^1$  defined as:

$$\mathcal{S}^1(r(\pi_0)) \equiv \{(\phi_1, \phi_2) \in \mathcal{S}^1 \mid r(\pi_0)\phi_1 + \sqrt{1 - r^2(\pi_0)}\phi_2 \geq 0\},$$

and  $r(\pi_0)$  equal to the correlation coefficient of the structural covariance matrix  $\Sigma_0$  associated to  $\pi_0$ . Consider the following weights for the canonical parameters. Assume  $(\rho, \phi, \omega)$  are independent. Let  $\phi, \omega$  have uniform distributions on their domains:

$$(2.6) \quad \phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))), \quad \omega \sim \mathcal{U}(\mathcal{S}^{k-1}).$$

<sup>8</sup>The (unorthodox) notation “ $\pi$ ” for the coefficient of the endogenous regressor and “ $\beta$ ” for the first-stage coefficient follows the recent treatment of [Andrews and Cheng \(2012\)](#).

<sup>9</sup>Under the canonical parameterization:

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

<sup>10</sup>That is,  $\mathcal{S}^m = \{x \in \mathbb{R}^{m+1} \mid \|x\| = 1\}$ , for any  $m \in \mathbb{N}$

and consider the following weight over the parameter  $\rho$  :

$$(2.7) \quad \rho \sim \sqrt{\chi_k^2}.$$

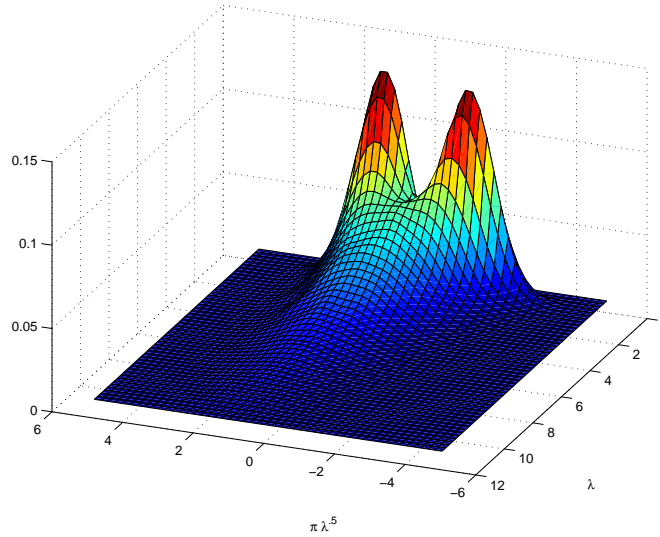
REMARK 5: Since the canonical parameterization is a one-to-one transformation of  $(\beta, \pi)$ , the weights over  $(\rho, \phi, \omega)$  induce weights over any function of  $(\beta, \pi)$  through the inverse mapping:

$$(2.8) \quad \pi - \pi_0 = \left[ (b_0' \Omega b_0)^{1/2} \phi_1 \right] / \left[ ([0, 1] \Omega [0, 1]')^{1/2} \left( r(\pi_0) \phi_1 + \sqrt{1 - r^2(\pi_0)} \phi_2 \right) \right]$$

$$(2.9) \quad (Z'Z)^{1/2} \beta = \rho \omega / \left[ (\pi - \pi_0, 1) \Sigma_0^{-1} (\pi - \pi_0, 1)' \right]^{1/2}$$

For instance, the Monte-Carlo exercises in [Andrews et al. \(2006\)](#) depend on the parameters  $(\sqrt{\lambda}\pi, \lambda)$ , where  $\lambda$  denotes the concentration parameter  $\beta(Z'Z)\beta$ . The probability density function of the induced weights for these parameters is given in Figure 1 below.

**Figure 1:** Weights for  $(\sqrt{\lambda}\pi, \lambda)$  induced by the weights on  $(\phi, \omega, \rho)$



DESCRIPTION: The figure is based on weights (2.6), (2.7) for the canonical parameters  $(\phi, \omega, \rho)$  and the formulas in equations (2.8), (2.9). The matrix  $\Omega$  is assumed to have unit diagonal elements and correlation parameter  $r = .5$ . The null hypothesis is  $\pi_0 = 0$ . There are 4 instruments,  $k = 4$ . The bivariate density is generated by a Monte-Carlo exercise with 50,000 independent draws from  $(\rho^2, \phi)$  and the Matlab bivariate density estimator `gkde2`.

**RESULT 1 (ECS TEST FOR IV):** *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \pi = \pi_0$  vs.  $\mathbf{H}_1 : \pi \neq \pi_0$  in an over-identified IV model with a single endogenous regressor and weights over  $(\beta, \pi)$  induced by (2.6), (2.7) rejects the null hypothesis if the statistic:*

$$(S'S - T'T) + 8 \ln \left[ I_0 \left( \frac{1}{8} \left[ (S'S - T'T)^2 + 4(S'T)^2 \right]^{1/2} \right) \right].$$

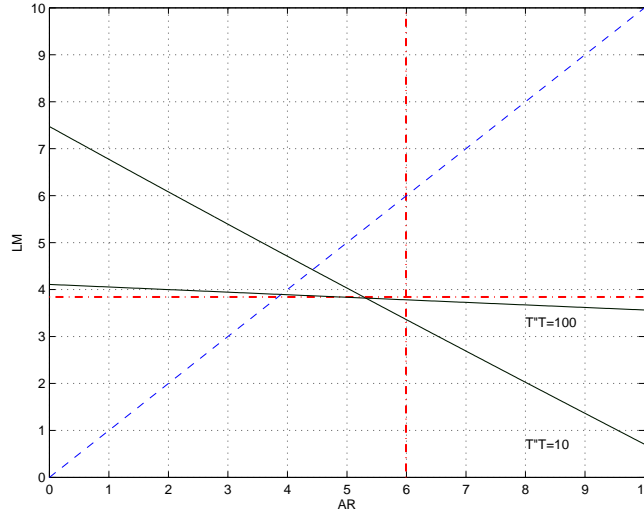
*exceeds the critical value function  $c(T, \alpha)$ , defined as the  $(1 - \alpha)$  quantile of the distribution of the statistic above with  $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$  and  $T$  fixed. The function  $I_0(\cdot)$  is the modified Bessel function of the first kind of order zero defined in Section 9.6, p. 375 of Abramowitz and Stegun (1964).*

PROOF: See Appendix C.

*Q.E.D.*

CONDITIONAL REJECTION REGION: Let  $AR \equiv S'S$  denote the Anderson and Rubin (1949) statistic for the over-identified IV model.<sup>11</sup> 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 1 is measurable with respect to the triplet  $(AR, LM, T'T)$ , hence it is invariant (see Andrews et al. (2006)). It is natural to ask whether the ECS test rejects the null hypothesis when both the AR and LM do.

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



(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  $\chi_2^2$  and  $\chi_1^2$ , respectively.

<sup>11</sup>This is a slight abuse of notation as  $AR = S'S/k$ ; see Andrews et al. (2006).

Figure 2 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 line traces the boundary of the rejection region of the ECS test for a given value of  $T'T \in \{10, 100\}$ .<sup>12</sup> The black solid line close to the LM critical value corresponds to the highest realization of  $T'T$ . The ECS test adjusts the  $\chi_k^2$  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  $\chi_{k,5\%}^2$  critical value for the AR is adjusted upwards and the null hypothesis is rejected only if  $AR > 9.7 > \chi_{k,5\%}^2$ . If, however,  $T'T = 100$  the adjustment 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  $LM > \chi_{1,5\%}^2$ , provided  $T'T$  is large.

POWER COMPARISON: Figure 3 below presents a simple power comparison of the ECS test in Result 1 against the AR, LM, and CLR. The null hypothesis is  $\pi_0 = 0$ . The reduced-form covariance matrix,  $\Omega$ , is assumed to have unit diagonal elements and correlation parameter  $r = .5$ . The number of instruments is  $k = 4$ . The distribution (under the alternative) of all the tests under comparison depends only on the parameters  $(\lambda^{1/2}\pi, \lambda)$  where  $\lambda = \beta'(Z'Z)^{1/2}\beta$ .<sup>13</sup>

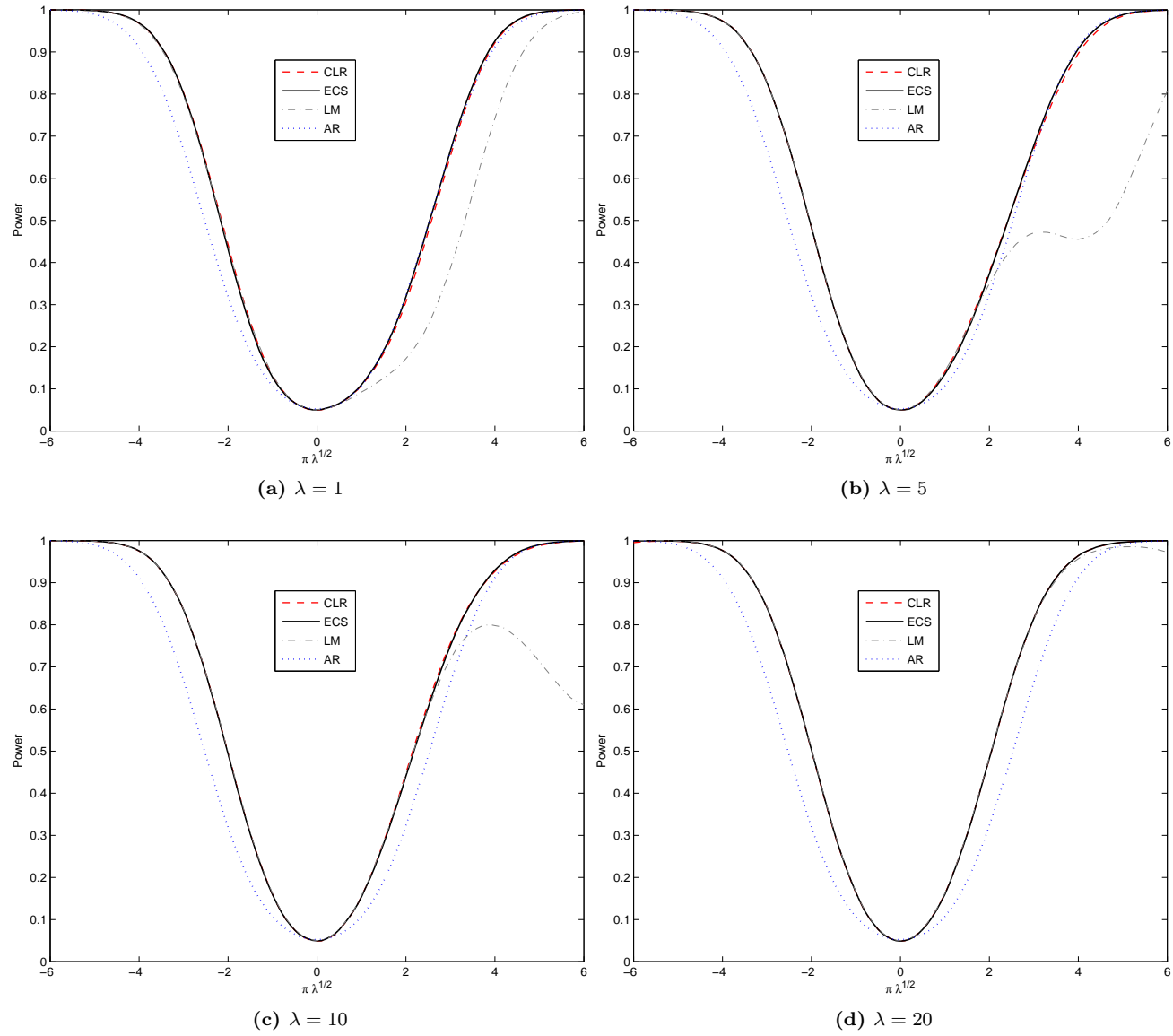
1. ECS VS. AR-LM: Result 1 implies there must be values in the parameter space for which the power curve of the ECS test lies above the AR, LM, CLR, and the power curve of any other similar/non-similar test. Figure 3 suggests that for low values of  $\lambda$  ( $\lambda = 1, 5$ ) the ECS test performs better than the LM for almost every positive alternative value of  $\pi$  and better than the AR for almost every negative alternative value (this results flips if the correlation parameter  $r$  takes a different sign). For large values of  $\lambda$ , the power curve of the ECS test is above the AR for most of the alternatives. The ECS and LM become almost indistinguishable for  $\lambda = 20$ .
2. ECS VS CLR: It has been shown, by means of numerical arguments, that the CLR power curve lies on the power envelope for two-sided invariant similar tests derived by Andrews et al. (2006). Figure 3 suggests that the ECS power curve is almost the same as the CLR power curve. Hence, in this small-scale power comparison, the new ECS test seems to share CLR’s numerical optimality properties. In fact, the admissibility result implies that there will be points in the parameter space for which the ECS power curve will be necessarily closer to any power envelope attained by the CLR.
3. EFFICIENCY: The WAP property of the ECS tests facilitates numerical comparison of testing procedures: one can use the weights to report differences in weighted average power or simply report power curves with nuisance parameters “integrated out”. For instance, in the context of Figure 3 below, the WAP comparison using weights (2.6) and (2.7) is as follows: 20.22% for the ECS, 20.13% for the CLR, 18.45% for the AR and 16.70% for the LM.

---

<sup>12</sup>The command `ezplot` in Matlab is used to graph the solution to the equation  $z(\text{AR}, \text{LM}, T'T) - c(T'T; \alpha) = 0$ .

<sup>13</sup>This follows from the fact that all tests are measurable with respect to maximal invariant  $(S'S, S'T, T'T)$ .

**Figure 3: ECS vs. AR, LM, CLR**  
( $k = 4$ )



### 3. LARGE-SAMPLE APPLICATIONS

This section is concerned with the application of the ECS tests to the extremum problems of AC12. More precisely, this section presents tests for:

$$\mathbf{H}_0 : \pi = \pi_0 \quad vs. \quad \mathbf{H}_1 : \pi \neq \pi_0,$$

inside a class of extremum problems with criterion function:

$$Q_n(\beta, \pi),$$

where  $n$  denotes the sample size.<sup>14</sup> The parameter  $\beta \in \mathbb{R}^{d_\beta}$  is neither specified nor restricted under the null hypothesis.<sup>15</sup>

The class of *extremum problems* studied by Andrews and Cheng (2012) has, broadly speaking, three main features. First, the parameter of interest ( $\pi$ ) is unidentified in some region of the parameter space. Second, the criterion function depends on a nuisance parameter that determines the strength of identification ( $\beta$ ). Third, the criterion function satisfies an asymptotic quadratic expansion around the point of lack of identification ( $\beta = \mathbf{0}$ ). This class of extremum problems includes, but it is not limited to, a probit model with endogeneity estimated by Conditional Marginal Maximum Likelihood (CMML), a nonlinear regression model estimated by nonlinear Least Squares (NLLS), an IV model with heteroskedasticity estimated by Quasi-Maximum Likelihood (QML).

This section argues that the finite-sample optimality properties of the ECS tests have large-sample efficiency consequences. In particular, Theorem 1 in Müller (2011) is used to show that “sample analogues” of the ECS tests applied to a “limiting statistical model” for extremum problems are, in a certain sense, *asymptotically efficient similar* under local-to-identification asymptotics.

The main conceptual challenge of this section is as follows. The ECS tests were defined for testing problems with an underlying statistical model,  $f(x, \theta)$ . The appeal of extremum problems is, however, that they do not require knowledge of the full distribution of the data. In fact, in AC12 the true distribution that generates the data is indexed by a parameter  $\gamma = (\beta, \pi, \phi) \in \Theta \times \Phi$ ; where the element  $\phi \in \Phi$  (possibly, of infinite dimension) denotes an unknown part of the data generating process. Thus, to apply the ECS tests to extremum problems it is necessary to map criterion functions,  $Q_n(\beta, \pi)$ , to a statistical model with a boundedly-complete, boundary-sufficient statistic.

The rest of this section is organized as follows. Section 3.1 presents an overview of the main ideas concerning the application of ECS tests to extremum problems. Section 3.2 introduces basic notation and presents a simple running example: a just-identified probit model with endogeneity estimated via Conditional Marginal Maximum Likelihood (CMML). Section 3.3 presents a summary of the results concerning the running example. Section 3.4 presents the main assumptions on the extremum problems under consideration [Assumption Q1-Q4.3]. Section 3.5 shows how to map the extremum problems under study to a “limiting” statistical model with a boundedly-complete, boundary-sufficient statistic. The main results of this section are Proposition 1 and 2. Section 3.6 shows how to build tests for the null hypothesis of interest using an ECS test for the “limiting” statistical model. This section also establishes the asymptotic properties of the recommended testing procedure. Section

<sup>14</sup>Explicit dependence of the criterion function on the data is ignored by notational convenience.

<sup>15</sup>AC12 also allow the criterion function to depend on an additional nuisance parameter  $\zeta$ , where  $\zeta$  unrelated to the identification of  $\pi$ . To keep the notation as simple as possible, this additional nuisance parameter is ignored in the examples presented in this section.

3.7 presents a small-scale Monte-Carlo exercise for the probit model with endogeneity. All proofs are collected in the Appendix.

### 3.1. Overview of the main results

LIMITING STATISTICAL MODEL FOR EXTREMUM PROBLEMS: This section introduces two new statistics for extremum problems in AC12: the  $\widehat{\beta}$ -score star and  $\widehat{\beta}$ -star, denoted

$$\begin{pmatrix} D_{\widehat{\beta}}^* \\ \widehat{\beta}^* \end{pmatrix}$$

This section states assumptions under which “local-to-unidentification” sequences of parameters ( $\sqrt{n}\beta_n \rightarrow b$ ) yield the following weak convergence result:

$$(3.1) \quad \widehat{C}_n \begin{pmatrix} \sqrt{n}D_{\widehat{\beta}}^* \\ \sqrt{n}\widehat{\beta}^* \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{2d_\beta} \left( \begin{pmatrix} \mu_1^*(\pi_0, b, \pi; \phi) \\ \mu_2^*(\pi_0, b, \pi; \phi) \end{pmatrix}, \mathbb{I}_2 \right),$$

where  $\widehat{C}_n$  is a  $2d_\beta \times 2d_\beta$  matrix that rotates and standardizes  $\widehat{\beta}$ -score star and  $\widehat{\beta}$ -star. The statistic  $X_2$  is shown to be boundedly complete, boundary sufficient for the problem  $\mathbf{H}_0 : \pi = \pi_0$  vs  $\mathbf{H}_1 : \pi \neq \pi_0$  in the statistical model above with parameters  $(b, \pi)$  and with elements  $\mu_1^*, \mu_2^*, \phi$  assumed known. The result in equation (3.1) is formalized by Propositions 1 and 2 in Section (3.5).

ECS FOR THE LIMITING STATISTICAL MODEL: ECS tests are available for the gaussian statistical model in equation (3.1) with sample space  $(X_1, X_2)$  and parameter space  $(b, \pi)$ . For example, it is shown that in the probit model with endogeneity introduced in Section (3.2) the test that rejects whenever  $X_1^2 > \chi_{1,1-\alpha}^2$  is an  $\alpha$ -level ECS test for some selection of weights over the parameters  $(b, \pi)$  in the limiting statistical model.

ECS EVALUATED AT SAMPLE ANALOGUES: The ECS test for the limiting statistical model is applied to extremum problems by evaluating the procedure at sample analogues. For instance, in the probit model with endogeneity the sample analogue of the test that rejects whenever  $X_1^2 > \chi_{1,1-\alpha}^2$  is given by the test that rejects if:

$$\left( \sqrt{n}D_{\widehat{\beta}}^*(0, \pi_0)/\widehat{\sigma}_1 \right)^2 > \chi_{1,1-\alpha}^2$$

where  $\widehat{\sigma}_1$  is an estimator for the asymptotic variance of  $\widehat{\beta}$ -score star.

ASYMPTOTIC EFFICIENCY: Theorem 1 in Müller (2011) is used to show that the limiting WAP of any asymptotically  $\alpha$ -similar test (under the weak convergence assumption (3.1)) is no larger than the limiting WAP of the ECS test evaluated at sample analogues.

### 3.2. Notation and Running Example

PARAMETER SPACE FOR EXTREMUM PROBLEMS: The extremum problems with criterion function  $Q_n(\beta, \pi)$  have the following parameter space:

$$\Theta \equiv \{ (\beta, \pi) \in \mathbb{R}^{d_\beta} \times \mathbb{R}^{d_\pi} \mid (\beta, \pi) \in B \times \Pi \subseteq \mathbb{R}^{d_\beta} \times \mathbb{R}^{d_\pi} \}.$$

The parameter of interest is  $\pi \in \Pi$ . The parameter  $\beta$  is associated to the identification of  $\pi$ . In particular, it is assumed that  $Q_n(0, \pi) = Q_n(0, \pi')$  for all  $\pi, \pi' \in \Pi$ .

DATA: A sample of size  $n$  is observed. The data, denoted  $W^n$ , is considered the  $n$ -th row of a (row-wise) i.i.d. triangular array:

$$W^n \equiv (W_{1n}, W_{2n}, \dots, W_{nn}),$$

where  $\{W_{in}\}_{i=1}^n$  is i.i.d. and  $W_{in} \in \mathcal{W} \subseteq \mathbb{R}^{d_s}$ . All the random vectors of the triangular array are defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .<sup>16</sup> As mentioned before, the distribution of the random vector  $W_{in}$  is fully specified by a tuple  $(\beta_n, \pi_n, \phi) \in \Theta \times \Phi$ , where  $\phi$  is, perhaps, an infinite dimensional parameter. The *true* parameter space, as defined in AC12, is denoted by  $\Gamma$  with elements  $\gamma = (\theta, \phi) \in \Gamma$ .

RUNNING EXAMPLE—PROBIT WITH ENDOGENEITY: Let  $\mathbf{y}_n^*, \mathbf{z}_n, \mathbf{u}_n, \mathbf{v}_n$  denote  $\mathbb{R}^n$ -valued random vectors with  $i$ -th elements  $y_{in}^*, z_{in}, u_{in}, v_{in}$ , respectively. Consider the econometric model:

$$\begin{aligned} \mathbf{y}_n^* &= \beta\pi\mathbf{z}_n + \mathbf{u}_n, \\ \mathbf{x}_n &= \beta\mathbf{z}_n + \mathbf{v}_n, \\ \mathbf{y}_n &= \mathbf{1}(\mathbf{y}_n^* > 0), \end{aligned}$$

where

$$\text{vec}[\mathbf{u}_n, \mathbf{v}_n] \sim \mathcal{N}_2\left(\mathbf{0}, \begin{bmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{bmatrix} \otimes \mathbb{I}_n\right),$$

The parameter space of this problem is  $\Theta = \mathbb{R}^2$ , or any subset of the plane that includes zero. The elements of the parameter space are  $(\beta, \pi) \in \mathbb{R}^2$ . It is assumed that  $\{u_{in}, v_{in}, z_{in}\}_{i=1}^n$  is an i.i.d. collection of random vectors and also that  $(u_{in}, v_{in})$  is independent of  $z_{in}$  for any sample size  $n$ .

The individual units observed by the econometrician are  $W_{in} = (y_{in}, x_{in}, z_{in})' \in \mathbb{R}^3$ . The distribution of  $W_{in}$  is fully specified by the tuple:

$$\gamma = (\beta, \pi, \phi) \quad \text{and} \quad \phi = (\rho, \sigma_v, P_z)$$

where  $P_z$  is the distribution of  $z_{in}$ .<sup>17</sup>The Conditional Marginal Maximum Likelihood criterion function is given by:<sup>18</sup>

$$(3.2) \quad Q_n(\beta, \pi) = \frac{1}{n} \sum_{i=1}^n \left( y_{in} \log \left[ L(\beta\pi z_{in}) \right] + (1 - y_{in})(1 - \log \left[ L(\beta\pi z_{in}) \right]) \right)$$

<sup>16</sup>One of the advantages of working with a triangular array is that the distribution of the data can depend on the sample size. This will be relevant when analyzing the behavior of tests statistics under drifting sequences of parameter values as in [Andrews and Cheng \(2012\)](#).

<sup>17</sup>The class of data generating processes (DGP) under consideration is oftentimes restricted by imposing moment conditions on the elements of  $\phi$  that guarantee the application of a Central Limit Theorem. For instance, see Appendix E.1 for common restrictions on the set  $\Phi$ .

<sup>18</sup>The adjective ‘‘Conditional’’ is used to denote conditioning with respect to the instrumental variable  $z_{in}$ . The adjective ‘‘Marginal’’ is used to remember the reader that the likelihood is based on the distribution of  $y_{in}|z_{in}$  and not on the joint conditional distribution  $(y_{in}, x_{in}) | z_{in}$ .



where  $L(\cdot)$  denotes the standard normal c.d.f.

### 3.3. *Probit Model with Endogeneity: Summary of results*

For the sake of concreteness, this section presents a brief summary of the results that will be obtained from the general theory presented in Sections 3.4 and 3.5.

The **limiting statistical model** for the probit with endogeneity is given by:

$$\begin{aligned} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} &\equiv \begin{pmatrix} D_\beta^*/\sigma_D \\ [\beta^*/\sigma_\beta - \rho D_\beta^*/\sigma_D]/\sqrt{1-\rho^2} \end{pmatrix} \\ &\sim \mathcal{N}_2 \left( \begin{array}{c} \mathbb{E}_\phi[z^2] \dot{L}(0)(\pi - \pi_0)b/\sigma_D \\ \mathbb{E}_\phi[z^2] b [1/\sigma_\beta - \rho \dot{L}(0)(\pi - \pi_0)/\sigma_D]/\sqrt{1-\rho^2} \end{array}, \mathbb{I}_2 \right) \end{aligned}$$

The **boundedly-complete, boundary-sufficient statistic** for the this model is given by:

$$X_2 = [\beta^*/\sigma_\beta - \rho D_\beta^*/\sigma_D]/\sqrt{1-\rho^2}.$$

where  $D_\beta^*$ ,  $\beta^*$  are defined in Section 3.4 as the limit of the new statistics  $\widehat{\beta}$ -score-star and  $\widehat{\beta}$ -star.

Given **weights**

$$\begin{aligned} \pi &\sim \frac{\gamma_1}{\dot{L}(0)\gamma_2} + \pi_0, \\ b &\sim \gamma_2 \end{aligned}$$

where  $(\gamma_1, \gamma_2)' \sim \mathcal{N}_2(\mathbf{0}, \lambda^{-2}\Sigma^0)$  and  $\lambda^2$  is a precision parameter selected by the econometrician, the **ECS test statistic** for the limiting statistical model is given by,

$$\exp\left(\frac{\mathbb{E}_\phi[z^2]^2}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_1^2\right) \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_2^2\right)$$

and the conditional  $\alpha$ -level **critical value function** equals:

$$\exp\left(\frac{\mathbb{E}_\phi[z^2]^2}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}\chi_{1,1-\alpha}^2\right) \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_2^2\right)$$

Therefore, the **ECS test for the limiting statistical model** is given by:

$$\phi_{\text{ECS}}(x_1, x_2) = \begin{cases} 1 & \text{if } x_1^2 > \chi_{1-\alpha}^2 \\ 0 & \text{i.o.c.} \end{cases}$$

and the **sample analogue of the ECS test** is:

$$\phi_{\text{ECS}}^n(\widehat{x}_1, \widehat{x}_2) = \begin{cases} 1 & \text{if } \widehat{x}_1^2 > \chi_{1-\alpha}^2 \\ 0 & \text{i.o.c.} \end{cases}$$

where

$$\hat{x}_1 = \sqrt{n} D_{\hat{\beta}}^* Q_n(0, \pi_0) / \hat{\sigma}_D = (1/\sqrt{n}) \sum_{i=1}^n z_{in} [y_{in} - L(\hat{\beta} \pi_0 z_{in})] / \hat{\sigma}_D$$

and

$$\begin{aligned} \hat{\sigma}_D^2 &= a_0 \left( \frac{1}{n} \sum_{i=1}^n \left[ \begin{pmatrix} y_{in} - L(\hat{\beta} \pi_0 z_{in}) \\ \hat{v}_{in} \end{pmatrix} \begin{pmatrix} y_{in} - L(\hat{\beta} \pi_0 z_{in}) \\ \hat{v}_{in} \end{pmatrix}' \right] \frac{1}{n} \sum_{i=1}^n z_{in}^2 \right) a_0' \\ a_0 &= \left[ 1, -\pi_0 \frac{1}{n} \sum_{i=1}^n z_{in}^2 L(\hat{\beta} \pi_0 z_{in}) \right], \\ \hat{v}_{in} &= x_{in} - \hat{\beta} z_{in}. \end{aligned}$$

and  $\hat{\beta}$  is the first-stage identification statistic:

$$\hat{\beta} = \left( \frac{1}{n} \sum_{i=1}^n z_{in}^2 \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_{in} x_{in}$$

Hence, Result 3 implies that the test  $\phi_{\text{ECS}}^n$  that rejects for large values of the  $\hat{\beta}$ -score-star is asymptotically efficient similar

### 3.4. Local-to-zero asymptotics and Assumptions Q1-Q4

This section presents Assumptions Q1-Q4. These assumptions are used to map the extremum problems in AC12 to a parametric statistical model. The “Q” denotes restrictions on the criterion functions  $Q_n(\beta, \pi)$  under consideration. This section also introduces two new statistics: the  $\hat{\beta}$ -score star and  $\hat{\beta}$ -star. The ECS tests will be functions of these statistics.

Let

$$(3.3) \quad \Gamma^* = \{ \{ \gamma_n \} \in \Gamma \mid \gamma_n \rightarrow \gamma, \text{ for some } \gamma \in \Gamma \}$$

$$(3.4) \quad \Gamma^*(\gamma, 0) = \{ \{ \gamma_n \} \in \Gamma^* \mid \gamma_n \rightarrow \gamma = (0, \pi, \phi) \text{ and } \sqrt{n} \beta_n \rightarrow b, b \in \mathbb{R}^{d_\beta} \}$$

Throughout this section  $\pi_0$  denotes the null hypothesis. A sequence  $\{ \gamma_n \} \in \Gamma^*(\gamma, 0)$  will be called a “weak sequence” or a “local-to-unidentification” sequence. A *null* sequence  $\{ \gamma_n \} \in \Gamma^*$  imposes  $\pi_0: \gamma_n = (\beta_n, \pi_0, \phi_n)$  for all  $n \in \mathbb{N}$ . A *b-weak* sequence  $\gamma_n \in \Gamma$  has a *drifting* parameter  $b$ ; i.e.,  $\sqrt{n} \beta_n \rightarrow b$ , with  $b \in \mathbb{R}^{d_\beta}$ . A  *$\beta$ -strong* sequence satisfies  $\beta_n \rightarrow \beta \neq 0$ , with  $\beta \in B$ . A *strong* sequence is  $\beta$  strong sequence for some  $\beta \in \mathbb{R}^{d_\beta}$ .

#### 3.4.1. ASSUMPTIONS Q1-Q3

**ASSUMPTION Q1 ( $\beta$ -score):**  $Q_n(\beta, \pi_0)$  is differentiable in  $\beta$  for a.e.  $W^n \in \mathcal{W}^n$ . Furthermore,

$$D_\beta Q_n(0, \pi_0) \equiv D_\beta Q_n(\beta, \pi) \Big|_{\beta=0, \pi=\pi_0} = g_1(W^n) \left[ \frac{1}{n} \sum_{i=1}^n m(W_{in}, \pi_0) \right],$$

with  $m(w, \pi_0) \neq 0$  for almost every  $w \in \mathcal{W}$ .

Assumption Q1 is closely related to Assumption C2 in AC12. This assumptions restricts the form of the  $\beta$ -score to guarantee that a standard Central Limit Theorem (CLT) for triangular arrays applies. To keep notation and intuition simple, Assumption Q1 dispenses the use of *generalized* first derivative in Assumption C2 and instead imposes differentiability for almost every realization of the data. Furthermore, Assumption Q1 does not require the additional moment restriction in part ii) of Assumption C2. Assumption Q1 is easily verified for the running example.

EXAMPLE 1—PROBIT WITH ENDOGENEITY: In Appendix E.2 it is shown that Assumption Q1 is satisfied with:

$$g_1(W^n) = \frac{\pi_0 \dot{L}(0)}{L(0)(1-L(0))}, \quad m(W_{in}, \pi_0) \equiv z_{in} [y_{in} - L(0)].$$

**ASSUMPTION Q2 (Identification Statistic)** There exists an “identification” statistic—possibly dependent on  $\pi_0$  :

$$\widehat{\beta}(\pi_0) \equiv g_2(W^n) \frac{1}{n} \sum_{i=1}^n f(W_{in}, \pi_0),$$

such that for any weak sequence  $\gamma_n = (\beta_n, \pi_n, \phi_n) \in \Gamma^*$  with limit  $\gamma = (0, \pi, \phi)$  :

$$g_2(W^n) \xrightarrow{P} a(\pi_0, \gamma) \in \mathbb{R}^{d_\beta \times d_\beta}$$

and for every null weak sequence  $\gamma_n = (\beta_n, \pi_0, \phi_n) \in \Gamma^*$  with limit  $\gamma = (0, \pi_0, \phi)$ :

$$\sqrt{n} \left( \widehat{\beta}(\pi_0) - \beta_n \right) \xrightarrow{d} \mathcal{N}_{d_\beta}(\mathbf{0}, V(\gamma)),$$

where  $V(\gamma)$  is a positive definite variance matrix.

Assumption Q2 plays an important role on the application of the ECS tests. In the extremum models in AC12 the identification of  $\pi$  is controlled by the finite-dimensional parameter  $\beta$ . Assumption Q2 requires the existence of an asymptotically normal estimate for  $\beta$  under any null weak sequence. The identification statistic is required to have a representation involving a sum of random variables. This will simplify the derivation of the limiting statistical model. Note that the identification statistic is allowed to depend on  $\pi_0$  (see the identification statistic for the nonlinear regression model in the Appendix F.3).

REMARK 6: The identification statistic is used in AC12 for *identification category selection* (pg. 2190). The identification statistic is the basis of their Type 1/Type 2 robust procedures.

EXAMPLE 1—CONTINUED: In Appendix E.3 it is shown that the standard first-stage identification statistic for the probit model satisfies Assumption Q2 with

$$g_2(W^n) = \left( \frac{1}{n} \sum_{i=1}^n z_{in}^2 \right)^{-1}, \quad f(W_{in}, \pi_0) \equiv z_{in} x_{in}, \quad a(\pi_0, \gamma) \equiv \mathbb{E}_\gamma [z^2]^{-1}.$$

**ASSUMPTION Q3 (Central Limit Theorem)** There exists  $\mathbb{R}^{d_\beta}$ -valued  $\beta$  differentiable functions:

$$m^*(W_{in}, \beta, \pi) \quad \text{and} \quad f^*(W_{in}, \beta, \pi)$$

such that for every *weak* sequence  $(\beta_n, \pi_n, \phi_n) \in \Gamma^*(\gamma, 0)$

$$\left( \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n m(W_{in}, \pi_0) - m^*(W_{in}, \beta_n, \pi_n) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n f(W_{in}, \pi_0) - f^*(W_{in}, \beta_n, \pi_n) \end{array} \right) \xrightarrow{d} \mathcal{N}_{2d_\beta}(\mathbf{0}, W(\pi_0, \gamma)),$$

where  $W(\pi_0, \gamma)$  is a full-rank covariance matrix.

The asymptotic properties—under weak sequences—of the extremum estimators analyzed in AC12 depend crucially on a weak convergence assumption for the  $\beta$ -score (see Assumption C3 in AC12). Since both the  $\beta$ -score and the identification statistic enter the robust procedures proposed in AC12, Assumption Q3 imposes a minimal weak convergence assumption for the tuple.

Assumption Q3 can be interpreted as follows. The terms that guarantee a CLT behavior for the  $\beta$ -score and the identification statistic are the functions  $m(W_{in}, \pi_0)$  and  $f(W_{in}, \pi_0)$ . The problem, though, is that these functions need not be properly centered at zero for an arbitrary selection of  $\gamma \in \Gamma$ . Assumption Q3 requires the existence of functions  $m^*(W_{in}, \beta, \pi)$  and  $f^*(W_{in}, \beta, \pi)$  that “recenter”  $m$  and  $f$  appropriately around zero, so that a multivariate CLT applies for every weak sequence.<sup>19</sup>

**EXAMPLE 1—CONTINUED:** In Appendix E.4 it is shown that Assumption Q3 is satisfied with

$$m^*(W_{in}, \beta, \pi) \equiv z_{in}[L(\beta\pi z_{in}) - L(0)] \quad \text{and} \quad f^*(W_{in}, \beta, \pi) \equiv \beta z_{in}^2$$

Hence,

$$m(W_{in}, \beta, \pi) - m^*(W_{in}, \beta, \pi) = z_{in}[y_{in} - L(\beta\pi z_{in})],$$

$$f(W_{in}, \beta, \pi) - f^*(W_{in}, \beta, \pi) = z_{in}(x_{in} - \beta z_{in}).$$

As shown in the Appendix, Assumption Q3 is established by verifying the assumptions of a Central Limit Theorem for triangular arrays.

Note that the known functions  $m^*$  and  $f^*$  depend on the nuisance parameter  $\beta$ . The key observation of this section is that one can estimate  $m^*$  by:

---

<sup>19</sup>In the appendix, it is shown that a stronger version of Assumption Q3 holds: weak convergence for every sequence  $\{\gamma_n\} \in \Gamma^*$ . Such assumption will play an important role in establishing the uniform validity of testing procedures.

$$m^*(W_{in}, \widehat{\beta}(\pi_0), \pi_0)$$

This observation motivates the following definitions. For a given pair:

$$(D_\beta Q_n(0, \pi_0), \widehat{\beta}(\pi_0))$$

consider the following statistics:

**DEFINITION 2.1 ( $\widehat{\beta}$ -Score-Star)** For given functions  $m(W_{in}, \pi_0)$ ,  $m^*(W_{in}, \beta, \pi)$  and the identification statistic  $\widehat{\beta}(\pi_0)$ , define the  $\widehat{\beta}$ -Score-Star as

$$D_\beta^* Q_n(0, \pi_0) \equiv \frac{1}{n} \sum_{i=1}^n m(W_{in}, \pi_0) - m^*(W_{in}, \widehat{\beta}(\pi_0), \pi_0)$$

**DEFINITION 2.2 ( $\widehat{\beta}$ -Star)** Let  $\widehat{\beta}$ -star be defined as

$$\widehat{\beta}^*(\pi_0) \equiv \frac{1}{n} \sum_{i=1}^n f^*(W_{in}, \pi_0)$$

Tests for the null hypothesis  $\pi_0$  will be based on  $\widehat{\beta}$ -score star and  $\beta$ -star.<sup>20</sup>

### 3.4.2. ASSUMPTIONS Q4.1-Q4.3

Let

$$m_\beta^*(W_{in}, \beta, \pi) \equiv D_\beta m^*(W_{in}, \beta, \pi) \quad \text{and} \quad f_\beta^*(W_{in}, \beta, \pi) \equiv D_\beta f^*(W_{in}, \beta, \pi),$$

The conceptual key to derive the asymptotic behavior of  $\widehat{\beta}(\pi_0)$ -score-star are the “estimation errors” that move  $\widehat{\beta}(\pi_0)$ -score-star away from the asymptotic behavior of Assumption Q3. The *first* “estimation error”—comes from the fact that  $\pi_0$  need not be the true parameter. This error, assuming away the estimation of  $\beta$ , is captured by:

$$\Delta_{m^*}^\pi(\pi_n, \pi_0) \equiv \frac{1}{n} \sum_{i=1}^n m^*(W_{in}, \beta_n, \pi_n) - m^*(W_{in}, \beta_n, \pi_0)$$

Note that even if  $\pi_0$  is the true parameter,  $\beta$  is unknown. Hence, the *second* estimation error comes from the following difference:

$$\Delta_{m^*}^\beta(\beta_n, \widehat{\beta}(\pi_0)) = \frac{1}{n} \sum_{i=1}^n m^*(W_{in}, \beta_n, \pi_0) - m^*(W_{in}, \widehat{\beta}(\pi_0), \pi_0)$$

Assumptions Q4.1, Q4.2, Q4.3 hold for every sequence  $\gamma_n = (\beta_n, \pi_n, \phi_n) \in \Gamma^*(\gamma, 0)$  with  $\sqrt{n}\beta_n \rightarrow b$  and  $\gamma = (0, \pi, \phi)$ .

---

<sup>20</sup>Note that given the data, the identification statistic, and the functions  $m^*$  the mapping between the  $\beta$ -score and the  $\beta$ -score star is one-to-one.

**ASSUMPTION Q4.1 (First Estimation Error)**  $\sqrt{n}\Delta_{m^*}^\pi(\pi_n, \pi_0) \xrightarrow{p} \mathbb{E}_\gamma \left[ m_\beta^*(w, 0, \pi) - m_\beta^*(w, 0, \pi_0) \right]' b$

Assumption Q4.1 requires the first estimation error to be, in the limit, a linear transformation of the drifting parameter  $b$ .

**REMARK 7:** The assumption can be motivated via a simple Taylor expansion. Suppose  $\beta \in \mathbb{R}$ . Since the model is unidentified at zero, the expectation of the  $\beta$ -score at any value of  $\pi$  should be zero when  $\beta = 0$ . Suppose, for simplicity, that  $m^*(w, 0, \pi) = m^*(w, 0, \pi_0) = 0$ . Then, a first-order exact Taylor expansion around 0 gives:

$$m^*(W_{in}, \beta_n, \pi_n) - m^*(W_{in}, \beta_n, \pi_0) = 0 + [m_\beta^*(W_{in}, \tilde{\beta}_{in}, \pi_n) - m_\beta^*(W_{in}, \tilde{\beta}_{in}, \pi_0)]\beta_n,$$

where  $\beta_{in} \rightarrow 0$  and  $\beta_n = b/\sqrt{n}$ . Therefore,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n m^*(W_{in}, \beta_n, \pi_n) - m^*(W_{in}, \beta_n, \pi_0) = b \sum_{i=1}^n [m_\beta^*(W_{in}, \tilde{\beta}_{in}, \pi_n) - m_\beta^*(W_{in}, \tilde{\beta}_{in}, \pi_0)]$$

**EXAMPLE 1—CONTINUED:** In Appendix E.5 it is shown that Assumption Q4.1 is satisfied and the first estimation error behaves as:

$$\begin{aligned} \sqrt{n}\Delta_{m^*}^\pi(\pi_n, \pi_0) &\xrightarrow{p} \mathbb{E}_\gamma [z_{in}^2 \dot{L}(0)(\pi - \pi_0)] b \\ &= \mathbb{E}_\phi [z^2] \dot{L}(0)(\pi - \pi_0) b \end{aligned}$$

Note that the first estimation error is zero whenever  $\pi = \pi_0$ .

**ASSUMPTION Q4.2 (Second estimation error)**

$$\sqrt{n}\Delta_{m^*}^\beta(\beta_n, \hat{\beta}(\pi_0)) = -\mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)]' \sqrt{n}(\hat{\beta}(\pi_0) - \beta_n) + o_{p, \gamma_n}(1)$$

Assumption Q4.2 requires the second estimation error to be, in the limit, a transformation of  $\sqrt{n}(\hat{\beta}(\pi_0) - \beta_n)$ . Under null sequences,  $\sqrt{n}\Delta_{m^*}^\beta(\beta_n, \hat{\beta}(\pi_0))$  will only affect the asymptotic variance in AQ3. However, when the null hypothesis is not true, there might be additional bias: the limiting distribution  $\sqrt{n}(\hat{\beta}(\pi_0) - \beta_n)$  need not have mean zero.

**EXAMPLE 1—CONTINUED:** In Appendix E.6 it is shown that Assumption Q4.2 is satisfied and the second estimation error behaves as:

$$\begin{aligned} \sqrt{n}\Delta_{m^*}^\beta(\beta_n, \hat{\beta}(\pi_0)) &= -\sqrt{n}(\hat{\beta}(\pi_0) - \beta_n) \mathbb{E}_\gamma [\pi_0 z^2 \dot{L}(0)] + o_{p, \gamma_n}(1) \\ &= -\left(\frac{1}{n} \sum_{i=1}^n z_{in}^2\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} v_{in}\right) \pi_0 \mathbb{E}_\phi [z^2] \dot{L}(0) + o_{p, \gamma_n}(1). \end{aligned}$$

The last assumption concerns the centering function  $f^*$ . This will be relevant to guarantee that the  $\hat{\beta}$ -star behaves as a normal—centered around a non-zero value:

**ASSUMPTION Q4.3:**  $\frac{1}{\sqrt{n}} \sum_{i=1}^n f^*(W_{in}, \beta_n, \pi_n) \xrightarrow{P} \mathbb{E}_\gamma[f_\beta^*(w, 0, \pi)]'b$ .

EXAMPLE 1—CONTINUED: In Appendix E.7 it is shown that Assumption Q4.3 is satisfied and:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n f^*(W_{in}, \beta_n, \pi_n) &= \sqrt{n}\beta_n \frac{1}{n} \sum_{i=1}^n z_{in}^2 \\ &\xrightarrow{P} b\mathbb{E}_\phi[z_{in}^2] \end{aligned}$$

### 3.5. Limiting Statistical Model for Extremum Problems

This section maps criterion functions,  $Q_n(\beta, \pi)$ , satisfying assumptions AQ1-AQ4.3 to a “limiting” statistical model with a boundedly-complete, boundary-sufficient statistic.

The map between a criterion function and a statistical model is constructed in three main steps:

1.  $Q_n(\beta, \pi) \mapsto (\widehat{\beta}\text{-score-star}, \widehat{\beta}\text{-star})$ : The tests presented in this section will be functions of  $\widehat{\beta}\text{-score-star}$  and  $\widehat{\beta}\text{-star}$ . As observed before, these statistics correspond to a one-to-one transformation of the  $\beta$ -score and the identification statistic used in AC12.<sup>21</sup>
2.  $(\widehat{\beta}\text{-score-star}, \widehat{\beta}\text{-star}) \mapsto (D_\beta^*, \beta^*)$ : Assumptions Q1 to Q4.3 imply—under local-to-unidentification sequences—the weak convergence of  $(\widehat{\beta}\text{-score-star}, \widehat{\beta}\text{-star})$  to a random vector  $(D_\beta^{*'}, \beta^{*'})' \in \mathbb{R}^{2d_\beta}$ . This weak convergence result yields a gaussian “limiting” statistical model:

$$\begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix} \sim \mathcal{N}_{2d_\beta} \left( \begin{pmatrix} \mu_1(\pi_0, b, \pi; \phi) \\ \mu_2(\pi_0, b, \pi; \phi) \end{pmatrix}, \Sigma(\pi_0, \gamma) \right), \quad b \in \mathbb{R}^{d_\beta}, \pi \in \Pi \subseteq \mathbb{R}^{d_\pi},$$

indexed by finite-dimensional parameters  $(b, \pi)$ , and functions  $\mu_1, \mu_2, \Sigma$  that will be treated as fixed objects.

The idea of a limiting statistical model is based on the theory of efficiency under a weak convergence assumption of Müller (2011).<sup>22</sup>

3.  $(D_\beta^*, \beta^*) \mapsto C^*(D_\beta^*, \beta^*)$ : After a one-to-one transformation of the gaussian model above based on a specific matrix  $C^*(\pi_0)$ , the transformed statistical model admits a boundedly-complete, boundary-sufficient statistic. It is shown that there is  $C^*(\pi_0)$

<sup>21</sup>The tests proposed in AC12 are based on weak convergence assumptions for  $\beta$ -score and an *identification statistic*. On the one hand, the  $\beta$ -score of  $Q_n(\beta, \pi)$  evaluated at the null hypothesis  $(\pi_0)$  and the point of lack of identification  $(\beta = \mathbf{0})$  is the leading term in the asymptotic quadratic expansion that determines the behavior of extremum problems under weak identification. On the other hand, the *identification statistic* is the main element in the *identification category selection* procedure (ICS) proposed in AC12 (see pg. 2189-2193) to build uniformly valid procedures.

<sup>22</sup>Just as him, this section takes the weak convergence assumption above as starting point and studies efficiency in the class of tests that remain asymptotically valid for all models that induce the same weak limit.

(see Lemma M in Appendix A.5) such that:

$$(3.5) \quad \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \equiv C^* \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix} \sim \mathcal{N}_{2d_\beta} \left( \begin{matrix} \mu_1^*(\pi_0, b, \pi; \phi) \\ \mu_2^*(\pi_0, b, \pi; \phi) \end{matrix}, \mathbb{I}_{2d_\beta} \right),$$

with

$$\mu_1^*(\pi_0, b, \pi_0, \phi) = \mathbf{0}.$$

The statistic  $X_2$  is boundedly complete and boundary sufficient, provided the function  $\mu_2^*(\pi_0, \cdot; \phi)$  contains a  $d_\beta$ -dimensional rectangle.<sup>23</sup>

Propositions 1 and 2 formalize the statements above.

### 3.5.1. **Proposition 1:** *The map from $(\widehat{\beta}$ -score-star, $\widehat{\beta}$ -star) to $(D_\beta^*, \beta^*)$*

The first proposition shows that, whenever the parameter  $\beta$  is close to zero, the problem of deciding how to use the statistics  $\widehat{\beta}$ -score-star and  $\widehat{\beta}$ -star to test the null hypothesis of interest ( $\pi = \pi_0$ ) is related to a well-defined finite-sample testing problem: there is a gaussian statistical model for the observations  $(D_\beta^*, \beta^*)$  (Equation 3.6), parameterized by  $(b, \pi)$  with functions  $a^*(\pi_0, \gamma)$ ,  $\mathbb{E}_\gamma[m_\beta^*(w, 0, \pi)]$ ,  $\mathbb{E}_\gamma[f_\beta(W_{in}, 0, \pi)]$ ,  $\Sigma(\pi_0, \gamma)$ .

**PROPOSITION 1:** *Suppose that Assumption AQ1-AQ4.3 hold. Then, under any weak sequence  $\gamma_n \in \Gamma^*(\gamma, 0)$  with limit  $\gamma = (0, \pi, \phi)$ :*

$$\begin{pmatrix} \sqrt{n}D_\beta^*Q_n(0, \pi_0) \\ \sqrt{n}\widehat{\beta}^*(\pi_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix},$$

where

$$(3.6) \quad \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix} \sim \mathcal{N}_{2d_\beta} \left( \begin{matrix} \mu_1(\pi_0, b, \pi, \phi) \\ \mu_2(\pi_0, b, \pi, \phi) \end{matrix}, \Sigma(\pi_0, \gamma) \right),$$

$$\begin{aligned} \mu_1(\pi_0, b, \pi, \phi) &= \mathbb{E}_\gamma[m_\beta^*(w, 0, \pi)] - a^*(\pi_0, \gamma)\mathbb{E}_\gamma[m_\beta^*(w, 0, \pi_0)] \Big] b \\ \mu_2(\pi_0, b, \pi, \phi) &= \mathbb{E}_\gamma[f_\beta(W_{in}, 0, \pi)]' b. \end{aligned}$$

The positive definite covariance matrix  $\Sigma(\pi_0, \gamma)$  is given by:

$$\begin{bmatrix} \mathbb{I}_{d_\beta} & -\mathbb{E}_\gamma[m_\beta^*(w, 0, \pi_0)]' a(\pi_0, \gamma) \\ 0 & \mathbb{I}_{d_\beta} \end{bmatrix} W(\pi_0, \gamma) \begin{bmatrix} \mathbb{I}_{d_\beta} & -\mathbb{E}_\gamma[m_\beta^*(w, 0, \pi_0)]' a(\pi_0, \gamma) \\ 0 & \mathbb{I}_{d_\beta} \end{bmatrix}',$$

and

$$a^*(\pi_0, \gamma) = \mathbb{E}_\gamma[f_\beta(w, 0, \pi)] a(\pi_0, \gamma)' \in \mathbb{R}^{d_\beta \times d_\beta},$$

where  $a(\pi_0, \gamma)$  is defined in Assumption Q2.

PROOF: See Appendix B.1

Q.E.D.

<sup>23</sup>Lehmann and Romano (2005) pg.



REMARK 8: Assumption AQ2 implies that whenever  $\gamma = (0, \pi_0, \phi)$ ,  $a^*(\pi_0, \gamma) = \mathbb{I}_{d_\beta}$ . Therefore, under the null hypothesis  $D_\beta^*$  will be a mean zero gaussian random vector.

EXAMPLE 1—CONTINUED: In the probit model with endogeneity:

$$a(\pi_0, \gamma) = E_\phi[z^2]^{-1} \quad \text{and} \quad \mathbb{E}_\gamma[f_\beta^*(w, 0, \pi)] = \mathbb{E}_\phi[z^2]$$

Therefore,  $a^*(\pi_0, \gamma) = 1$ . The statistical model in Proposition 1 is given by:<sup>24</sup>

$$(3.7) \quad \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix} \sim \mathcal{N}_2 \left( \begin{array}{c} \mathbb{E}_\phi[z^2] \dot{L}(0)(\pi - \pi_0)b \\ \mathbb{E}_\phi[z^2]b \end{array}, \Sigma(\pi_0, \gamma) \right), \quad (b, \pi) \in \mathbb{R}^2$$

and the variance matrix  $\Sigma(\pi_0, \gamma)$  equals

$$\begin{bmatrix} 1 & -\pi_0 \mathbb{E}_\phi[z^2] \dot{L}(0) \\ 0 & 1 \end{bmatrix} W(\pi_0, \gamma) \begin{bmatrix} 1 & -\pi_0 \mathbb{E}_\phi[z^2] \dot{L}(0) \\ 0 & 1 \end{bmatrix}'$$

$$W(\pi_0, \gamma) = \mathbb{E}_\phi[z^2] \mathbb{E}_\phi \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix}' \right]$$

The formula follows from the independence of  $z$  and  $(u, v)$ . The parameter  $\pi$  does not enter the variance  $\Sigma(\pi_0, \gamma)$ . Therefore, in the context of the statistical model in Equation 3.7, the testing problem of interest is a gaussian shift problem with correlated components. To stay within the context of limiting gaussian shift problems, this section imposes the following assumption

**ASSUMPTION Q5:**  $W(\pi_0, 0, \pi, \phi) = W(\pi_0, 0, \pi', \phi)$  for all  $\pi, \pi' \in \Pi$ .

EXAMPLE 1—CONTINUED: In the probit model  $\phi = (\rho, \sigma_v, P_z)$  does not depend on  $\pi$ . Therefore

$$W(\pi_0, \gamma) = \mathbb{E}_\phi[z^2] \mathbb{E}_\phi \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix}' \right]$$

does not depend on  $\pi$ .

Appendix F.8 presents the statistical model for the nonlinear regression problem and verifies Assumption Q5.

3.5.2. **Proposition 2:** *The map from  $(D_\beta^*, \beta^*)$  to  $C^*(D_\beta^*, \beta^*)$*

The following proposition will show that a (one-to-one) transformation of the statistical model in Proposition 1 admits a boundedly-complete, boundary-sufficient statistic. Partition the covariance matrix, evaluated at  $\gamma = (0, \pi_0, \phi)$ , according to:

<sup>24</sup>Since the bias of the second estimation error is zero (as the identification statistic does not depend on the null hypothesis), the centrality parameter for the probit model equals the probability limit of  $\sqrt{n} \Delta_{m^*}^\pi(\pi_n, \pi_0)$ .

$$\Sigma^0 \equiv \Sigma(\pi_0, (0, \pi_0, \phi)) \equiv \begin{pmatrix} \Sigma_{DD}^0 & \Sigma_{D\beta}^0 \\ \Sigma_{\beta D}^0 & \Sigma_{\beta\beta}^0 \end{pmatrix} = \Sigma(\pi_0, (0, \pi, \phi)),$$

Define:

$$R_{12} \equiv \left(\Sigma_{DD}^0\right)^{-1/2} \Sigma_{D\beta}^0 \left(\Sigma_{\beta\beta}^0\right)^{-1/2} \quad \text{and} \quad B = \mathbb{I}_{d_\beta} - R'_{12} R_{12}$$

Lemma M in A.5 shows that the matrix

$$C^* = \begin{pmatrix} \left(\Sigma_{DD}^0\right)^{-1/2} & \mathbf{0} \\ -B^{-1/2} R'_{12} \left(\Sigma_{DD}^0\right)^{-1/2} & B^{-1/2} \left(\Sigma_{\beta\beta}^0\right)^{-1/2} \end{pmatrix}$$

satisfies  $C^* \Sigma^0 C^{*'} = \mathbb{I}_{2d_\beta}$  and  $C^{*'} C^* = (\Sigma^0)^{-1}$ . Consider the one-to-one linear transformation:<sup>25</sup>

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \equiv C^* \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix},$$

and define the functions:

$$\begin{aligned} \mu_1^*(\pi_0, \beta, \pi) &= \left(\Sigma_{DD}^0\right)^{-1/2} \mu_1(\pi_0, b, \pi, \phi) \\ \mu_2^*(\pi_0, \beta, \pi) &= B^{-1/2} \left[ \left(\Sigma_{\beta\beta}^0\right)^{-1/2} \mu_2(\pi_0, b, \pi, \phi) - R'_{12} \left(\Sigma_{DD}^0\right)^{-1/2} \mu_1(\pi_0, b, \pi, \phi) \right] \end{aligned}$$

EXAMPLE 1—CONTINUED: In the probit model with endogeneity the  $2 \times 2$  matrix  $\Sigma(\pi_0, (0, \pi, \phi))$  is partitioned as:

$$\Sigma^0 = \Sigma(\pi_0, (0, \pi, \phi)) \equiv \begin{pmatrix} \sigma_D^2 & \sigma_{D\beta} \\ \sigma_{D\beta} & \sigma_\beta^2 \end{pmatrix}.$$

Note that

$$R_{12} = \frac{\sigma_{D\beta}}{\sigma_D \sigma_\beta} \equiv \rho \quad \text{and} \quad B = 1 - \rho^2.$$

Therefore

$$\begin{aligned} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} &\equiv \begin{pmatrix} 1/\sigma_D & 0 \\ -\rho/\sqrt{1-\rho^2}\sigma_D & 1/\sqrt{1-\rho^2}\sigma_\beta \end{pmatrix} \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix} \\ &= \begin{pmatrix} D_\beta^*/\sigma_D \\ [\beta^*/\sigma_\beta - \rho D_\beta^*/\sigma_D]/\sqrt{1-\rho^2} \end{pmatrix} \end{aligned}$$

---

<sup>25</sup>Note that the inverse of this mapping is given by:

$$\Sigma^0 C^{*'} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$X_1$  corresponds to the standardized limit of  $\widehat{\beta}$ -score-star. The statistic  $X_2$  can be interpreted as the error of the projection of the standardized limit of  $\widehat{\beta}$ -star over  $X_1$ , scaled by  $\sqrt{1 - \rho^2}$ . Furthermore:

$$\begin{aligned}\mu_1^*(\pi_0, b, \pi, \phi) &= \dot{L}(0)(\pi - \pi_0)\mathbb{E}_\phi[z^2]b/\sigma_D, \\ \mu_2^*(\pi_0, b, \pi, \phi) &= \mathbb{E}_\phi[z^2]\left[b/\sigma_\beta - \rho L(0)(\pi - \pi_0)b/\sigma_D\right]/\sqrt{1 - \rho^2}.\end{aligned}$$

The following proposition is key in the application of the ECS tests to the extremum problems in AC12. The main observation is that  $X_2$  is a boundary-sufficient statistic in the model associated to the transformed variables  $(D_\beta^*, \beta^*)$ .

**PROPOSITION 2:** *If  $(D_\beta^*, \beta^*)$  has the distribution in Equation 3.6 with parameters  $(b, \pi) \in \mathbb{R}^{d_\beta} \times \Pi$ . Then*

$$(3.8) \quad \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{2d_\beta} \left( \begin{array}{c} \mu_1^*(\pi_0, b, \pi; \phi) \\ \mu_2^*(\pi_0, b, \pi; \phi) \end{array}, \mathbb{I}_2 \right).$$

The statistic given by:

$$X_2 \equiv B^{-1/2} \left[ (\Sigma_{\beta\beta}^0)^{-1/2} \beta^* - R'_{12} (\Sigma_{DD}^0)^{-1/2} D_\beta^* \right]$$

is boundary sufficient for the problem  $\mathbf{H}_0 : \pi = \pi_0$  vs  $\mathbf{H}_1 : \pi \neq \pi_0$  in the statistical model of Equation 3.8. Furthermore, if the image of the mapping  $\mu_2^*(\pi_0, \cdot, \pi_0; \phi) : \mathbb{R}^{d_\beta} \rightarrow \mathbb{R}^{d_\beta}$  contains a  $d_\beta$ -dimensional rectangle (as defined in Theorem 4.3.1 of [Lehmann and Romano \(2005\)](#)), then  $X_2$  is boundedly-complete, boundary-sufficient for the problem in (3.8), with  $\mu_1^*, \mu_2^*, \phi$  assumed known.

PROOF: See Appendix B.2

*Q.E.D.*

EXAMPLE 1—CONTINUED: The limiting statistical model for the probit model with endogeneity based on Proposition 2 is given by:

$$(3.9) \quad \begin{aligned} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} &\equiv \begin{pmatrix} D_\beta^*/\sigma_D \\ \left[ \beta^*/\sigma_\beta - \rho D_\beta^*/\sigma_D \right] / \sqrt{1 - \rho^2} \end{pmatrix} \\ &\sim \mathcal{N}_2 \left( \begin{array}{c} \mathbb{E}_\phi[z^2] \dot{L}(0)(\pi - \pi_0)b/\sigma_D \\ \mathbb{E}_\phi[z^2] b \left[ 1/\sigma_\beta - \rho \dot{L}(0)(\pi - \pi_0)/\sigma_D \right] / \sqrt{1 - \rho^2} \end{array}, \mathbb{I}_2 \right) \end{aligned}$$

The boundary-sufficient statistic for the this model is:

$$X_2 = \left[ \beta^*/\sigma_\beta - \rho D_\beta^*/\sigma_D \right] / \sqrt{1 - \rho^2}.$$

Furthermore, under the null hypothesis:

$$X_2 \sim \mathcal{N} \left( \mathbb{E}_\phi[z^2] b / \sigma_\beta \sqrt{1 - \rho^2}, 1 \right)$$

Note that the map

$$\mu_2^*(\pi_0, b, \pi_0; \phi) = \mathbb{E}_\phi[z^2]b / \sigma_\beta \sqrt{1 - \rho^2},$$

with argument  $b$  and additional fixed parameters  $\phi, \sigma_\beta, \sqrt{1 - \rho^2}$  is onto the real line, provided  $\mathbb{E}_\phi[z^2] > 0$ ,  $\sigma_\beta > 0$  and  $\rho^2 < 1$ . Since the map is onto, it contains any rectangle  $[a, b]$ . Hence, Theorem 4.3.1 in [Lehmann and Romano \(2005\)](#) implies  $X_2$  is boundedly complete.

### 3.6. ECS test applied to extremum problems

The statistical model in Equation 3.8 satisfies the conditions required for the application of the ECS tests. This section shows how to use an ECS test for the limiting statistical model to test:

$$\mathbf{H}_0 : \pi = \pi_0 \quad \text{vs.} \quad \mathbf{H}_1 : \pi \neq \pi_0$$

given an extremum problem  $Q_n(\beta, \pi)$ . This section also uses Theorem 1 in [Müller \(2011\)](#) to show that the ECS tests “evaluated at sample analogues” are *asymptotically efficient similar*, provided the asymptotic variance  $\Sigma(\pi_0, (0, \pi_0, \phi))$  can be estimated consistently.

#### 3.6.1. ECS for the limiting statistical model

For a given choice of weights  $P_1(b, \pi)$  over  $(b, \pi)$ , the ECS test for the limiting statistical model (3.5) rejects if:

$$\frac{f_1^*(x_1, x_2; \mu_1^*, \mu_2^*)}{f_{\text{Bd}}(x_1|x_2)} > c(x_2; \alpha, \mu_1^*, \mu_2^*).$$

where the boundary conditional likelihood,  $f_{\text{Bd}}(x_1|x_2)$ , for the limiting model is given by

$$\frac{1}{(2\pi)^{d_\beta/2}} \exp\left(-\frac{1}{2}x_1'x_1\right)$$

and  $f_1^*(x_1, x_2; \mu_1^*, \mu_2^*)$  is the integrated likelihood.

**Integrated likelihood:** The first step to derive an ECS tests for the limiting statistical model is to compute the integrated likelihood,  $f_1^*(x_1, x_2; \mu_1^*, \mu_2^*)$ , which is proportional to:

$$\int_{\mathbb{R}^{d_\beta} \times \Pi} \exp\left(-\frac{1}{2}\|x_1 - \mu_1^*(\pi_0, b, \pi; \phi)\|^2\right) \exp\left(-\frac{1}{2}\|x_2 - \mu_2^*(\pi_0, b, \pi; \phi)\|^2\right) dP_1(b, \pi).$$

Note that the evaluation of the integrated likelihood depends on the functions  $\mu_1^*$ ,  $\mu_2^*$  and  $\phi$ . The probit example shows that even if  $\phi$  is unknown to the econometrician it is possible to select weights for  $(b, \pi)$  that yield a closed-form expression for the integrated likelihood,  $f_1^*(x_1, x_2; \mu_1^*, \mu_2^*)$ .

**REMARK 9:** A more general procedure in which  $\mu_i^*$  can be replaced by consistent estimators—to allow, for example, numerical evaluation of the integrated likelihood—is covered in the Appendix F with an application to the nonlinear regression model estimated by nonlinear least squares in AC12.

EXAMPLE 1—CONTINUED: Consider the following weights:

$$(3.10) \quad \begin{aligned} \pi &\sim \frac{\gamma_1}{\dot{L}(0)\gamma_2} + \pi_0, \\ b &\sim \gamma_2 \end{aligned}$$

where

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \sim \mathcal{N}_2\left(0, \frac{1}{\lambda^2}\Sigma^0\right).$$

where  $\lambda^2$  is a precision parameter selected by the econometrician. Appendix E.9 shows that  $f_1^*(x_1, x_2; \mu_1^*, \mu_2^*)$  is proportional to:

$$\exp\left(-\frac{1}{2}x_1^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]^2}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_1^2\right) \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_2^2\right)$$

Therefore, the ECS test rejects whenever:

$$\exp\left(\frac{\mathbb{E}_\phi[z^2]^2}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_1^2\right) \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_2^2\right) > c(x_2; \alpha, \mu_1^*, \mu_2^*)$$

**Critical value function:** As discussed before, the critical value function is given by the  $x_2$ -conditional  $\alpha$  upper quantile of the ratio between the integrated likelihood and the boundary conditional likelihood.

EXAMPLE 1—CONTINUED: For a fixed  $x_2$ , the ECS test statistic is simply a monotone transformation of a  $\chi_1^2$  random variable. Therefore, the  $x_2$ -conditional  $\alpha$  upper quantile is given by:

$$\exp\left(\frac{\mathbb{E}_\phi[z^2]^2}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}\chi_{1,1-\alpha}^2\right) \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]}{2(\lambda^2 + \mathbb{E}_\phi[z^2])}x_2^2\right)$$

**RESULT 2:** Consider the limiting statistical model for the probit with endogeneity in Equation 3.9. The  $\alpha$ -ECS test for the weights in Equation 3.10 rejects whenever:

$$\left(D_\beta^*/\sigma_D\right)^2 > \chi_{1-\alpha}^2$$

where  $\chi_{1-\alpha}^2$  is the  $1-\alpha$  critical value of a central chi-square random variable with one degree of freedom.

PROOF: The result follows from the formula for the integrated likelihood in Appendix E.9. *Q.E.D.*

### 3.6.2. ECS for the extremum problem

The ECS is implemented by replacing  $(x_1, x_2)$  by its sample analogues:

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} \equiv \hat{C}_n \begin{pmatrix} D_{\hat{\beta}}^*(0, \pi_0) \\ \hat{\beta}^* \end{pmatrix}.$$

**ASSUMPTION Q6:** There exists a sequence of  $2d_\beta \times 2d_\beta$  matrices such under every sequence  $\{\gamma_n\} \in \Gamma^*(\gamma, 0)$ ,  $\hat{C}_n \xrightarrow{P} C^*$

EXAMPLE 1—CONTINUED: Appendix E.10 verifies Assumption Q6. The ECS test for the limiting problem in Equation 3.9, evaluated at sample analogues, rejects for large values of the standardized  $\hat{\beta}$ -score-star:

$$\left( \sqrt{n} D_{\hat{\beta}}^* Q_n(0, \pi_0) / \hat{\sigma}_D \right)^2 > \chi_{1-\alpha}^2$$

where

$$\begin{aligned} \hat{\sigma}_D^2 &= a_0 \frac{1}{n} \sum_{i=1}^n \left[ \begin{pmatrix} y_{in} - L(\hat{\beta}\pi_0 z_{in}) \\ \hat{v}_{in} \end{pmatrix} \begin{pmatrix} y_{in} - L(\hat{\beta}\pi_0 z_{in}) \\ \hat{v}_{in} \end{pmatrix}' \right] \frac{1}{n} \sum_{i=1}^n z_{in}^2 a'_0 \\ a_0 &= \left[ 1, -\pi_0 \frac{1}{n} \sum_{i=1}^n z_{in}^2 L(\hat{\beta}\pi_0 z_{in}) \right] \end{aligned}$$

and

$$\hat{v}_{in} = x_{in} - \hat{\beta} z_{in}.$$

### 3.6.3. Asymptotic Efficiency

GENERAL TESTS FOR EXTREMUM PROBLEMS: Tests for extremum problems are measurable functions

$$\phi^n : \mathbb{R}^{d_s \times n} \rightarrow [0, 1],$$

where  $\phi^n(w^n)$  indicates the probability of rejection after observing data  $w^n$ . The distribution of  $W^n$  is now written as  $F_n(b, \pi, \phi_n)$ , with  $\beta_n = b/\sqrt{n}$ .

The objective of this section is to compare the asymptotic performance of general tests  $\phi^n$  that depend on  $W^n$  against  $\phi_{\text{ECS}}$  tests, that depend only on the  $\hat{\beta}$ -score star and  $\hat{\beta}$ -star.

PARAMETER SPACE: The parameter space of the limiting statistical model for extremum problems is given by  $\Theta \equiv \mathbb{R}^{d_\beta} \times \Pi$ , with elements  $(b, \pi)$ .

STATISTICS: Consider the mapping  $h_n = \mathbb{R}^{d_s \times n} \rightarrow \mathbb{R}^{2d_\beta}$  given by:

$$h_n(W^n) \equiv \hat{C}_n \begin{pmatrix} D_{\hat{\beta}}^* \\ \hat{\beta}^* \end{pmatrix}$$

Let  $P_n(b, \pi, \phi_n)$  denote the distribution of  $X_n \equiv h_n(Y_n)$ . The ECS tests for extremum problems depend on the data only through  $h_n$ .

MODELS AND TESTS INTEREST: As mentioned before  $\phi$  denotes the part of the data generating process not indexed by  $(b, \pi)$ . Let  $\phi^* \in \Phi$  denote the “true” value of  $\phi$  in the extremum problem, which is unknown to the econometrician. Consider the set of all sequences  $\{\phi_n\}$  for which:

$$h_n(W_n) \xrightarrow{d} X \equiv C^* \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix} \sim \mathcal{N}_{2d_\beta} \left( \begin{array}{c} \mu_1^*(\pi_0, b, \pi; \phi^*) \\ \mu_2^*(\pi_0, b, \pi; \phi^*) \end{array}, \mathbb{I}_{2d_\beta} \right),$$

so that

$$P_n(b, \pi, \phi_n) \xrightarrow{d} P(b, \pi; \phi^*)$$

where  $P(b, \pi; \phi^*)$  is a multivariate normal probability measure with  $\mathbb{I}_{2d_\beta}$  covariance matrix and mean vector partitioned as  $\mu_1^*(\pi_0, b, \pi; \phi^*)$ ,  $\mu_2^*(\pi_0, b, \pi; \phi^*)$ . Define

$$\mathcal{M}(\phi^*) = \{ \{\phi_n\}_{i=1}^\infty \subset \Phi \mid P_n(b, \pi, \phi_n) \xrightarrow{d} P(b, \pi; \phi^*) \},$$

which corresponds to the set of all sequences under which the weak convergence assumption  $h(W^n) \xrightarrow{d} X$  holds. The elements of  $\mathcal{M}(\phi^*)$  are denoted by  $m$  (“models”). Consider the class  $\mathfrak{C}_{\alpha-s}$  of tests  $\phi^n$  that are  $\alpha$ -asymptotically similar under  $\mathcal{M}(\phi^*)$ :

$$(3.11) \quad \lim_{n \rightarrow \infty} \int (\phi^n(y) - \alpha) dF_n(b, \pi, \phi_n)(y) = 0 \quad \forall m \in \mathcal{M}(\phi^*), b \in \mathbb{R}^{d_\beta}$$

WEIGHTED AVERAGE POWER (WAP): The criterion for asymptotic efficiency is the limit (superior) of WAP. Given a model  $m \in \mathcal{M}(\phi)$ , the  $\text{WAP}_n$  of a test  $\phi^n$  is defined as:

$$\text{WAP}_n(\phi^n, m) \equiv \int_{\mathbb{R}^{d_\beta} \times \Pi} \left( \int_{\mathbb{R}^{d_s \times n}} \phi_n(w) dF_n(b, \pi, \phi_n)(w) \right) dP_1(b, \pi)$$

For a test  $\phi$  defined in the limiting statistical model (3.5) the WAP is denoted as:

$$\text{WAP}(\phi) \equiv \int_{\mathbb{R}^{d_\beta} \times \Pi} \left( \int_{\mathbb{R}^{2d_\beta}} \phi(x) dP(b, \pi; \phi^*)(x) \right) dP_1(b, \pi)$$

**RESULT 3:** *Let  $\phi_{\text{ECS}}$  be an ECS test for the statistical model in (3.5). Suppose  $\phi_{\text{ECS}}$  is lebesgue-almost everywhere continuous. Let  $\phi_{\text{ECS}}^n \equiv \phi_{\text{ECS}} \circ h_n$  be the sample analogue of the ECS test available for a sample of size  $n$ . Then any test  $\phi^n : \mathbb{R}^{d_s \times n} \rightarrow [0, 1] \in \mathfrak{C}_{\alpha-s}$ :*

$$\limsup_{n \rightarrow \infty} \text{WAP}_n(\phi^n, m) \leq \lim_{n \rightarrow \infty} \text{WAP}_n(\phi_{\text{ECS}}^n, m) = \text{WAP}(\phi_{\text{ECS}})$$

and  $\phi_{\text{ECS}}^n \in \mathfrak{C}_{\alpha-s}$ .

PROOF: The proof follows from Theorem 1 in Müller (2011). See Appendix B.3 for details. *Q.E.D.*

### 3.7. Monte-Carlo Simulation

This section presents a simple Monte-Carlo study to analyze the finite sample properties of the ECS test derived for the probit model with endogeneity (sample analogue of  $\phi_{\text{ECS}}$  in

Section 3.3). The main focus is on the finite-sample implications of the asymptotic efficiency and asymptotic similarity of the ECS tests.

The econometric model of interest is a probit with endogeneity. The matrix notation for the model is:

$$\begin{aligned} \mathbf{y}_n^* &= \beta\pi\mathbf{z}_n + \mathbf{u}_n, \\ \mathbf{x}_n &= \beta\mathbf{z}_n + \mathbf{v}_n, \\ \mathbf{y}_n &= \mathbf{1}(\mathbf{y}_n^* > 0), \end{aligned}$$

$$\text{vec}[\mathbf{u}_n, \mathbf{v}_n] \sim \mathcal{N}_2\left(\mathbf{0}, \begin{bmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{bmatrix} \otimes \mathbb{I}_n\right),$$

The distribution of the i.i.d. data is fully specified by  $(\beta, \pi)$  and the parameter  $\phi = (\sigma_v^2, \rho, P_z)$ , where  $P_z$  is the unknown distribution of the instrument,  $z_{in}$ . The sample size is  $n = 200$ ,  $\rho = .5$  and  $\sigma_v^2 = 1$ . The exercise considers  $I = 50,000$  Monte-carlo draws.

**ASYMPTOTIC SIMILARITY:** Figure 4 presents the rate of Type I error for two different specifications of  $\phi$ : normal and  $t$  distributed instrument. The asymptotic similarity of the ECS tests translates into good finite sample control of the rate of Type of I error. Note that for a sample of size  $n = 200$  the (Monte-carlo) rate of Type I error of the ECS test seem to be almost invariant to the value of  $\beta \in [-3, 3]$ .

**ASYMPTOTIC EFFICIENCY:** Figure 5 presents the power of the ECS for different values of  $\pi$ , ( $\pi - \pi_0 \in [-3, 3]$ ) and different values of the nuisance parameter  $\beta$  ( $\beta \in \{.1, .4, .7, 1\}$ ). Each graph in Figure 5 presents the Monte-Carlo power curve and also the power curve derived from the “local-to-zero” asymptotic approximation. As expected, the power of the test improves as  $|\beta|$  increases.

#### 4. CONCLUSION

This paper studies a class of hypothesis testing problems with a special feature: a boundary-sufficient statistic. The paper focuses on a two-sided testing problem with a nuisance parameter. The appendix shows that our framework also covers certain one-sided hypotheses.

The main contribution of this paper is a new class of tests: the Efficient Conditionally Similar (ECS) tests. In the two-sided testing problem with a boundedly-complete, boundary-sufficient statistic, the ECS tests are shown to be weighted average power (WAP) maximizers—with user-specified weights—subject to a similarity-*on-the-null* constraint. In one-sided problems, the ECS tests are *average risk minimizers* subject to a similarity-*on-the-boundary constraint*. The appeal of ECS tests is that they are essentially the only way to generate *admissible, similar-on-the-boundary* tests; two finite-sample properties oftentimes deemed desirable. An immediate application of the finite-sample theory developed in this paper is a new test for a gaussian Instrumental Variables model with homoskedastic errors, non-stochastic instruments, and known reduced-form variance matrix. The new test for the IV model is compared against the Conditional Likelihood Ratio (CLR) using a small scale Monte-Carlo exercise. The theory and simulations suggest that the ECS test does not lack any of the finite-sample properties that make the CLR the recommended procedure in the literature. In addition, the ECS is constrained efficient and admissible.



The second contribution of this paper is an application of the ECS test to the class of extremum problems studied in [Andrews and Cheng \(2012\)](#). The application has two steps. First, we map the criterion functions for extremum problems into a limiting statistical model satisfying bounded completeness and boundary sufficiency. The key element in this map is the introduction of a new statistic, the  $\hat{\beta}$ -score star, which has a pivotal asymptotic distribution under local-to-unidentification null sequences.<sup>26</sup> Second, we show that ECS tests for the limiting statistical model evaluated at sample analogues are, in a certain sense, asymptotically efficient and similar. The key element in this argument is the theory of efficient tests under a weak convergence assumption developed by [Müller \(2011\)](#).

The large-sample application of the ECS test to the extremum problems studied by [Andrews and Cheng \(2012\)](#) broadens the scope of the finite-sample theory developed in the paper. Appendix G also derives a limiting experiment for the weakly-identified models of [Stock and Wright \(2000\)](#).<sup>27</sup>

FINAL COMMENTS—RELATED LITERATURE: Two-sided problems with nuisance parameters have recently received renewed attention in the econometrics literature. Two relevant references are [Elliott et al. \(2013\)](#) and [Moreira and Moreira \(2010\)](#).

[Elliott et al. \(2013\)](#) propose tests that maximize WAP subject to a size control constraint; i.e, the maximal rate of Type I error is restricted to be smaller than or equal to certain threshold,  $\alpha$ . Their recommended procedure is very attractive because it applies to a general class of problems with a composite null and a composite alternative; for instance, they do not require the existence of a boundary-sufficient statistic. A potential problem with these tests is that their implementation can be numerically involved. The ECS tests will also require numerical methods, but only to compute the critical value function evaluated at the sample value of the boundary-sufficient statistic.

One could worry, of course, that in models with a boundedly-complete, boundary-sufficient statistic the WAP of an [Elliott et al. \(2013\)](#) test could (and will) generally exceed the WAP of an ECS test: after all, every  $\alpha$ -similar test has size  $\alpha$ . The probit model with endogeneity provides a simple, albeit interesting, example of an ECS test that also maximizes WAP subject to a size control constraint (simply note that the gaussian weight for the parameter  $b$  in pg. 27 is least favorable).

[Moreira and Moreira \(2010\)](#) present a general theory to approximate the power function of the test that maximizes WAP subject to a similarity constraint. Their approximation results are applicable to models with or without a boundary-sufficient statistic. Note that in two-sided problems satisfying boundary sufficiency but violating bounded completeness, the WAP of the approximation test proposed by [Moreira and Moreira \(2010\)](#) could, in principle, be larger than the WAP of ECS tests. An interesting exercise would be to compare the WAP loss in models that do not satisfy bounded completeness, such as the linear regression model with persistent regressors of [Jansson and Moreira \(2006\)](#) or the Behrens-Fisher problem described in [Linnik \(1968\)](#). This could allow the researcher to decide whether the gains in WAP compensate the computation intensity required to approximate [Moreira and Moreira \(2010\)](#)'s efficient similar test.

---

<sup>26</sup>In fact, by strengthening Assumption Q3 the pivotal asymptotic distribution will hold under every null sequence.

<sup>27</sup>The limiting model of a Structural Vector Autoregression identified with external instruments also admits a boundary-sufficient statistic, as shown in [Montiel Olea, Stock, and Watson \(2012\)](#).

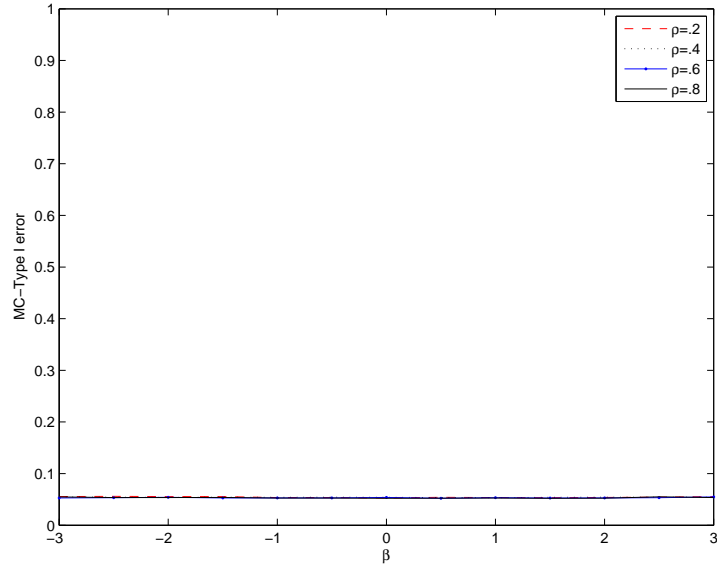
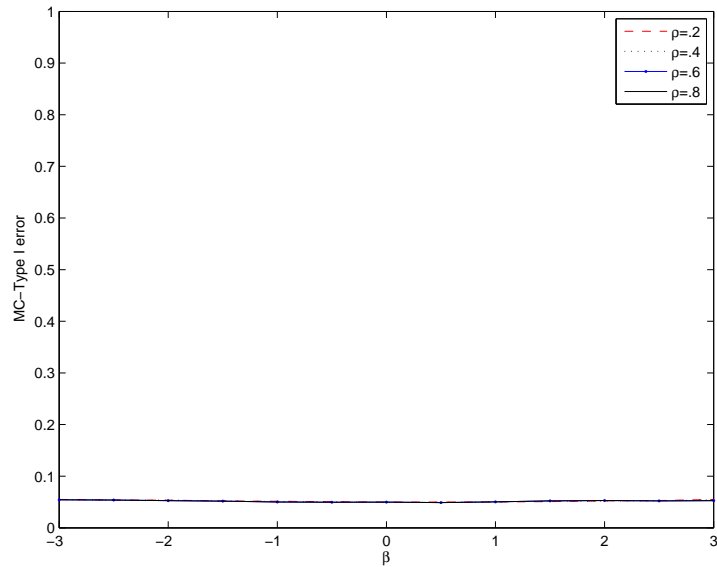
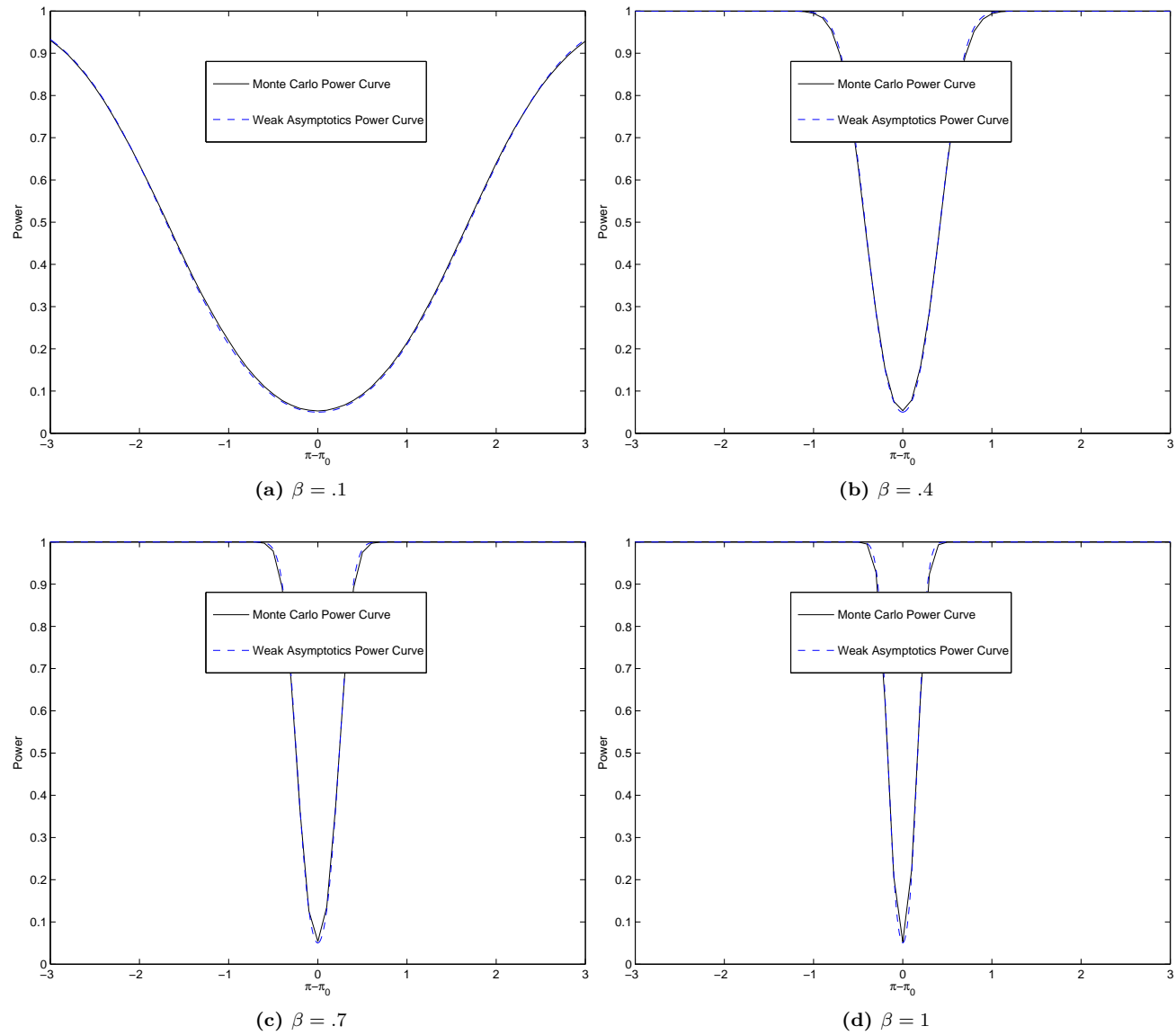
**Figure 4:** Monte-Carlo Type I error for ECS test**(a)**  $z_{in} \sim N(0, 1)$ **(b)**  $z_{in} \sim t_5 / (\sqrt{5/3})$ 

FIGURE 1: The MC simulation is based on 50,000 draws. The true parameter  $\pi = .1$  and  $\beta$  ranges from  $[-3, 3]$ . The distribution of  $z_{in}$  is either a  $N(0, 1)$  or a  $t$  random variable with 5 degrees of freedom. Each graph presents the rate of Type I error for different values of  $\rho$ , the correlation between  $u_{in}$  and  $v_{in}$ .

**Figure 5:** Monte-Carlo Power Curve for ECS test



## 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, 3rd ed.
- 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. (2012): "Similar-on-The-Boundary Tests for Moment Inequalities Exist, but Have Poor Power," *Discussion Paper No. 1815R, Cowles Foundation*.
- ANDREWS, D., M. MOREIRA, AND J. STOCK (2006): "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–52.
- (2007): "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 139, 116–132.
- (2008): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 146, 241–254.
- ANDREWS, D. W. AND X. CHENG (2012): "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.
- BILLINGSLEY, P. (1995): *Probability and Measure*, John Wiley & Sons, New York, 3rd ed.
- (1999): *Convergence of Probability Measures*, Wiley Series in Probability and Statistics, Wiley, New York, 2nd ed.
- CHAMBERLAIN, G. (2007): "Decision Theory Applied to an Instrumental Variables Model," *Econometrica*, 75, 609–652.
- CHESNEAU, C. (2009): "A Tail Bound for Sums of Independent Random Variables and Application to the Pareto Distribution," *Applied Mathematics E-Notes*, 9, 300–306.
- DURRETT, R. (2010): *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 4th ed.
- ELLIOTT, G., U. MÜLLER, AND M. WATSON (2013): "Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis," Working Paper.
- FERGUSON, T. (1967): *Mathematical Statistics: A Decision Theoretic Approach*, vol. 7, Academic Press New York.
- FIELLER, E. C. (1954): "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 175–185.
- JANSSON, M. AND M. MOREIRA (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74, 681–714.
- KLEIBERGEN, F. AND S. MAVROEIDIS (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics*, 27, 293–311.
- LE CAM, L. (1986): *Asymptotic Methods in Statistical Decision Theory*, Springer Verlag.
- LEHMANN, E. AND J. ROMANO (2005): *Testing Statistical Hypotheses*, Springer Texts in Statistics, Springer Verlag.
- LINNIK, J. (1968): *Statistical Problems with Nuisance Parameters*, vol. 20, American Mathematical Society.
- MARDIA, K. AND P. JUPP (2000): *Directional Statistics*, Wiley Series in Probability and Statistics, Wiley.
- 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.
- MOREIRA, M. AND H. MOREIRA (2010): "Contributions to the Theory of Similar Tests," Working Paper, December 2010.
- MÜLLER, U. K. (2011): "Efficient Tests under a Weak Convergence Assumption," *Econometrica*, 79, 395–435.
- MUNKRES, J. (2000): *Topology*, Prentice Hall Inc, 2nd ed.
- 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, McGraw-Hill, New York.
- (2006): *Real and Complex Analysis*, Tata McGraw-Hill Education, 3rd ed.
- 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 J. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.
- STROOCK, D. (1999): *A Concise Introduction to the Theory of Integration*, Birkhauser, 2nd ed.
- WALD, A. (1950): "Statistical Decision Functions," *New York*, 660.

## APPENDIX A: FINITE-SAMPLE THEORY

A.1. Lemma 1: Weak\* compactness of  $\mathcal{C}(\alpha\text{-sb})$ 

DESCRIPTION: The first lemma of this appendix shows that the both the set of  $\alpha$ -conditionally similar-on-the-boundary tests, denoted  $\mathcal{C}(\alpha\text{-csb})$ , and the set of  $\alpha$ -similar-on-the-boundary-tests, denoted  $\mathcal{C}(\alpha\text{-sb})$ , are compact relative to the space of essentially bounded measurable functions endowed with the weak\* topology.

RELEVANCE OF LEMMA 1: This lemma will be used to prove part ii) of Theorem 1. The weak\* compactness of  $\mathcal{C}(\alpha\text{-csb})$  will allow the application of an essentially complete class Theorem [See Theorem 3, pg. 87, Chapter 2 in Ferguson (1967)]. This lemma is also used to show the *existence* of a very general class of admissible and similar procedures based on a criterion function that trades-off Type I and Type II error. Such class of tests will be described in Lemma 2.

**PRELIMINARIES 1** ( $L^1$  and  $L^\infty$ ): 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  $\sigma$ -finite measure space, where  $\lambda^s$  denotes the Lebesgue measure in  $\mathbb{R}^s$  restricted to  $\mathbf{X}$ . 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 10: 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  $\{f(x, \theta)\}_{\theta \in \Theta}$  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  $\theta \in \Theta$ .

**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  $\Lambda$  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  $\Lambda$  with a unique element (up to equivalence)  $g \in L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ , and vice versa: for  $f \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)^*$ , the functional  $\Lambda \in [L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$  is of the form:

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

**PRELIMINARIES 3** (weak\* topology on  $L^\infty$ ): 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

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

Let  $(\mathbf{X}, \Theta, f, \Theta_0)$  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

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

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 compactness 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  $\lambda^s$  is  $\sigma$ -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  $\phi(x) \leq 1$   $\lambda^s$ -almost everywhere in  $\mathbf{X}$ . An analogous argument yields  $\phi(x) \geq 0$   $\lambda^s$ -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  $\theta \in \text{Bd}\Theta_0$   $f(\cdot; \theta)$  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

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

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

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

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

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

Note that  $\mathcal{C}(\alpha\text{-}\mathcal{G}) \subseteq \mathcal{C} \subseteq K$ , as for any test  $\left| \int_{\mathbf{X}} f(x)\phi(x) dx \right| \leq \int_{\mathbf{X}} |f(x)|\phi(x) dx \leq \int_{\mathbf{X}} |f(x)| dx \leq 1$ . 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  $\mathcal{C}(\alpha)$  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  $\alpha$ -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  $\alpha$ -conditionally similar-on-the boundary tests,  $\mathcal{C}(\alpha\text{-csb})$  is weak\* compact.

PROOF: Set  $\mathcal{G} = \{g \mid g(x_1, x_2) = 1 \text{ if } x_2 \in \mathcal{F}; g(x_1, x_2) = 0 \text{ if } x_2 \notin \mathcal{F} \text{ for some } \mathcal{F} \in \mathcal{B}(\mathbf{X}_2)\}$ . 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  $P_\theta$  is the measure induced by  $f(x; \theta)$ . Note that

$$\int_{\mathbf{X}} (\phi(X_1, X_2) - \alpha)g(X_1, X_2)f(X_1, X_2, \theta) = \alpha \quad \forall g \in \mathcal{G}$$

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  $P_\theta$ . And this holds for every  $\theta \in \text{Bd}\Theta_0$ .

*Q.E.D.*

*A.2. Lemma 2: Some General Results concerning the existence of Admissible (Conditionally) Similar-on-the-boundary tests*

DESCRIPTION: Let  $(\mathbf{X}, \Theta, f, \Theta_0)$  be a hypothesis testing problem with a product sample space  $(\mathbf{X}_1, \mathbf{X}_2)$ .  $\Theta_0$  is allowed to be any closed set relative to  $(\Theta, \mathcal{T})$  (hence, the formulation will cover one-sided and two-sided problems with nuisance parameters). Let  $\mathcal{G}$  be a collection of bounded measurable functions,  $g : \mathbf{X} \rightarrow \mathbb{R}$ . This Lemma 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 weights. Second, trade off the average rates of Type I/Type II error using a *strictly monotone function*  $\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$ , while imposing the constraints in  $\mathcal{C}(\alpha\text{-}\mathcal{G})$ .<sup>28</sup> Lemma 3 shows that ECS tests defined in (A.9) are a particular case of this approach with a linear tradeoff function  $\mathbf{W}(x, y) = \tau x + (1 - \tau)y$ .

RELEVANCE OF LEMMA 2: Lemma 2 will be used to establish a general property that will imply part i) of Theorem 1. Lemma 2 also illustrates the relevance of part ii) of Theorem 1: there are different ways to generate admissible (conditionally) similar-on-the-boundary tests; however, all of them are, in a sense, equivalent to extended ECS. The statement of Lemma 2 is general enough to include testing problems where similarity on the boundary is of interest, but in which there is no boundary-sufficient statistic. For example, moment inequality models.

**LEMMA 2:** Let  $w_i(\theta)$  denote a full-support probability measures over  $\text{Int } \Theta_i$ , for  $i = \{0, 1\}$ , and let  $\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a continuous, strictly monotone function. Define

$$M(\mathbf{W}, w_1, w_0, \mathcal{G}) \equiv \arg \min_{\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})} \mathbf{W} \left( \int_{\text{int } \Theta_1} R(\phi, \theta) dw_1(\theta), \int_{\text{Int } \Theta_0} R(\phi, \theta) dw_0(\theta) \right).$$

and suppose that Assumption F0 holds. Then,

**L2a:** If the sample space  $\mathbf{X}$  is topologically separable:

$$M(\mathbf{W}, w_1, w_0, \mathcal{G}) \neq \emptyset.$$

**L2b:** If  $g^* \in \mathcal{G}$ , where  $g^*(x) = 1$  for all  $x \in \mathbf{X}$ . Under Assumption F0,

$$\phi^* \in M(\mathbf{W}, w_1, w_0, \mathcal{G}) \implies \phi^* \text{ is admissible in } \mathcal{C}(\alpha\text{-}\mathcal{G}).$$

**L2c:** If  $\mathcal{G}^* \equiv \{g^*\}$ , so that  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*) = \mathcal{C}(\alpha\text{-sb})$

$$\phi^* \in M(\mathbf{W}, w_1, w_0, \mathcal{G}^*) \implies \phi^* \text{ is admissible in } \mathcal{C}.$$

**Proof of L2a:** For simplicity, assume that  $w_0$  and  $w_1$  have associated pdfs  $p_0$  and  $p_1$ . 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 L2a it will be sufficient to show that the objective function

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

is continuous in the weak\* topology.

---

<sup>28</sup> $\mathbf{W}(x, y)$  is strictly monotone if whenever  $x \leq x'$ ,  $y \leq y'$  (with either  $x < x'$  or  $y < y'$ ), then  $\mathbf{W}(x, y) < \mathbf{W}(x', y')$ .

**L2a-Step 1** (Fubini's Theorem:) Since the image of any test  $\phi \in \mathcal{C}$  is contained in the interval  $[0, 1]$   $\lambda^s$ -a.e. and  $f(x; \theta) \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  for all  $\theta$ , it follows that  $\left( \int_{\mathbf{X}} \phi(x) f(x; \theta) dx \right) \leq 1$  for every  $\theta \in \Theta$ . Furthermore, since  $p_1(x)$  and  $p_0(x)$  are also probability density functions on  $\text{Int}\Theta_1$  and  $\text{Int}\Theta_0$  the following holds

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

and

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

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

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

and

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

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

$$(A.2) \quad f_1^*(x) \equiv \int_{\text{int } \Theta_1} f(x; \theta) p_1(\theta) d\theta, \quad f_0^*(x) \equiv \int_{\text{int } \Theta_0} f(x; \theta) p_0(\theta) d\theta.$$

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  $\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$  induces a functional  $\mathcal{W}^*$  over  $\mathcal{D}$ :

$$(A.3) \quad \mathcal{W}^*(\phi) \equiv \mathbf{W} \left( \int_{\mathbf{X}} (1 - \phi(x)) f_1^*(x) dx, \int_{\mathbf{X}} \phi(x) f_0^*(x) dx \right)$$

**L2a-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  $\phi_n \rightarrow^* \phi$ . 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

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

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

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

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  $M(\mathbf{W}, w_1, w_0, \mathcal{G}) \neq \emptyset$ .

**L2b** : Let  $\phi^* \in M(\mathbf{W}, w_1, w_0, \mathcal{G})$ . I show that if  $\phi' \in \mathcal{C}(\alpha\text{-}\mathcal{G})$  satisfies

$$(A.4) \quad \mathbb{E}_{\theta}[\phi'(X)] \leq \mathbb{E}_{\theta}[\phi^*(X)] \quad \forall \theta \in \Theta_0$$

and

$$(A.5) \quad \mathbb{E}_{\theta}[\phi'(X)] \geq \mathbb{E}_{\theta}[\phi^*(X)] \quad \forall \theta \in \Theta_1$$

then

$$(A.6) \quad \mathbb{E}_{\theta}[\phi'(x)] = \mathbb{E}_{\theta}[\phi^*(x)] \quad \forall \theta \in \Theta = \Theta_0 \cup \Theta_1.$$

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



Suppose (A.4) and (A.5) hold, but (A.6) 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  $\Theta_1$ . Since  $w_1(\theta)$  has full support on  $\text{Int}\Theta_1$ ,  $\int_A dw_1(\theta) > 0$  for any open set  $A$  contained in  $\Theta_1$ . Note that  $\Delta_{\phi', \phi^*}(\theta) > 0$  for all  $\theta \in \mathcal{S}_{\tilde{\theta}}$  and equations (A.4)-(A.5) imply

$$\int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi'(x)f(x; \theta)dx \right) dw_0(\theta) \leq \int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi^*(x)f(x; \theta)dx \right) dw_0(\theta).$$

and

$$\int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi'(x))f(x; \theta)dx \right) dw_1(\theta) < \int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi^*(x))f(x; \theta)dx \right) dw_1(\theta)$$

The monotonicity of  $\mathbf{W}$  implies that  $\mathcal{W}^*(\phi') < \mathcal{W}(\phi^*)$ . This contradicts the fact that  $\phi^* \in M(\mathbf{W}, w_1, w_0, \mathcal{G})$ . 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  $\mathcal{S}_{\tilde{\theta}}$  contained in  $\text{Int}\Theta_0$  such that  $\Delta_{\phi', \phi^*}(\theta) < 0$  for all  $\theta \in \mathcal{S}_{\tilde{\theta}}$ . Since this set has positive probability under  $w_0$ , this implies

$$\int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi'(x)f(x; \theta)dx \right) w_0(\theta)d\theta < \int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi^*(x)f(x; \theta)dx \right) w_0(\theta)d\theta$$

and

$$\int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi'(x))f(x; \theta)dx \right) w_1(\theta)d\theta \leq \int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi^*(x))f(x; \theta)dx \right) w_1(\theta)d\theta.$$

Which, once again, contradicts the fact that  $\phi^* \in M(\mathbf{W}, w_1, w_0, \mathcal{G})$ .

Therefore, (A.4) and (A.5) imply (A.6). I conclude that  $\phi^*$  is admissible in  $\mathcal{C}(\alpha\text{-}\mathcal{G})$ .

**L2c (Outline):** Let  $\mathcal{G}^* \equiv \{g : \mathbf{X} \rightarrow \mathbb{R} \mid g(x) = 1 \ \forall x \in \mathbf{X}\}$ , so that the class  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$  coincides with  $\mathcal{C}(\alpha\text{-sb})$ . I show that a test  $\phi^* \in M(\mathbf{W}, w_1, w_0, \mathcal{G}^*)$  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

$$(A.7) \quad \mathbb{E}_{\theta}[\phi'(X)] \leq \mathbb{E}_{\theta}[\phi^*(X)] \quad \forall \theta \in \Theta_0$$

and

$$(A.8) \quad \mathbb{E}_{\theta}[\phi'(X)] \geq \mathbb{E}_{\theta}[\phi^*(X)] \quad \forall \theta \in \Theta_1$$

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

Let  $\mathcal{C}_{ns} \subset \mathcal{C}$  be the class of tests that are not similar on the boundary of  $\Theta_0$ . 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

$\mathcal{C} \equiv \mathcal{C}_{ns} \cup (\mathcal{C} \setminus \mathcal{C}_{ns})$ . Take any test  $\phi' \in \mathcal{C}_{ns}$  that satisfies (A.4). Since  $\phi'$  is an element of  $\mathcal{C}_{ns}$  and  $\Theta_0$  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  $\theta$ , there exists an open neighborhood  $\tau_\theta \in \mathcal{T}$  such that  $\Delta_{\phi', \phi^*}(\theta) < 0$  for all  $\theta \in \tau_\theta$ . Since  $\theta$  is an element of  $\text{Bd}\Theta_0$ , then  $\tau_\theta \cap \Theta_1 \neq \emptyset$ . The latter implies there exists  $\theta_1 \in \Theta_1$  such that  $\Delta_{\phi', \phi^*}(\theta_1) = \mathbb{E}_{\theta_1}[\phi'(X)] - \mathbb{E}_{\theta_1}[\phi^*(X)] < 0$ . Therefore, equation (A.4) and (A.5) cannot hold. We conclude there is no test  $\phi' \in \mathcal{C}_{ns}$  that satisfies (A.4) and (A.5). Since  $\mathcal{C}_{ns}$  partitions  $\mathcal{C}$ , a test  $\phi' \in \mathcal{C}$  that satisfies (A.4) and (A.5) must be an element of  $\mathcal{C} \setminus \mathcal{C}_{ns}$  (as  $\phi' \notin \mathcal{C}_{ns}$ ). Equation (A.4) implies  $\phi'$  is  $\alpha'$ -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  $\phi'$  will violate (A.5). We conclude that any test that satisfies (A.4) and (A.5) must be  $\alpha$ -similar on  $\text{Bd}\Theta_0$ .

**STEP 2:** Since  $\phi^* \in M(\mathbf{W}, w_1, w_0, \mathcal{G}^*)$ ,  $\phi^*$  is admissible in  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$ . Therefore, there is no  $\alpha$ -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  $\phi' \in \mathcal{C}$  that satisfies (A.4) and (A.5) must be  $\alpha$ -similar on  $\text{Bd}\Theta_0$ , I conclude  $\phi^*$  is admissible in  $\mathcal{C}$

### A.3. Lemma 3: ECS are Average Risk Minimizers

DESCRIPTION: (*Generalized version of ECS tests for problems in which  $\Theta_0$  need not coincide with  $\text{Bd}\Theta_0$* ) Let  $w_i(\theta)$  denote probability measures over  $\text{Int}\Theta_i$ , for  $i = \{0, 1\}$ . Lemma 3 will show that the test that rejects whenever

$$(A.9) \quad z(x_1, x_2) \equiv \left[ \tau f_{w_1}^*(x_1, x_2) + (1 - \tau) f_{w_0}^*(x_1, x_2) \right] / f_{\text{Bd}}(x_1 | x_2) > c(x_2; \alpha, w),$$

is an element of

$$M(\tau, w_1, w_0, \alpha\text{-csb}) \equiv \arg \min_{\phi \in \mathcal{C}(\alpha\text{-csb})} \tau \int_{\text{Int}\Theta_1} R(\phi, \theta) dw_1(\theta) + (1 - \tau) \int_{\text{Int}\Theta_0} R(\phi, \theta) dw_0(\theta),$$

provided  $c(x_2; \alpha, w)$  is the  $1 - \alpha$  quantile of  $z(X_1, x_2)$  with  $X_1 \sim f_{\text{Bd}}(X_1 | x_2)$ . In a slight abuse of notation the tests in (A.9) will be denoted  $\phi_{\text{ECS}}$  as well.

REMARK 11: Note that when  $\Theta_0 = \text{Bd}\Theta_0$  the tests coincide with the definition of ECS tests in equation (2.2).

RELEVANCE OF LEMMA 3: From Theorem 4.3.2 in Lehmann and Romano (2005) it follows that if  $X_2$  is boundedly complete then  $\mathcal{C}(\alpha\text{-sb}) = \mathcal{C}(\alpha\text{-csb})$ .<sup>29</sup> This implies, by Lemma 2 part c), that the tests in (A.9) are admissible and similar-on-the-boundary.

REMARK 12: If  $\Theta_0 = \text{Bd}\Theta_0$ , Lemma 3 implies that the  $\alpha$ -level ECS test in (A.9) is an element of

$$\arg \min_{\phi \in \mathcal{C}(\alpha\text{-sb})} \tau \int_{\text{int}\Theta_1} R(\phi, \theta) dw_1(\theta) + (1 - \tau) \int_{\text{Int}\Theta_0} \alpha dw_0(\theta),$$

hence, the  $\alpha$ -level ECS test maximizes

$$\text{WAP}(\phi, w) = \int_{\Theta_1} \left( \phi(x) f(x, \theta) \right) dw_1(\theta)$$

subject to

$$\int_{\mathbf{X}} \phi(x) f(x, \theta) dx = \alpha \quad \forall \theta \in \Theta_0$$

**LEMMA 3:** Let  $\phi_{\text{ECS}}$  be defined as in (A.9), then  $\phi_{\text{ECS}} \in M(\tau, w_1, w_0)$ .

<sup>29</sup>See also the Lehmann and Scheffe's Theorem (Linnik (1968) Chapter 4, p. 67).

PROOF: Let

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

Fubini's theorem (L2a-Step 1) implies that  $\phi^* \in M(\tau, w_1, w_0, \alpha\text{-csb})$  if and only if  $\phi^*$  solves the problem:

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

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

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

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

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

I need to 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  $M(\tau, w_1, w_0)$ —provided  $c(x_2, \alpha)$  is defined as the  $(1-\alpha)$  quantile of the random variable  $z(x_1, x_2)$ ,  $x_1 \sim f_{\text{Bd}}(x_1|x_2)$  for every  $x_2 \in \mathbf{X}_2$ . 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

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

subject to

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

except, perhaps, for  $x_2$  that belong to a set of measure zero under every  $h(x_2, \theta)$ ,  $\theta \in \text{Bd}\Theta_0$ . Hence, to show that  $\phi^*(x_1, x_2) \in M(\tau, w_1, w_0)$  it only remains to prove that  $\phi^*(x_1, x_2)$  is measurable. That is,  $\phi^*(x_1, x_2) \in \mathcal{C}(\alpha\text{-csb})$ . Assumption F0 implies that  $\phi^*(x_1, x_2)$  is continuous in  $x_1$ , for every  $x_2$ . Furthermore, since  $c(\cdot, \alpha)$  is measurable, then  $\phi^*(x_1, x_2)$  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  $\phi^* : \mathbf{X} \rightarrow [0, 1]$  is measurable. *Q.E.D.*

REMARK 13: Lemma 3 implies that when  $\text{Bd}\Theta_0 = \Theta_0$ ,  $\phi_{\text{ECS}}$  in (2.2) solves the problem:

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

subject to

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

which is equivalent to

$$\max_{\phi \in \mathcal{C}} \int_X \phi(x) \left( \int_{\theta} f(x, \theta) dw_1(\theta) \right) dx$$

subject to

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

and by Fubini's Theorem, this coincides with

$$\max_{\phi \in \mathcal{C}} \int_X \left( \phi(x) f(x, \theta) dx \right) dw_1(\theta) \equiv \text{WAP}(\phi, w_1)$$

subject to

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

#### A.4. Proof of Theorem 1

I present a proof of Theorem 1 for the tests in (A.9). That is, I consider a general hypothesis testing problem  $(\mathbf{X}, \Theta, f, \Theta_0)$  with  $\Theta_0$  closed and a boundary-sufficient statistic  $x_2$ . Given probability measures  $w_i$  over  $\text{int}\Theta_i$  define:

$$\phi_{\text{ECS}}(x_1, x_2) = \begin{cases} 1 & \text{if } z(x_1, x_2) \equiv \left[ \tau f_{w_1}^*(x_1, x_2) - (1 - \tau) f_{w_0}^*(x_1, x_2) \right] / f_{\text{Bd}}(x_1|x_2) > c(x_2; \alpha, w) \\ 0 & \text{i.o.c.} \end{cases},$$

with

$$f_{w_i}^*(x_1, x_2) \equiv \int_{\text{int}(\Theta_i)} f(x_1, x_2; \theta) dw_i(\theta), \quad i = 0, 1$$

and

$$c(x_2; \alpha) \equiv \arg \min_{q \in \mathbb{R}} \int \rho_{(1-\alpha)}(z(x_1, x_2) - q) f_{\text{Bd}}(x_1|x_2) dx_1,$$

where  $\rho_{1-\alpha}(u) = u[(1-\alpha) - \mathbb{1}\{u < 0\}]$  denotes the check function.

REMARK 14: Theorem 1 for the ECS tests in (2.2) defined for two-sided problems follows by noting that for two-sided testing problems with a nuisance parameter  $f_0^* = 0$  as  $\text{int}\Theta_0 = \emptyset$ .

PROOF OF PART I): Lemma 3 implies that  $\phi_{\text{ECS}}$  in (A.9) is an element of the set:

$$M(\tau, w_1, w_0, \alpha\text{-csb}) \equiv \arg \min_{\phi \in \mathcal{C}(\alpha\text{-csb})} \tau \int_{\text{int}\Theta_1} R(\phi, \theta) dw_1(\theta) + (1 - \tau) \int_{\text{int}\Theta_0} R(\phi, \theta) dw_0(\theta),$$

Theorem 4.3.2 in Lehmann and Romano (2005) and Assumption F2 imply  $\mathcal{C}(\alpha\text{-sb}) = \mathcal{C}(\alpha\text{-csb})$ . Hence,  $\phi_{\text{ECS}}$  is an element of:

$$M(\tau, w_1, w_0, \alpha\text{-sb}) \equiv \arg \min_{\phi \in \mathcal{C}(\alpha\text{-sb})} \tau \int_{\text{int}\Theta_1} R(\phi, \theta) dw_1(\theta) + (1 - \tau) \int_{\text{int}\Theta_0} R(\phi, \theta) dw_0(\theta),$$

If  $w_i(\beta, \pi)$  has full-support on  $\text{int}\Theta_i$ , Lemma 2c implies the ECS in (A.9) is admissible and similar on the boundary.

*Q.E.D.*

PROOF OF PART II): The proof is based on the essentially complete class theorem. See Theorem 2.9.2 and 2.10.3 in Ferguson (1967). See also Le Cam (1986), Chapter 2, Theorem 1.

OUTLINE: The main idea is that any average risk minimizers inside the class of similar-on-the-boundary tests is an ECS. Since  $\mathcal{C}(\alpha\text{-sb})$  is compact (in an appropriate topology) and the risk function is continuous (in the same topology) the essentially complete class theorem applies: every admissible similar-on-the-boundary test is risk equivalent to an extended Bayes test. Extended Bayes tests on  $\mathcal{C}(\alpha\text{-sb})$  are the same as extended ECS tests

PROOF: Note first that the class of all tests,  $\mathcal{C}$ , is essentially complete (as it contains all tests). Note that under assumptions F1, F2, the set  $\mathcal{C}(\alpha\text{-sb}) = \mathcal{C}(\alpha\text{-sb}) \subset \mathcal{C}$  is weak\* compact by Lemma 1. In addition, the risk function of the testing problem  $R(\phi, \theta)$  in Section 2 is—by definition of weak\* topology—continuous

(in  $\phi$ ) for all  $\theta \in \Theta$ . This verifies the assumptions of Theorem 2.9.2 in Ferguson (1967).

Theorem 2.10.3 in Ferguson (1967) implies that the set of *extended Bayes tests* in  $\mathcal{C}(\alpha - \text{sb})$  is essentially complete. This is, for any other test  $\phi \in \mathcal{C}$  there is a test  $\phi^*$  extended Bayes in  $\mathcal{C}(\alpha - \text{sb})$  such that:

$$R(\phi^*, \theta) \leq R(\phi, \theta)$$

for all  $\theta$ . Now if  $\phi$  is admissible  $R(\phi^*, \theta) \leq R(\phi, \theta)$  for all  $\theta$  implies that  $R(\phi^*, \theta) = R(\phi, \theta)$ . It remains to show that the extended Bayes test in  $\mathcal{C}(\alpha - \text{sb})$  are extended ECS.

By definition,  $\phi$  is an extended Bayes test in  $\mathcal{C}(\alpha - \text{sb})$  if for all  $\epsilon$  there exists a measure  $w_\epsilon(\beta, \pi)$  supported on a non-empty subset of  $\Theta \setminus \text{Bd}\Theta_0$  such that:

$$\int_{\Theta \setminus \text{Bd}\Theta_0} R(\phi) dw_\epsilon(\theta) \leq \left[ \inf_{\phi \in \mathcal{C}(\alpha - \text{sb})} \int_{\Theta \setminus \text{Bd}\Theta_0} R(\phi, \theta) dw_\epsilon(\theta) \right] + \epsilon$$

Suppose  $\text{int}\Theta_0 \neq \emptyset$  and  $w_\epsilon(\text{int}\Theta_0) > 0$ . Note that:

$$\begin{aligned} \int_{\Theta \setminus \text{Bd}\Theta_0} R(\phi, \theta) dw_\epsilon(\theta) &= \int_{\Theta_1 \cup \text{int}\Theta_0} R(\phi, \theta) dw_\epsilon(\theta) \\ &= \int_{\text{int}\Theta_1 \cup \text{int}\Theta_0} R(\phi, \theta) dw_\epsilon(\theta) \\ &= \int_{\text{int}\Theta_1} R(\phi, \theta) dw_\epsilon(\theta) + \int_{\text{int}\Theta_0} R(\phi, \theta) dw_\epsilon(\theta) \\ &= w_\epsilon(\text{int}\Theta_1) \int_{\text{int}\Theta_1} R(\phi, \theta) \frac{dw_\epsilon(\theta)}{w_\epsilon(\text{int}\Theta_1)} + w_\epsilon(\text{int}\Theta_0) \int_{\text{int}\Theta_0} R(\phi, \theta) \frac{dw_\epsilon(\theta)}{w_\epsilon(\text{int}\Theta_0)} \\ &= \tau \int_{\text{int}\Theta_1} R(\phi, \theta) dw_{1,\epsilon}(\theta) + (1 - \tau) \int_{\text{int}\Theta_0} R(\phi, \theta) dw_{0,\epsilon}(\theta) \end{aligned}$$

where  $\tau \equiv w_\epsilon(\text{int}\Theta_1)$  and  $w_{i,\epsilon}(A) \equiv w_\epsilon(A \cap \text{int}\Theta_i) / w_\epsilon(\text{int}\Theta_i)$ , for measurable  $A$ . Lemma 3 thus implies that if  $\phi$  is extended Bayes in  $\mathcal{C}(\alpha - \text{csb})$  then for all  $\epsilon$  there exists probability measures  $w_{1,\epsilon}, w_{0,\epsilon}$  supported on subsets of  $\text{int}\Theta_1$  and  $\text{int}\Theta_0$  such that:

$$\begin{aligned} \int_{\Theta \setminus \text{Bd}\Theta_0} R(\phi) dw_\epsilon(\theta) &\leq \left[ \inf_{\phi \in \mathcal{C}(\alpha - \text{sb})} \tau \int_{\text{int}\Theta_1} R(\phi, \theta) dw_{1,\epsilon}(\theta) + (1 - \tau) \int_{\text{int}\Theta_0} R(\phi, \theta) dw_{0,\epsilon}(\theta) \right] + \epsilon \\ &= \tau \int_{\text{int}\Theta_1} R(\phi_{\text{ECS}}^{w_{i,\epsilon}}, \theta) dw_{1,\epsilon}(\theta) + (1 - \tau) \int_{\text{int}\Theta_0} R(\phi_{\text{ECS}}^{w_{i,\epsilon}}, \theta) dw_{0,\epsilon}(\theta) + \epsilon \end{aligned}$$

where  $\phi_{\text{ECS}}^{w_{i,\epsilon}}$  denotes the ECS test in (A.9), given weights  $w_{i,\epsilon}$ ,  $i = 0, 1$ .

*Q.E.D.*

#### A.5. Lemma M

Let  $W$  be a  $s \times s$ ,  $s \in \mathbb{N} \setminus \{1\}$ , 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  $B^{-1/2}BB^{-1/2} = \mathbb{I}_{s_2}$ . 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 M:**  $DWD' = \mathbb{I}_s$

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.*

## APPENDIX B: LIMITING STATISTICAL MODEL FOR EXTREMUM PROBLEMS

### B.1. Proof of Proposition 1

Intuition: The centrality parameter for  $D_\beta^*$  is derived by adding up the first and second ‘‘estimation errors’’. By assumption AQ4.1,  $\sqrt{n}\Delta_{m^*}^\pi(\pi_n, \pi_0)$  converges in probability to:

$$\mathbb{E}_\gamma \left[ m_\beta^*(w, 0, \pi) - m_\beta^*(w, 0, \pi_0) \right]' b,$$

while the asymptotic bias term associated to  $\sqrt{n}\Delta_{m^*}^\beta(\beta_n, \widehat{\beta}(\pi_0))$  is given by:

$$-\mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)]' [a^*(\pi_0, \gamma)' b - b].$$

The centrality parameter of the gaussian model in Equation 3.6 corresponds to the sum of the two previous equations.

The centrality parameter for  $\beta^*$  is derived using Assumption AQ4.3. The adjustment in the asymptotic variance of  $(D_\beta^*, \beta^*)$  comes from AQ4.2.

The intuition above is formalized in the proof of Proposition 1.

PROOF: Define

$$\begin{pmatrix} \eta_{1n} \\ \eta_{2n} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n m(W_{in}, \pi_0) - m^*(W_{in}, \beta_n, \pi_n) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n f(W_{in}, \pi_0) - f^*(W_{in}, \beta_n, \pi_n) \end{pmatrix}$$

Let

$$a^*(\pi_0, \gamma) \equiv \mathbb{E}_\gamma [f_\beta(w, 0, \pi)] a(\pi_0, \gamma)' \in \mathbb{R}^{d_\beta \times d_\beta}$$

Note first that

$$\begin{aligned}
\sqrt{n}\widehat{\beta}^*(\pi_0) &= g_2(W^n)\eta_{2n} + g_2(W_2)\frac{1}{\sqrt{n}}\sum_{i=1}^n f^*(W_{in}, \beta_n, \pi_n) \\
&= a(\pi_0, \gamma)\eta_{2n} + a(\pi_0, \gamma)\mathbb{E}_\gamma[f_\beta(W_{in}, 0, \pi)]'b + o_{p, \gamma_n}(1) \\
&\quad (\text{where I have used AQ4.3}) \\
\text{(B.1)} \quad &= a(\pi_0, \gamma)\eta_{2n} + a^*(\pi_0, \gamma)'b + o_{p, \gamma_n}(1)
\end{aligned}$$

Note then that

$$\begin{aligned}
\sqrt{n}D_\beta^*Q_n(0, \pi_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n m(W_{in}, \pi_0) - m^*(W_{in}, \widehat{\beta}(\pi_0), \pi_0) \\
&= \left[ \frac{1}{\sqrt{n}}\sum_{i=1}^n m(W_{in}, \pi_0) - m^*(W_{in}, \beta_n, \pi_n) \right] + \sqrt{n}\Delta_{m^*}^\pi(\pi_n, \pi_0) + \sqrt{n}\Delta_{m^*}^\beta(\beta_n, \widehat{\beta}(\pi_0)) \\
&\quad (\text{i.e., AQ3 plus the first and second estimation errors}) \\
&= \eta_{1n} + \mathbb{E}_\gamma \left[ m_\beta^*(w, 0, \pi) - m_\beta^*(w, 0, \pi_0) \right]' b - \mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)]' \sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n) + o_{p, \gamma_n}(1) \\
&\quad (\text{where I have used AQ4.1 and AQ4.2}) \\
&= \eta_{1n} + \mathbb{E}_\gamma \left[ m_\beta^*(w, 0, \pi) - m_\beta^*(w, 0, \pi_0) \right]' b - \mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)]' a(\pi_0, \gamma)\eta_{2n} \\
&\quad - \mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)]' [a^*(\pi_0, \gamma)'b - b] + o_{p, \gamma_n}(1) \\
&\quad (\text{where I have used Equation B.1})
\end{aligned}$$

Therefore,

$$\begin{aligned}
\begin{pmatrix} \sqrt{n}D_\beta^*Q_n(0, \pi_0) \\ \sqrt{n}\widehat{\beta}^*(\pi_0) \end{pmatrix} &= \begin{bmatrix} \mathbb{I}_{d_\beta} & -\mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)]' a^*(\pi_0, \gamma) \\ 0 & \mathbb{I}_{d_\beta} \end{bmatrix} \begin{pmatrix} \eta_{1n} \\ \eta_{2n} \end{pmatrix} \\
&\quad + \begin{pmatrix} \mathbb{E}_\gamma [m_\beta^*(w, 0, \pi)] - a^*(\pi_0, \gamma)\mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)] \\ \mathbb{E}_\gamma [f_\beta(W_{in}, 0, \pi)]' b \end{pmatrix}' b \\
&\quad + o_{p, \gamma_n}(1)
\end{aligned}$$

The result then follows. Q.E.D.

## B.2. Proof of Proposition 2

PROOF: The random vector  $(D_\beta^*, \beta^*)$  is multivariate normal with centrality parameter

$$\begin{pmatrix} \left[ \mathbb{E}_\gamma [m_\beta^*(w, 0, \pi)] - a^*(\pi_0, \gamma)\mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)] \right]' b \\ \mathbb{E}_\gamma [f_\beta(W_{in}, 0, \pi)]' b \end{pmatrix}$$

and covariance matrix  $\Sigma(\pi_0, \gamma)$ . Hence, the linear transformation given by  $C_0(D_\beta^*, \beta^*)$  is multivariate normal with centrality parameter:

$$C_0 \begin{pmatrix} \left[ \mathbb{E}_\gamma [m_\beta^*(w, 0, \pi)] - a^*(\pi_0, \gamma)\mathbb{E}_\gamma [m_\beta^*(w, 0, \pi_0)] \right]' b \\ \mathbb{E}_\gamma [f_\beta(W_{in}, 0, \pi)]' b \end{pmatrix},$$

and covariance matrix  $C_0\Sigma(\pi_0, \gamma)C_0'$ , where

$$C_0 \equiv \begin{pmatrix} \left( \Sigma_{DD}^0 \right)^{-1/2} & \mathbf{0} \\ -B^{-1/2}R'_{12} \left( \Sigma_{DD}^0 \right)^{-1/2} & B^{-1/2} \left( \Sigma_{\beta\beta}^0 \right)^{-1/2} \end{pmatrix}.$$

Simple algebra shows that

$$C_0 \left( \begin{bmatrix} \mathbb{E}_\gamma[m_\beta^*(w, 0, \pi)] - a^*(\pi_0, \gamma)\mathbb{E}_\gamma[m_\beta^*(w, 0, \pi_0)] \\ \mathbb{E}_\gamma[f_\beta(W_{in}, 0, \pi)]'b \end{bmatrix}' b \right) = \begin{pmatrix} \mu_1^*(\pi_0, b, \pi) \\ \mu_2^*(\pi_0, b, \pi) \end{pmatrix}$$

The statistical model in Equation 3.8 evaluated at the boundary of the null set ( $\pi = \pi_0$ ) equals:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{2d_\beta} \left( \begin{pmatrix} \mathbf{0} \\ \mu_2^*(\pi_0, b, \pi_0) \end{pmatrix}, \mathbb{I}_{2d_\beta} \right).$$

as

$$\left[ \mathbb{E}_\gamma[m_\beta^*(w, 0, \pi_0)] - a^*(\pi_0, (0, \pi_0, \phi))\mathbb{E}_\gamma[m_\beta^*(w, 0, \pi_0)] \right] = \mathbf{0}.$$

and, by construction,  $C_0\Sigma(\pi_0, (0, \pi_0, \gamma))C_0' = \mathbb{I}_{2d_\beta}$ . Therefore, when  $\pi = \pi_0$ ,  $X_1 \sim \mathcal{N}_{2d_\beta}(\mathbf{0}, \mathbb{I}_{d_\beta})$  independently of  $X_2$ . The latter implies  $X_2$  is boundary sufficient.

Note also that whenever  $\pi = \pi_0$ ,  $X_2 \sim \mathcal{N}_{d_\beta}(\mu_2^*(\pi_0, b, \pi_0))$ . Hence if the image of the function

$$\mu_2^*(\pi_0, \cdot, \pi_0) : \mathbb{R}^{d_\beta} \rightarrow \mathbb{R}^{d_\beta}$$

contains a set of the form:

$$\{\mu_2 \in \mathbb{R}^{d_\beta} \mid \text{s.t. } a < \mu_{2,j} < b \quad \forall j = 1 \dots d_\beta\}$$

Theorem 4.3.1 of [Lehmann and Romano \(2005\)](#) implies  $X_2$  is boundedly complete.

*Q.E.D.*

### B.3. Proof of Result 3

PROOF: Let  $P_n(b, \pi, \phi_n)$  denote the distribution of the random vector  $h_n(W_n)$  that takes values in  $\mathcal{S} \equiv \mathbb{R}^{2d_\beta}$ , a complete metric separable space as required by [Müller \(2011\)](#). For a fixed  $\phi^*$ , the relevant set of models—or sequences  $\phi_n$ —is given by

$$\mathcal{M}(\phi^*) = \{\{\phi_n\}_{i=1}^\infty \subset \Phi \mid P_n(b, \pi, \phi_n) \xrightarrow{d} P(b, \pi; \phi^*)\},$$

where the weak convergence holds pointwise for all  $(b, \pi)$ . As mentioned before,  $P(b, \pi; \phi^*)$  is a multivariate normal probability measure with  $\mathbb{I}_{2d_\beta}$  covariance matrix and mean vector partitioned as  $\mu_1^*(\pi_0, b, \pi; \phi^*)$ ,  $\mu_2^*(\pi_0, b, \pi; \phi^*)$ .

For a fixed  $\phi^*$ ,  $P(b, \pi; \phi^*)$  is equivalent to  $P(b', \pi'; \phi^*)$ , as both measures are equivalent to the Lebesgue measure in  $\mathbb{R}^{2d_\beta}$ . Therefore,  $\mu_p$  as defined in [Müller \(2011\)](#) (pg. 400) can be any probability measure with parameters  $(b_0, \pi_0)$ . Consequently, if  $\phi_{\text{ECS}}$  is lebesgue almost everywhere continuous it follows that it is also  $\mu_p$ -almost everywhere continuous. Therefore, the premise of Theorem 1 in [Müller \(2011\)](#) is verified.

By construction, the ECS test maximizes WAP subject to a similarity constraint. Therefore, part *i* of Theorem 1 implies that ECS test, evaluated at sample analogues, is asymptotically similar. Part *i*) also implies that:

$$\lim_{n \rightarrow \infty} \text{WAP}_n(\phi_{\text{ECS}}^n, m) = \text{WAP}(\phi_{\text{ECS}})$$

Part *ii*) of Theorem 1 in [Müller \(2011\)](#) yields the remaining inequality.

*Q.E.D.*



SUPPLEMENTARY MATERIAL  
Efficient Conditionally Similar Tests:  
Finite-Sample Theory and Large-Sample Applications

## APPENDIX C: GAUSSIAN IV EXAMPLE

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

$$\rho = (a'\Omega^{-1}a)^{1/2}(\beta'Z'Z\beta)^{1/2}, \quad \phi = C_0a/(a'\Omega^{-1}a)^{1/2}, \quad \omega = (Z'Z)^{1/2}\beta/(\beta'Z'Z\beta)^{1/2}$$

where  $a \equiv [\pi, 1]'$ , and

$$C_0 \equiv \begin{pmatrix} (b'_0\Omega b_0)^{-1/2}b'_0 \\ (a'_0\Omega^{-1}a_0)^{-1/2}a'_0\Omega^{-1} \end{pmatrix}$$

$$b_0 = [1, -\pi_0]' \quad a_0 = [\pi_0, 1]'$$

REMARK: The value  $\pi_0$  and the reduced-form covariance matrix  $\Omega$  impose a restriction on the possible values for the parameter  $\phi$ . Note first that  $\Omega C'_0 C_0 = \mathbb{I}_2$ . Therefore:

$$\begin{aligned} (0, 1)\Omega C'_0 \phi &= (0, 1)\Omega C'_0 C_0 a / a'\Omega^{-1}a'^{1/2} \\ &= (0, 1)\mathbb{I}_2(\beta, 1)' / (a\Omega^{-1}a')^{1/2} \\ &= 1 / (a\Omega^{-1}a')^{1/2} \geq 0 \end{aligned}$$

Hence, the parameter  $\phi$  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,

$$\begin{aligned} (0, 1)\Omega C'_0 &= \left( (0, 1)\Omega b_0 / (b'_0\Omega b_0)^{1/2}, (0, 1)a_0 / (a'_0\Omega^{-1}a_0)^{1/2} \right) \\ &= \left( (0, 1)\Omega b_0 / (b'_0\Omega b_0)^{1/2}, (\omega_1^2\omega_2^2 - \omega_{12}^2)^{1/2} / (b'_0\Omega b_0)^{1/2} \right) \\ &= \left( (0, 1)\Omega b_0 / (b'_0\Omega b_0)^{1/2}, (\omega_1^2\omega_2^2 - (\omega_{12} - \pi_0\omega^2 + \pi_0\omega_s^2)^2)^{1/2} / (b'_0\Omega b_0)^{1/2} \right) \\ &= \left( \omega_2 r(\pi_0), \omega_2 \sqrt{1 - r^2(\pi_0)} \right) \end{aligned}$$

where  $r(\pi_0)$  corresponds to the structural correlation implied by  $\pi_0$ :

$$r(\pi_0) = (0, 1)\Omega b_0 / \left[ (b'_0\Omega b_0)^{1/2}\omega_2 \right]$$

Thus, the domain for the parameter  $\phi$  in the canonical model is given by:

$$\Theta = \left\{ (\rho^2, \phi) \in \mathbb{R}_+ \times \mathcal{S}^1 : r(\pi_0)\phi_1 + \sqrt{1 - r^2(\pi_0)}\phi_2 \geq 0 \right\}$$

**DERIVATION OF THE INTEGRATED LIKELIHOODS:** Let:

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

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

**STEP 1:** (Integrate  $\omega$ ) Note that:

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

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,

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

$c_2$  is a non-negative constant.

**STEP 2:** (Integrate  $\rho$ ) By assumption  $\rho \sim \sqrt{\chi_k^2}$  independently of  $\phi$  and  $\omega$ . The latter implies that the density of  $\rho$  is given by:

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

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

$$\begin{aligned} & \int_{\mathbb{R}^+} \tilde{f}(S, T; \rho, \phi) m_1(\rho) d\rho \\ &= a_2(Q) \int_{\mathbb{R}^+} \left( \exp(-\rho^2/2) \int_{\mathcal{S}^{k-1}} \exp\left(\left[(S, T)\rho\phi\right]'\omega\right) d\lambda_{\mathcal{S}^{k-1}}(\omega) \right) m_1(\rho) d\rho \\ &= a_2(Q) \int_{\mathcal{S}^{k-1}} \left( \int_{\mathbb{R}^+} \exp\left(\left[(S, T)\rho\phi\right]'\omega\right) m_1(\rho) \exp(-\rho^2/2) d\rho \right) d\lambda_{\mathcal{S}^{k-1}}(\omega) \\ &= a_3(Q) \int_{\mathcal{S}^{k-1}} \left( \int_{\mathbb{R}^+} \exp\left(\left[(S, T)\rho\phi\right]'\omega\right) \exp(-\rho^2/2b^2)\rho^{k-1} dp \right) d\lambda_{\mathcal{S}^{k-1}}(\omega) \\ & \hspace{15em} \text{(by definition of } m_1, b^2 \equiv [1/2] \text{)} \\ &= a_3(Q) \int_{\mathcal{S}^{k-1}} \left( \int_{\mathbb{R}^+} \exp\left(\left[(S, T)\phi\right]'\rho\omega\right) \exp(-(\rho\omega)'(\rho\omega)/2b^2)\rho^{k-1} dp \right) d\lambda_{\mathcal{S}^{k-1}}(\omega) \end{aligned}$$

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

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

Theorem 5.2.2, p. 86 in [Stroock \(1999\)](#) implies:

$$\begin{aligned} \int_{\mathbb{R}^+} \tilde{f}(S, T; \rho, \phi) m_1(\rho) d\rho &= a_3(Q) \int_{\mathbb{R}^k} \exp\left(\left[(S, T)\phi\right]'\omega\right) \exp(-\omega'\omega/2b^2) d\omega \\ &= a_4(Q) \exp\left(\frac{b^2}{2}\phi'Q\phi\right) \end{aligned}$$

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  $\phi$ ) For simplicity, I will assume throughout the remaining part of this section that  $r(\pi_0) \geq 0$ . Consider the mapping  $m: [-\pi i, \pi i] \rightarrow \mathcal{S}^1$  given by  $m(\theta) = [-\sin(\theta), \cos(\theta)]$ . Note that  $m(\cdot)$  evaluated at  $-\pi i$  (-3.1415 ...) gives the point (0, 1) in the unit circle.<sup>30</sup> As  $\theta$  increases, the mapping  $m(\cdot)$  traces  $\mathcal{S}^1$  counter-clock wise. Therefore,

$$\mathcal{S}^1(r(\pi_0)) = \{\phi \in \mathcal{S}^1 \mid r(\pi_0)\phi_1 + \sqrt{1-r^2(\pi_0)}\phi_2 \geq 0\}$$

can be expressed as:

$$\begin{aligned} & \left\{ \theta \in [-\pi i, \pi i] : r(\pi_0)(-\sin(\theta)) + \sqrt{1-r^2(\pi_0)}\cos(\theta) \geq 0 \right\} \\ &= \left[ \tan^{-1}\left(\sqrt{1-r(\pi_0)^2}/r(\pi_0)\right) - \pi, \tan^{-1}\left(\sqrt{1-r(\pi_0)^2}/r(\pi_0)\right) \right] \\ &= [\pi i_0, \pi i_0 + \pi i], \quad \text{where } \pi i_0 \equiv \tan^{-1}\left(\sqrt{1-r(\pi_0)^2}/r(\pi_0)\right) - \pi \end{aligned}$$

and  $\pi i_0 < 0$  when  $r(\pi_0) > 0$ . The parameter  $\phi \sim \mathcal{U}(\mathcal{S}^1(r(\pi_0)))$  if and only if  $\theta \sim \mathcal{U}[\pi i_0, \pi i_0 + \pi i]$ . Define:

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

<sup>30</sup>Throughout the following pages, " $\pi i$ " will be used to denote 3.1416... This is done to avoid confusion with the parameter of interest,  $\pi$

The following Lemma is crucial for the derivation of the two-sided ECS test in the IV model. Let

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

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

**LEMMA 1-IV:** . Let  $pi_l, pi_u$  belong to the interval  $[pi_0, pi + pi_0]$ . Then

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

where  $\Phi_{[0, 2\pi]}^{VM}$  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*

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

$I_0(\cdot)$  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:

$$e_{max} \equiv \begin{cases} (L, S'T)' / \sqrt{L^2 + (S'T)^2} & \text{if } r(\pi_0)L + \sqrt{1 - r(\pi_0)}S'T > 0 \\ -(L, S'T)' / \sqrt{L^2 + (S'T)^2} & \text{if } r(\pi_0)L + \sqrt{1 - r(\pi_0)}S'T \leq 0 \end{cases}$$

Note that  $e_{max}$  is the maximum eigenvalue of the matrix  $Q$  adjusted to belong to the domain  $\mathcal{S}^1(r(\pi_0))$ . Define  $\widehat{\theta} \in [pi_0, pi]$  implicitly 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_{[p_i_l, p_i_u]}^*(S, T) &= a_4(Q) \frac{1}{p_i_u - p_i_l} \int_{p_i_l}^{p_i_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}{p_i_u - p_i_l} \int_{\widehat{\theta}_{max} - p_i_u}^{\widehat{\theta}_{max} - p_i_l} \exp\left(\frac{b^2}{2} \left[ \zeta_{max} \cos^2(\theta) + \zeta_{min} \sin^2(\theta) \right]\right) d\theta \\
&\quad \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} - p_i_u}^{\widehat{\theta}_{max} - p_i_l} \exp\left(\frac{b^2}{2} \left[ (\zeta_{max} - \zeta_{min}) \cos^2(\theta) \right]\right) d\theta \\
&\quad \text{(as } \sin^2(\theta) + \cos^2(\theta) = 1\text{)} \\
&= 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} - p_i_u}^{\widehat{\theta}_{max} - p_i_l} \exp\left(\frac{b^2}{4} (\zeta_{max} - \zeta_{min}) \cos(2\theta)\right) d\theta \\
&\quad \text{(as } 2 \cos^2(\theta) - 1 = \cos(2\theta)\text{)} \\
&= 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(p_i_u - p_i_l)} \\
&= \int_{2(\widehat{\theta}_{max} - p_i_u)}^{2(\widehat{\theta}_{max} - p_i_l)} \exp\left(\kappa(Q) \cos(\theta)\right) d\theta \\
&\quad \text{(where we have used the change of variable } \tilde{\theta} = 2\theta\text{)} \\
&\quad \left(\kappa(Q) \equiv \frac{b^2}{4} (\zeta_{max} - \zeta_{min})\right) \\
&= a_4(Q) \exp\left(\frac{b^2}{4} (\zeta_{max} + \zeta_{min})\right) \frac{1}{2(p_i_u - p_i_l)} \\
&\quad \int_{2(p_i_l - p_i_0)}^{2(p_i_u - p_i_0)} \exp\left(\kappa(Q) \cos(\theta - 2(\widehat{\theta}_{max} - p_i_0))\right) du \\
&\quad \text{(where we have used the change of variable } u = 2(\widehat{\theta}_{max} - p_i_0) - \theta\text{)}
\end{aligned}$$

Note that  $\widehat{\theta}_{max} - p_i_0 \in [0, p_i]$ . Therefore,  $\mu(Q) \equiv 2(\widehat{\theta}_{max} - p_i_0) \in [0, 2p_i]$ . Using the definition of the Von-Mises distribution (supported on  $[0, 2p_i]$ ) in [Mardia and Jupp \(2000\)](#) it follows that:

$$\begin{aligned}
f_{[p_i_l, p_i_u]}^*(S, T) &= a_4(Q) \exp\left(\frac{b^2}{4} (\zeta_{max} + \zeta_{min})\right) \frac{p_i}{p_i_u - p_i_l} I_0(\kappa(Q)) \left[ \Phi_{[0, 2p_i]}^{VM} \left( 2(\pi_u - \pi_0) | \kappa(Q), \mu(Q) \right) \right. \\
&\quad \left. - \Phi_{[0, 2p_i]}^{VM} \left( 2(p_i_l - p_i_0) | \kappa(Q), \mu(Q) \right) \right]
\end{aligned}$$

$\mu(Q)$  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 1:** From Lemma 1-IV above it follows that the integrated likelihood for independent weights:

$$\phi \sim \mathcal{U}(S^1(r(\pi_0))) \quad \omega \sim \mathcal{U}(S^k) \quad \rho \sim \sqrt{\chi_k^2}$$

is given by:

$$f_w^*(S, T) = f_{[p_i_0, p_i_0 + p_i]}^* = a_1 \exp\left(-\frac{1}{2} [S'S + T'T]\right) \exp\left(\frac{b^2}{4} (S'S + T'T)\right) I_0(\kappa(Q))$$

where  $a_1$  is a non-negative constant,  $b^2 = 1/2$ , 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:

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

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:

$$S'S - T'T + 8 \ln \left[ I_0 \left( \frac{1}{8} \left( (S'S - T'T)^2 + 4(S'T)^2 \right)^{1/2} \right) \right]$$

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  $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$ .

#### APPENDIX D: LINEAR REGRESSION MODEL WITH A SIGN RESTRICTION

[Elliott et al. \(2013\)](#) consider inference about a linear regression coefficient ( $\beta$ ) when the sign of a control coefficient ( $\delta$ ) is known.

**a) ECONOMETRIC MODEL:** Let  $(z_{i1}, z_{i2}, \epsilon_i)$  be a triplet of real-valued random variables. The model of interest is given by:

$$(D.1) \quad y_i = \pi z_{i1} + \beta z_{i2} + \epsilon_i, \quad \{z_{i1}, z_{i2}, \epsilon_i\}_{i=1}^n \text{ i.i.d.}$$

with matrix form representation:

$$(D.2) \quad Y = Z \alpha + \varepsilon, \quad \alpha = (\pi, \beta)'$$

The objective is to test  $\pi \leq \pi_0$  or  $\pi = \pi_0$ , taking into account the sign restriction of the nuisance parameter  $\beta \geq 0$ .

**b) DISTRIBUTIONAL ASSUMPTIONS FOR LINEAR REGRESSION:** To work within the context of a finite-sample statistical model let  $(z_{i1}, z_{i2})$  be a vector of non-stochastic regressors and let  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ , with  $\sigma^2$  known.

**c) CANONICAL REPRESENTATION OF THE PROBLEM:** Using a sufficiency argument, the inference problem can be embedded into the following statistical model

$$\begin{pmatrix} \widehat{\pi}_{\text{OLS}} - \pi_0 \\ \widehat{\delta}_{\text{OLS}} \end{pmatrix} \sim \mathcal{N}_2 \left( \begin{pmatrix} \pi - \pi_0 \\ \beta \end{pmatrix}, \Omega \right),$$

where

$$(\widehat{\pi}_{\text{OLS}}, \widehat{\beta}_{\text{OLS}})' \equiv (Z'Z)^{-1} Z'Y, \quad \Omega \equiv \sigma^2 (Z'Z)^{-1}.$$

Thus, testing in the linear regression model admits the canonical representation:

$$(D.3) \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \equiv \begin{bmatrix} (\widehat{\pi}_{\text{OLS}} - \pi_0)/\omega_1 \\ ((\widehat{\beta}_{\text{OLS}}/\omega_2) - r(\widehat{\pi}_{\text{OLS}} - \pi_0/\omega_1))/(\sqrt{1-r^2}) \end{bmatrix} \sim \mathcal{N}_2 \left( \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \mathbb{I}_2 \right),$$

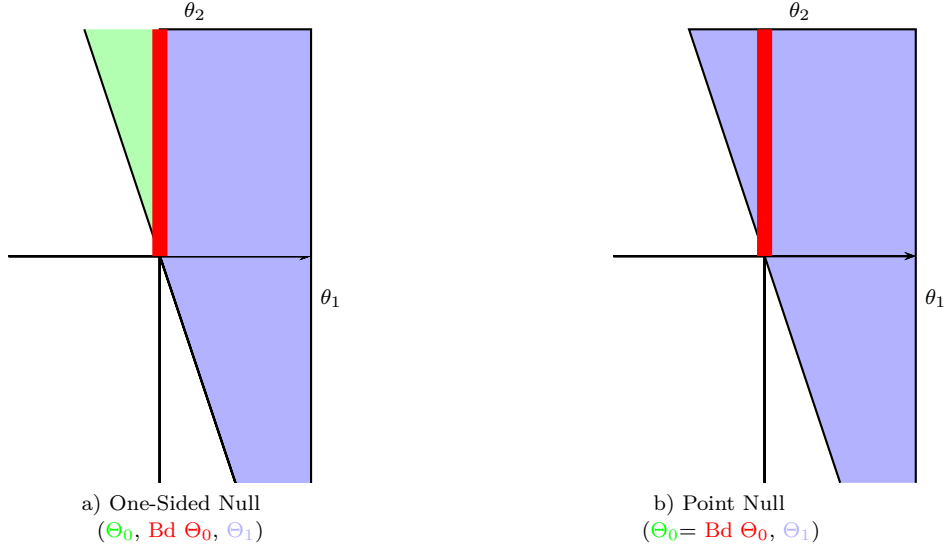
with  $r = \omega_{12}/\omega_1\omega_2$ .<sup>31</sup> Note that I have used the rotation matrix:

$$\begin{pmatrix} 1/\omega_1 & 0 \\ -r/\omega_1\sqrt{1-r^2} & 1/\omega_2\sqrt{1-r^2} \end{pmatrix}.$$

to standardize the OLS coefficients. The sample space in the canonical model is  $\mathbb{R}^2$  and the parameter space is given by:

$$\Theta \equiv \{(\theta_1, \theta_2) \in \mathbb{R}^2 \mid r\theta_1 + \sqrt{1-r^2}\theta_2 \geq 0\}.$$

Graphically, the one-sided null ( $\theta_1 \leq 0$ ) and the two-sided null ( $\theta_1 = 0$ ):



**d) BOUNDARY SUFFICIENCY:** When the likelihood function of the canonical model is evaluated at any point of  $\text{Bd } \Theta_0$ , the following decomposition applies

$$f(x_1, x_2; \theta_1 = 0, \theta_2) = \underbrace{(2\pi)^{-1/2} \exp\left(-x_1^2/2\right)}_{f_{\text{Bd}}(x_1|x_2)} \underbrace{(2\pi)^{-1/2} \exp\left(-(x_2 - \theta_2)^2/2\right)}_{h(x_2, \theta_2)}.$$

Hence,  $x_2$  is a boundary-sufficient statistic.

The boundary-sufficient statistic  $x_2$  in this model is also boundedly complete. Note that  $\{\theta_2 \in \mathbb{R} \mid (\theta_1, \theta_2) \in \text{Bd}\Theta_0\} = \mathbb{R}_+$ . Consider the (exponential) family of probability distributions over  $\mathbb{R}$  given by  $\{\mathcal{N}_1(\theta_2, 1)\}_{\theta_2 \in \mathbb{R}_+}$ . Since  $\mathbb{R}_+$  contains a one-dimensional rectangle, Theorem 4.3.1 in [Lehmann and Romano \(2005\)](#) implies the family of distributions is complete.

**e1) WEIGHTS:** First, consider the problem

$$\mathbf{H}_0 : \theta_1 = 0 \quad \text{vs.} \quad \mathbf{H}_1 : \theta_1 \neq 0,$$

and the weights (in the canonical model) given by:

$$p_i(\theta_1, \theta_2; \lambda^2) \propto \exp(-\lambda^2\theta_1^2/2) \exp(-\lambda^2\theta_2^2/2) \mathbb{1}\{\Theta\},$$

That is, the restriction of the normal measure (centered at zero) with independent components to the parameter space  $\Theta$ .

<sup>31</sup>Since  $\sigma^2$  is known and  $Z$  is a non-stochastic matrix, then  $\Omega$ . The diagonal components of  $\Omega$  are denoted  $\omega_1^2, \omega_2^2$ . The off-diagonal component is  $\omega_{12}$ .

**RESULT 4:** *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \theta_1 = 0$  vs.  $\mathbf{H}_1 : \theta_1 \neq 0$  rejects the null hypothesis if the random variable*

$$x_1^2 + 2(1 + \lambda^2) \ln \left[ 1 - \Phi \left( - [rx_1 + \sqrt{1 - r^2}x_2] / \sqrt{1 + \lambda^2} \right) \right]$$

*is larger than its  $(1-\alpha)$  quantile (conditional on  $x_2$ ) evaluated under the distribution  $x_1 \sim \mathcal{N}(0, 1)$ .*

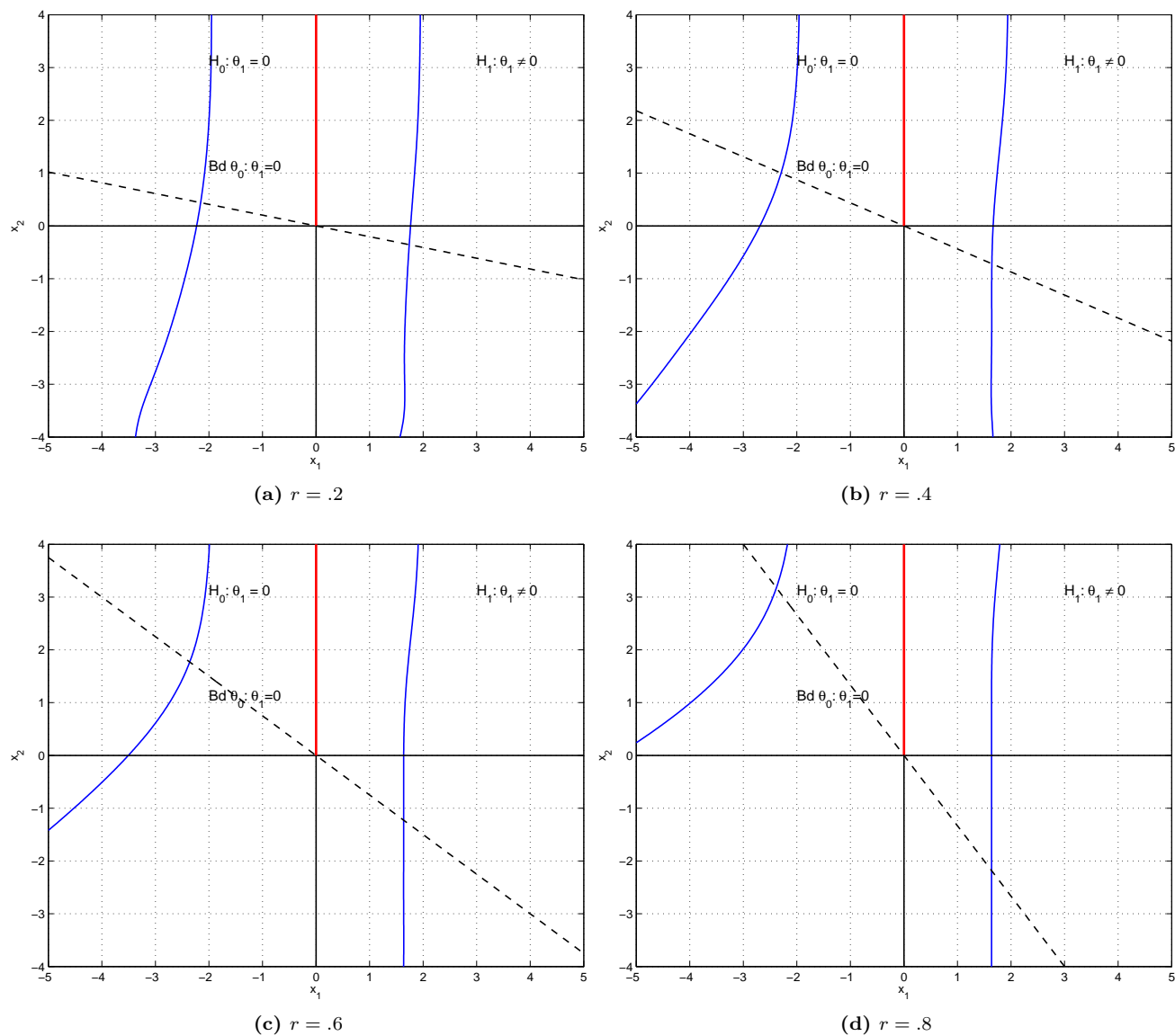
PROOF: See Sections D.1 and D.2

*Q.E.D.*

CRITICAL REGION FOR THE TWO-SIDED TEST Figure 6 presents the critical region of the point-null ECS test. The ECS test provides power gains by taking into account the shape of the parameter space. For example, when  $r = 0.8$  the rejection region for negative values of the boundary sufficient statistic is close to the region  $x_1 > 1.64$ .



**Figure 6:** Critical Region for the two-sided 5%-ECS Test



(RED) Boundary of the null hypothesis. (BLUE) A 10th degree polynomial is fitted to the conditional critical value functions using the command `polyfit` in Matlab. The domain is restricted to  $x_2 \in [-5, 5]$ . The test statistic is based on the expression in Appendix D.2. The command `ezplot` is used to graph the solution of the equation  $z(x_1, x_2; \lambda^2, r) - c(x_2; \lambda^2, r) = 0$ . (BLACK, DASHED) Boundary of the parameter space.

**e2)** WEIGHTS ONE-SIDED PROBLEM: Consider now the problem

$$\mathbf{H}_0 : \theta_1 \leq 0 \quad \text{vs.} \quad \mathbf{H}_1 : \theta_1 > 0,$$

and the weights (in the canonical model) given by the normal measure in **e1)** restricted to the null and to the alternative.

**RESULT 5:** *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \theta_1 \leq 0$  vs.  $\mathbf{H}_1 : \theta_1 > 0$  rejects the null hypothesis if the random variable  $z(x_1, x_2, \tau, \lambda^2)$*

$$\begin{aligned} & \tau \int_0^\infty \left( 1 - \Phi \left[ -\frac{\sqrt{1+\lambda^2}}{\sqrt{1-r^2}} \theta_1 r - \frac{x_2}{\sqrt{1+\lambda^2}} \right] \right) \exp(x_1 \theta_1) \exp\left(-\theta_1^2(1+\lambda^2)/2\right) d\theta_1 - \\ & (1-\tau) \int_{-\infty}^0 \left( 1 - \Phi \left[ -\frac{\sqrt{1+\lambda^2}}{\sqrt{1-r^2}} \theta_1 r - \frac{x_2}{\sqrt{1+\lambda^2}} \right] \right) \exp(x_1 \theta_1) \exp\left(-\theta_1^2(1+\lambda^2)/2\right) d\theta_1 \end{aligned}$$

is larger than its  $1-\alpha$  quantile (conditional on  $x_2$ ) evaluated under the distribution  $x_1 \sim \mathcal{N}(0, 1)$ .

PROOF: See Section D.3.

Q.E.D.

REMARK 15:  $(\tau, \lambda^2, r)$  The parameters  $\tau \in (0, 1)$  and  $\lambda^2 \in (0, \infty)$  are selected by the researcher. The parameter  $\tau$  controls the relative importance of the average rates of Type I and Type II error. The parameter  $\lambda^2$  controls the precision of the gaussian weight used to generate the ECS test. The scalar  $r$ —which corresponds to the correlation coefficient implied by the matrix  $\Omega$ —is taken as a primitive of the model.

REMARK 16: (*Implementation*) The implementation of the ECS test in Result 2 requires the following numerical routine. First,  $x_2$  is set equal to the sample realization of the boundary sufficient statistic  $X_2$ . Therefore,  $x_2$  is treated as a constant in the expression  $z(x_1, x_2, \tau, \lambda^2)$ . Second, the conditional quantiles of the the random variable  $z(X_1, x_2, \tau, \lambda^2)$  are obtained via Monte-Carlo. That is, one takes  $I$  independent draws using the *boundary conditional likelihood* and for each draw  $x_1^i$  one computes the statistic

$$z(x_1^i, x_2, \tau, \lambda^2).$$

In this model, the boundary conditional likelihood corresponds to the density of  $\mathcal{N}(0, 1)$ . So, sampling from this distribution is a simple task. Note that in order to evaluate  $z(x_1^i, x_2, \lambda^2, \tau)$  it is necessary to compute the following integrals

$$\begin{aligned} & \int_0^\infty \left( 1 - \Phi \left[ -\frac{\sqrt{1+\lambda^2}}{\sqrt{1-r^2}} \theta_1 r - \frac{x_2}{\sqrt{1+\lambda^2}} \right] \right) \exp(x_1^i \theta_1) \exp\left(-\theta_1^2(1+\lambda^2)/2\right) d\theta_1, \\ & \int_{-\infty}^0 \left( 1 - \Phi \left[ -\frac{\sqrt{1+\lambda^2}}{\sqrt{1-r^2}} \theta_1 r - \frac{x_2}{\sqrt{1+\lambda^2}} \right] \right) \exp(x_1^i \theta_1) \exp\left(-\theta_1^2(1+\lambda^2)/2\right) d\theta_1. \end{aligned}$$

CRITICAL REGION FOR THE ONE-SIDED TEST: The test statistic  $z(x_1, x_2, \tau, \lambda^2)$  and the approximation to the *critical value function*  $\widehat{c}(x_2, \tau, \lambda^2)$  induce the following critical region:

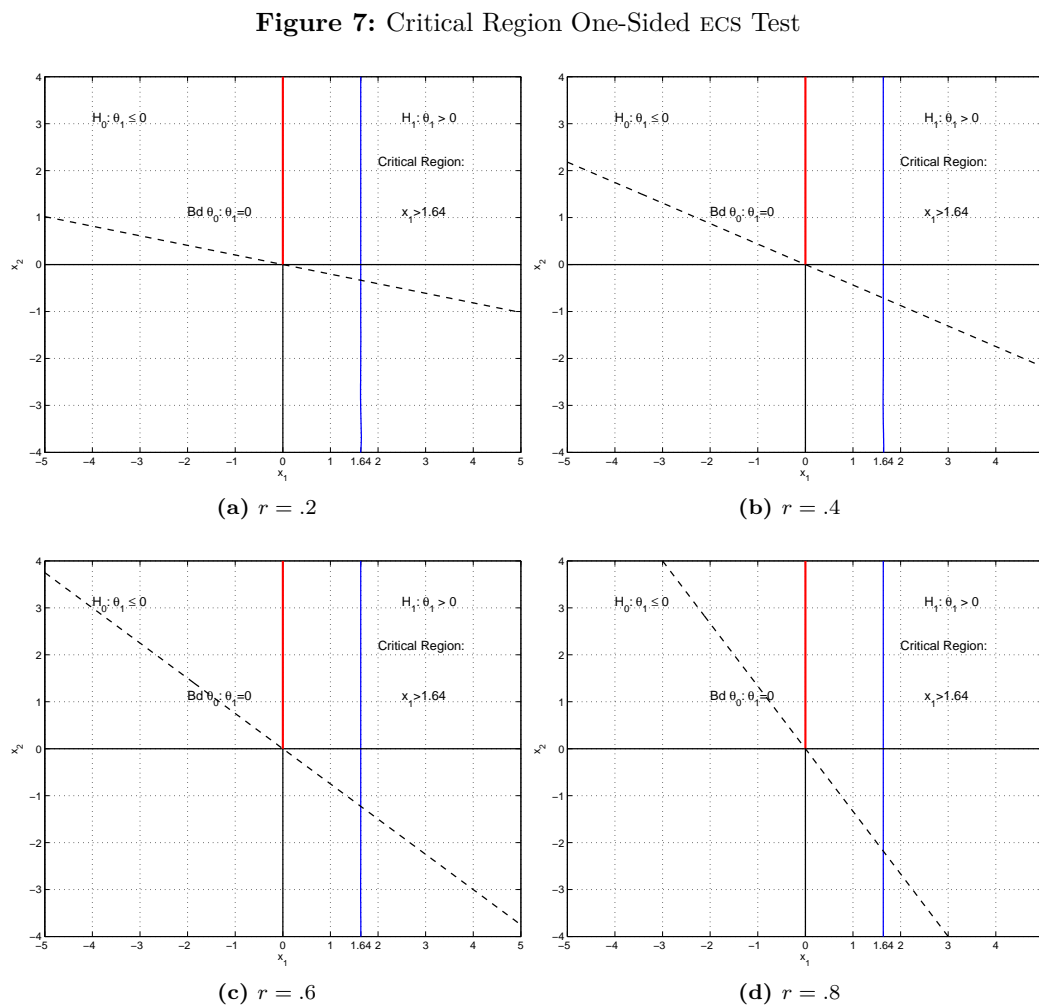
$$\{(x_1, x_2) \in \mathbb{R}^2 \mid z(x_1, x_2, \lambda^2, \tau) - c(x_2, \lambda^2, \tau) > 0\}.$$

The boundary of the critical region is defined as

$$\{(x_1, x_2) \in \mathbb{R}^2 \mid z(x_1, x_2, \lambda^2, \tau) - c(x_2, \lambda^2, \tau) = 0\},$$

where  $c(x_2, \lambda^2, \tau)$  denotes the population conditional quantiles of  $z(x_1, x_2, \lambda^2, \tau)$ , which we approximate with the Monte-Carlo conditional quantiles  $\widehat{c}(x_2, \lambda^2, \tau)$ .

Figure 7 presents an approximation to the boundary of the critical region for the one-sided test. In the figure,  $\lambda^2 = 1$  and  $\tau = 12\%$ .



(RED) Boundary of the null hypothesis. (BLUE) A 10th degree polynomial is fitted to the critical value functions using the command `polyfit` in Matlab. The domain is restricted to  $x_2 \in [-5, 5]$ . An anonymous function is defined to compute the test statistic for the one-sided test  $z(x_1, x_2; \lambda^2, \tau, r)$ . The command `ezplot` is used to graph the solution of the equation  $z(x_1, x_2; \lambda^2, \tau, r) - c_{os}(x_2; \lambda^2, \tau, r) = 0$ , which lies very close to  $x_1 = 1.64$ . (BLACK, DASHED) Boundary of the parameter space.

D.1. *Integrated Likelihoods for the Problem*

Define  $s \equiv -r/\sqrt{1-r^2}$ . First, it will be convenient to derive an expression for the integrated likelihood over the set

$$\left\{ (\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 \in [a, b] \text{ and } \theta_2 \in [\theta_1 s, \infty) \right\}$$

We are interested in three different values for the integrated likelihood:

$$f^*(x_1, x_2) \equiv f_{[-\infty, \infty]}^*(x_1, x_2), \quad f_1^*(x_1, x_2) \equiv f_{[0, \infty]}^*(x_1, x_2), \quad f_0^*(x_1, x_2) \equiv f_{[-\infty, 0]}^*(x_1, x_2)$$

The weight is denoted by  $p(\theta_1, \theta_2; \lambda^2)$  and it is assumed to be the product of two independent gaussian random variables (mean zero, variance  $\lambda^{-2}$ ) restricted to the subset of interest. We will use  $a_i$  to denote integration constants that do not depend on  $(x_1, x_2)$  or  $(\theta_1, \theta_2)$ . Note that:

$$\begin{aligned} f_{[a, b]}^*(x_1, x_2) &\equiv \int_a^b \left( \int_{\theta_1 s}^{\infty} f(x_1, x_2; \theta_1, \theta_2) p(\theta_1, \theta_2; \lambda^2) d\theta_2 \right) d\theta_1 \\ &= a_1 \int_a^b \left( \int_{\theta_1 s}^{\infty} \exp\left(-\frac{1}{2}(x_1 - \theta_1)^2\right) \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) p(\theta_1, \theta_2; \lambda^2) d\theta_2 \right) d\theta_1 \\ &\quad \text{(where we have used the definition of } f) \\ &= a_2 \int_a^b \left( \int_{\theta_1 s}^{\infty} \exp\left(-\frac{1}{2}(x_1 - \theta_1)^2\right) \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) \exp\left(-\frac{\lambda^2 \theta_1^2}{2}\right) \right. \\ &\quad \left. \exp\left(-\frac{\lambda^2 \theta_2^2}{2}\right) d\theta_2 \right) d\theta_1 \\ &\quad \text{(where we have used the definition of } p) \\ &= a_2 \exp\left(-\frac{x_1^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \int_a^b \left( \int_{\theta_1 s}^{\infty} \left[ \exp(x_1 \theta_1) \exp(x_2 \theta_2) \right. \right. \\ &\quad \left. \left. \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) \exp\left(-\frac{\theta_2^2}{2\sigma^2}\right) \right] d\theta_2 \right) d\theta_1, \quad \text{where } \sigma^2 \equiv 1/(1 + \lambda^2) \\ &\quad \text{(where we have expanded the squared terms)} \\ &= a_2 \exp\left(-\frac{x_1^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \exp\left(\frac{x_2^2 \sigma^2}{2}\right) \\ &\quad \int_a^b \left( \int_{\theta_1 s}^{\infty} \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}[\theta_2 - x_2 \sigma^2]^2\right) d\theta_2 \right) d\theta_1 \\ &\quad \left( \text{where we have used } -\frac{\theta_2^2}{2\sigma^2} + x_2 \theta_2 = -\frac{1}{2\sigma^2}[\theta_2^2 - 2\theta_2 x_2 \sigma^2 + x_2^2 \sigma^4 - x_2^2 \sigma^4] \right) \\ &= a_2 \exp\left(-\frac{x_1^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \exp\left(\frac{x_2^2 \sigma^2}{2}\right) \\ &\quad \int_a^b \left( \left[ \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) \right] \left[ \int_{\theta_1 s}^{\infty} \exp\left(-\frac{1}{2\sigma^2}[\theta_2 - x_2 \sigma^2]^2\right) d\theta_2 \right] \right) d\theta_1 \\ &= a_3 \exp\left(-\frac{x_1^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \exp\left(\frac{x_2^2 \sigma^2}{2}\right) \\ &\quad \int_a^b \left( 1 - \Phi\left[\frac{\theta_1 s}{\sigma} - x_2 \sigma\right] \right) \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 \end{aligned} \tag{D.4}$$

where  $\Phi$  is c.d.f. of a standard normal and the last equality is derived using the following observation:

$$\mathbb{P}\left(\theta_2 \geq \theta_1 s\right) = \int_{\theta_1 s}^{\infty} \exp\left(-\frac{1}{2\sigma^2}[\theta_2 - x_2 \sigma^2]^2\right) d\theta_2, \quad \text{if } \theta_2 \sim \mathcal{N}\left(x_2 \sigma^2, \sigma^2\right)$$

D.2. ECS test for  $\theta_1 = 0$ 

The results in Section 2 imply that the ECS test for  $\theta_1 = 0$  rejects if

$$\int_{-\infty}^{\infty} \left(1 - \Phi \left[ \frac{\theta_1 s}{\sigma} - x_2 \sigma \right]\right) \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 > c_2(x_2; \sigma^2, r), \quad s \equiv (-r/\sqrt{1-r^2}), \sigma = (1+\lambda^2)^{-1/2}$$

We will use equation (D.4) and change of variables formula to show that the ECS test for the two-sided problem rejects if and only if:

$$\exp\left(\frac{\sigma^2}{2} x_1^2\right) \left[1 - \Phi\left(-\sigma[r x_1 + \sqrt{1-r^2} x_2]\right)\right] > c(x_2; \sigma^2, r)$$

Consider the change of variables given by the linear transformation  $T: \Theta \rightarrow \mathbb{R}^2$ :

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \equiv T(\theta) \equiv \begin{pmatrix} T_1(\theta_1, \theta_2) \\ T_2(\theta_1, \theta_2) \end{pmatrix} = \begin{pmatrix} \sqrt{1-r^2} & -r \\ r & \sqrt{1-r^2} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

Note that  $(\theta_1, \theta_2) \in \mathbb{R}^2$  and  $r\theta_1 + \sqrt{1-r^2}\theta_2 \geq 0$  imply  $T(\Theta) = \mathbb{R} \times \mathbb{R}_+$ . The Jacobian of the transformation is constant and equal to 1. Let  $T^{-1}(u)$  denote the inverse of  $T$  with coordinates  $T_1^{-1}(u_1, u_2) = \sqrt{1-r^2}u_1 + ru_2$ ,  $T_2^{-1}(u_1, u_2) = -ru_1 + \sqrt{1-r^2}u_2$ .

Consider the measurable function  $f: \mathbb{R} \times \mathbb{R}_+ \rightarrow [0, \infty]$ :

$$f(u_1, u_2) = \exp\left(-\frac{1}{2\sigma^2}[T_1^{-1}(u_1, u_2) - x_1\sigma^2]^2\right) \exp\left(-\frac{1}{2\sigma^2}[T_2^{-1}(u_1, u_2) - x_2\sigma^2]^2\right)$$

This function corresponds to the integrand in (D.4) after completing the square for the terms involving  $x_1$  and  $\theta_1$ :

$$\begin{aligned} f^*(x_1, x_2) &= a_2 \exp\left(-\frac{z_1^2}{2}\right) \exp\left(-\frac{z_2^2}{2}\right) \exp\left(\frac{z_1^2 \sigma^2}{2}\right) \exp\left(\frac{z_2^2 \sigma^2}{2}\right) \\ &\quad \int_{-\infty}^{\infty} \left( \int_{\theta_1 s}^{\infty} \exp\left(-\frac{1}{2\sigma^2}[\theta_1 - z_1\sigma^2]^2\right) \exp\left(-\frac{1}{2\sigma^2}[\theta_2 - z_2\sigma^2]^2\right) d\theta_2 \right) d\theta_1 \end{aligned}$$

Note that  $[T_1^{-1}(u_1, u_2) - x_1\sigma^2]^2 + [T_2^{-1}(u_1, u_2) - x_2\sigma^2]^2$  equals:

$$\begin{aligned} &= (\sqrt{1-r^2}u_1 + ru_2 - x_1\sigma^2)^2 + (-ru_1 + \sqrt{1-r^2}u_2 - x_2\sigma^2)^2 \\ &\quad \text{(where we have used the definition of } T_i^{-1}, i = 1, 2) \\ &= r^2 u_2^2 - 2ru_2 x_1 \sigma^2 + 2r\sqrt{1-r^2}u_1 u_2 + (\sqrt{1-r^2}u_1 - x_1\sigma^2)^2 \\ &+ (1-r^2)u_2^2 - 2\sqrt{1-r^2}u_2 x_2 \sigma^2 - 2r\sqrt{1-r^2}u_2 u_1 + (ru_1 + x_2\sigma^2)^2 \\ &\quad \text{(where we have expanded the squared binomial)} \\ &= u_2^2 - 2u_2\sigma^2(rx_1 + \sqrt{1-r^2}x_2) + (\sqrt{1-r^2}u_1 - x_1\sigma^2)^2 + (ru_1 + x_2\sigma^2)^2 \\ &\quad \text{(where we have cancelled terms)} \\ &= u_2^2 - 2u_2\sigma^2(rx_1 + \sqrt{1-r^2}x_2) + (1-r^2)u_1^2 - 2\sqrt{1-r^2}u_1 x_1 \sigma^2 + x_1^2 \sigma^4 \\ &+ r^2 u_1^2 + 2ru_1 x_2 \sigma^2 + x_2^2 \sigma^2 \\ &\quad \text{(where we have expanded the squared binomials)} \\ &= u_1^2 + u_2^2 - 2u_2\sigma^2(rx_1 + \sqrt{1-r^2}x_2) - 2u_1\sigma^2(\sqrt{1-r^2}x_1 - rx_2) \\ &+ x_1^2 \sigma^4 + x_2^2 \sigma^4 \end{aligned}$$

Let  $f_{\text{Bd}}(x_1|x_2) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x_1^2\right)$ . By the Change-of-variables theorem (Theorem 7.26 p. 153 Rudin (2006)) and Fubini's theorem:

$$f^*(x_1, x_2)/f_{\text{Bd}}(x_1|x_2)$$

$$\begin{aligned}
&= a_2 \exp\left(-\frac{1}{2}x_2^2\right) \left(\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[u_1^2 - 2u_1\sigma^2(\sqrt{1-r^2}x_1 - rx_2)\right]\right) du_1\right) \\
&\quad \left(\int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[u_2^2 - 2u_2\sigma^2(rx_1 + \sqrt{1-r^2}x_2)\right]\right) du_2\right) \\
&\quad \text{(where we have used the definition of } f^*(x_1, x_2) \text{ and changed variables)} \\
&= a_3 \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\sigma^2}{2}\left[\sqrt{1-r^2}x_1 - rx_2\right]^2\right) \\
&\quad \exp\left(\frac{\sigma^2}{2}\left[rx_1 + \sqrt{1-r^2}x_2\right]^2\right) \left(\int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[u_2 - \sigma^2(rx_1 + \sqrt{1-r^2}x_2)\right]^2\right) du_2\right) \\
&\quad \text{(where we have solved for the gaussian integral in the first line above)} \\
&= a_4 \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\sigma^2}{2}\left[\sqrt{1-r^2}x_1 - rx_2\right]^2\right) \exp\left(\frac{\sigma^2}{2}\left[rx_1 + \sqrt{1-r^2}x_2\right]^2\right) \\
&\quad \left[1 - \Phi\left(-\sigma[rx_1 + \sqrt{1-r^2}x_2]\right)\right] \\
&\quad \text{(where we have completed the square and used the definition of the normal cdf } \Phi) \\
&= a_4 \exp\left(-\frac{1}{2}x_2^2\right) \exp\left(\frac{\sigma^2}{2}(x_1^2 + x_2^2)\right) \left[1 - \Phi\left(-\sigma[rx_1 + \sqrt{1-r^2}x_2]\right)\right], \quad \sigma = (1 + \lambda^2)^{-1/2}
\end{aligned}$$

Note that the second to last equality follows from the fact that

$$\left(\int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[u_2 - \sigma^2(rx_1 + \sqrt{1-r^2}x_2)\right]^2\right) du_2\right)$$

is proportional to the probability of the set  $[0, \infty]$  computed under the probability measure  $\mathcal{N}_1(-\sigma^2(rx_1 + \sqrt{1-r^2}x_2), \sigma^2)$ .

### D.3. ECS-Test for $\theta_1 \leq 0$

Since  $g(x_1, x_2) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x_1^2\right)$ , the one-sided ECS-test in (??) and definition (??) is given by:

$$\begin{aligned}
\left(\tau f_1^*(x_1, x_2) - (1-\tau)f_0^*(x_1, x_2)\right)/g(x_1, x_2) &= a_4 \exp\left(-\frac{x_2^2}{2}\right) \exp\left(\frac{x_2^2\sigma^2}{2}\right) \\
&\quad \left[\tau \int_0^{\infty} \left(1 - \Phi\left[\frac{\theta_1 s}{\sigma} - x_2\sigma\right]\right) \exp\left(x_1\theta_1\right) \right. \\
&\quad \left. \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1\right] - \\
&\quad (1-\tau) \left[\int_{-\infty}^0 \left(1 - \Phi\left[\frac{\theta_1 s}{\sigma} - x_2\sigma\right]\right) \exp\left(x_1\theta_1\right) \right. \\
&\quad \left. \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1\right] \\
&> c_1(x_2; \sigma^2, \tau, r)
\end{aligned}$$

This is, the test rejects if:

$$\begin{aligned}
&\left[\tau \int_0^{\infty} \left(1 - \Phi\left[\frac{\theta_1 s}{\sigma} - x_2\sigma\right]\right) \exp(x_1\theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1\right] \\
&- (1-\tau) \int_{-\infty}^0 \left(1 - \Phi\left[\frac{\theta_1 s}{\sigma} - x_2\sigma\right]\right) \exp(x_1\theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 \\
&> c^*(x_2, \sigma^2, \tau, r)
\end{aligned}$$

The critical value function  $c^*(\cdot; \sigma^2, \tau, r)$  is parameterized by  $\sigma^2$ ,  $\tau$  and  $r$ . The parameter  $\sigma^2$  enters the prior through the definition:  $\sigma^2 \equiv (1 + \lambda^2)^{-1}$ . The parameter  $\tau$  enters the linear weighting function  $\mathcal{W}$  used

to trade-off average power and average size. The parameter  $r$  is a primitive of the problem.

**CALIBRATION OF  $\tau$ :** We now present a brief discussion to motivate the selection of the parameter  $\tau$  in the one sided testing problem that arises in the Linear Regression Model with a Sign Restriction. Note first that  $r = 0$  implies  $s = 0$ . Therefore, the  $\mathcal{W}$ -test for the normal prior rejects iff:

$$\left[ \tau \int_0^\infty \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 - (1 - \tau) \int_{-\infty}^0 \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 \right] > \tilde{c}^*(x_2, \sigma^2, \tau, r)$$

The distribution of the random variable on the left hand side of the expression above does not depend on  $x_2$ . Consequently,  $\tilde{c}^*$  is constant in  $x_2$ .<sup>32</sup> We now argue that when  $s = 0$  the  $\mathcal{W}$ -test coincides with an Integrated Likelihood Ratio test for an appropriate choice of  $\tau$ . Let  $\tau^*(\alpha, \sigma^2)$  be defined as:

$$\mathbb{P}\left(\psi_{\text{ILR}}(x_1; \sigma^2) > (1 - \tau^*(\alpha, \sigma^2))/\tau^*(\alpha, \sigma^2)\right) = \alpha, \quad x_1 \sim \mathcal{N}(0, 1)$$

where

$$\psi_{\text{ILR}}(x_1; \sigma^2) \equiv \left[ \int_0^\infty \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 / \int_{-\infty}^0 \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 \right]$$

The test  $\psi_{\text{ILR}}$  corresponds to the Integrated Likelihood Ratio or Bayes Test (Lehmann and Romano (2005), Theorem 6.7.2 pg. 236; Chamberlain (2007) pg. 643-645) that rejects whenever the integrated likelihood ratio  $f_1^*(x_1, x_2; \sigma^2)/f_0^*(x_1, x_2; \sigma^2)$  is bigger than a constant threshold. The value  $\tau^*(\alpha)$  guarantees that the rejection probability on the boundary of the null hypothesis (which corresponds to the highest rejection probability of the test  $\psi_{\text{ILR}}$  over  $\mathbf{H}_0 : \theta_1 \leq 0$ ) is equal to  $\alpha$ . Therefore,

$$\begin{aligned} \mathbb{P}\left(\left[ \tau^*(\alpha, \sigma^2) \int_0^\infty \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 - (1 - \tau^*(\alpha, \sigma^2)) \int_{-\infty}^0 \exp(x_1 \theta_1) \exp\left(-\frac{\theta_1^2}{2\sigma^2}\right) d\theta_1 \right] \right. \\ \left. > \tilde{c}^*(x_2, \sigma^2, \tau) \right) = \alpha, \quad x_1 \sim \mathcal{N}(0, 1) \end{aligned}$$

if and only if the function  $c^*(\cdot; \sigma^2, \tau) = 0$  for (lebesgue) almost every  $z_2 \in \mathbb{R}$  and for any given value of  $\sigma^2$ .

---

<sup>32</sup>Intuitively, when  $x_1$  and  $x_2$  are independent ( $r = 0$ ) conditioning on  $x_2$  does not any information, as the restrictions imposed on  $\theta_2$  are not important for inference on  $\theta_1$ .

## APPENDIX E: PROBIT WITH ENDOGENEITY

E.1. Restrictions on the set  $\Phi$ 

Let  $P_z$  be a Borel probability measure on the real line. The space of such measures is denoted  $\Delta_B(\mathbb{R})$ . Define

$$\begin{aligned} \Delta_B(\mathbb{R}, c_1) &\equiv \left\{ P_z \in \Delta_B(\mathbb{R}) \text{ s.t. the following restrictions hold} \right. \\ &\quad \mathbf{R1)} \mathbb{E}_{P_z}[z] = 0 \text{ and } \mathbb{E}_{P_z}[z^2] = 1, \\ &\quad \left. \mathbf{R2)} \mathbb{E}_{P_z}[|z|^{2+\epsilon}] < c_1 \text{ for some } \epsilon > 0 \right\} \end{aligned}$$

and consider the set:

$$\Phi \equiv \left\{ (\rho, \sigma_v, P_z) \mid P_z \in \Delta_B(\mathbb{R}, c_1), \sigma_v \in [a, b] \subseteq [0, c_2], \rho \in [-c_3, c_3], 0 < c_3 < 1 \right\}$$

A sequence  $\phi_n = (\rho_n, \sigma_{vn}, P_{zn})$  of elements in  $\Phi$  converges to  $(\rho, \sigma_v, P_z)$  if and only if:

$$\rho_n \rightarrow \rho, \quad \sigma_{vn} \rightarrow \sigma_v, \quad P_{zn} \xrightarrow{d} P_z.$$

E.2. Assumption Q1: Derivative of  $Q_n(\beta, \pi)$ 

Let  $L(x)$  denote the standard normal c.d.f. Let:

$$\dot{L}(x) \equiv \left. \frac{\partial L(x)}{\partial x} \right|_x.$$

Hence  $\dot{L}(x)$  is the standard normal p.d.f. evaluated at  $x$ .

$$\begin{aligned} D_\beta Q_n(0, \pi_0) &= \frac{1}{n} \sum_{i=1}^n \frac{y_{in} \dot{L}(0) \pi_0 z_i}{\dot{L}(0)} - \frac{(1 - y_{in}) \dot{L}(0) \pi_0 z_{in}}{1 - \dot{L}(0)} \\ &= \frac{1}{n} \sum_{i=1}^n \pi_0 z_{in} \left[ \frac{y_{in} \dot{L}(0)}{\dot{L}(0)} - \frac{(1 - y_{in}) \dot{L}(0)}{1 - \dot{L}(0)} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \pi_0 z_i \left[ \frac{\dot{L}(0)(y_{in} - L(0))}{L(0)(1 - L(0))} \right] \\ &= \frac{\pi_0 \dot{L}(0)}{L(0)(1 - L(0))} \frac{1}{n} \sum_{i=1}^n z_{in} [y_{in} - L(0)] \end{aligned}$$

## E.3. Assumption Q2: Identification Statistic

Let  $\{\gamma_n = (\beta_n, \pi_n, \phi_n)\}_{n \in \mathbb{N}}$  be an arbitrary sequence in  $\Gamma^*$  with limit  $(\beta, \pi_0, \phi)$ . Consider the following identification statistic:

$$\widehat{\beta}(\pi_0) \equiv \left( \frac{1}{n} \sum_{i=1}^n z_{in}^2 \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_{in} x_{in} = \beta_n + \left( \frac{1}{n} \sum_{i=1}^n z_{in}^2 \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_{in} v_{in}$$

Note that  $\{z_{in}^2\}_{i=1}^n, n \in \mathbb{N}$  is a triangular array of row-wise i.i.d. random variables. Since  $\gamma_n \in \Gamma^*$ , it follows that  $P_{zn} \xrightarrow{d} P_z$  for some  $P_z$  in  $\Delta_B(\mathbb{R}, c_1)$ . In addition, for arbitrary  $i$ , restriction **R2)** in the set  $\Delta_B(\mathbb{R}, c_1)$  implies

$$\sup_{n \geq i} \mathbb{E}_{\gamma_n}[|z_{in}|^{2+\epsilon}] < c_1.$$

Therefore, from Corollary 1

$$g_2(W^n) = \frac{1}{n} \sum_{i=1}^n z_{in}^2 \rightarrow \mathbb{E}_\gamma[z^2],$$



where the last equality follows from the fact that  $\gamma \in \Gamma$  implies  $\mathbb{E}_{P_z}[z^2] = 1$  by restriction **R1**). Consider now the row-wise i.i.d. triangular array given by  $\{z_{in}v_{in}\}_{i=1}^n, n \in \mathbb{N}$ . Note that for arbitrary  $i$ ,

$$\{z_{in}v_{in}\}_{n=i}^{\infty} \xrightarrow{d} zv$$

where  $z \sim P_z, v \sim \mathcal{N}(0, \sigma_v^2)$ , and  $(z, v)$  are independent random variables. Furthermore, note that  $\gamma_n \in \Gamma$  implies

$$\sup_{n \geq i} \mathbb{E}_{\gamma_n} [(z_{in}v_{in})^{2+\epsilon}] = \sup_{n \geq i} \mathbb{E}_{\gamma_n} [|z_{in}|^{2+\epsilon}] \mathbb{E}_{\gamma_n} [|v_{in}|^{2+\epsilon}] \leq c_1 c^*(c_2),$$

where the last inequality follows by the definition of  $\Gamma$ . Corollary 2 implies:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in}v_{in} \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_{\phi}[z^2]\sigma_v^2).$$

Hence, for an arbitrary sequence  $\gamma_n \in \Gamma^*$  with limit  $(\beta, \pi_0, \phi)$ :

$$\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n) \xrightarrow{d} \mathcal{N}_1(0, \mathbb{E}_{\phi}[z^2]^{-1}\sigma_v^2)$$

#### E.4. Assumption Q3: $m^*(W_{in}, \beta, \pi)$ and $f^*(W_{in}, \beta, \pi)$

In the probit model the data  $W_{in} = (y_{in}, x_{in}, z_{in})' \in \mathbb{R}^3$ . Consider the real valued functions:

$$m^*(W_{in}, \beta, \pi) = z_{in}[L(\beta\pi z_{in}) - L(0)] \quad \text{and} \quad f^*(W_{in}, \beta, \pi) = \beta z_{in}^2$$

Note that:

$$\left( \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n m(W_{in}, \pi_0) - m^*(W_{in}, \beta_n, \pi_n) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n f(W_{in}, \pi_0) - f^*(W_{in}, \beta, \pi_n) \end{array} \right) = \left( \begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} [y_{in} - L(\beta_n \pi_n z_{in})] \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} v_{in} \end{array} \right).$$

It is now shown that for arbitrary  $a, b \in \mathbb{R}$ , the linear combination

$$(E.1) \quad a \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} [y_{in} - L(\beta_n \pi_n z_{in})] + b \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} v_{in}$$

follows a Central Limit Theorem under arbitrary sequences of true parameter values  $\gamma_n \in \Gamma^*$ .<sup>33</sup> The joint normality in Assumption Q3 will follow by the Cramer-Wold theorem in Durrett (2010), pg. 176. To establish the asymptotic normality of E.1, the assumptions of Corollary 2 are now verified. Note first that:

$$\mathbb{E}_{\gamma_n} [z_{in}[y_{in} - L(\beta_n \pi_n z_{in})]] = 0.$$

To verify this equality, consider the following conditional distribution:

$$(z_{in}[y_{in} - L(\beta_n \pi_n z_{in})]) \mid z_{in} = \begin{cases} 1 - L(\beta_n \pi_n z_{in}) & \text{with probability } L(\beta_n \pi_n z_{in}) \\ -L(\beta_n \pi_n z_{in}) & \text{with probability } 1 - L(\beta_n \pi_n z_{in}) \end{cases}$$

Therefore,  $\mathbb{E}_{\gamma_n} [z_{in}[y_{in} - L(\beta_n \pi_n z_{in})] \mid z_{in}] = 0$  for every  $z_{in}$ . Note also that,

$$\gamma_n = (\beta_n, \pi_n, \rho_n, \sigma_{vn}, P_{zn})$$

converges to

$$\gamma = (\beta, \pi, \rho, \sigma_v, P_z) \in \Gamma,$$

which implies—along with the independence of  $(u_{in}, v_{in})$  and  $z_{in}$ —that the sequence of random vectors

$$(u_{in}, v_{in}, z_{in}) \xrightarrow{d} (u, v, z)$$

<sup>33</sup>Assumption Q3 requires only joint normality under “weak” sequences; that is, sequences of parameter values  $\gamma_n \in \Gamma^*(\gamma, 0)$ . However, it is important to note that a stronger result is available: joint normality under any sequence in  $\Gamma^*$ . Since the proof is the same regardless of the sequences considered, the more general result is established in this Appendix.

where

$$(u, v) \sim \mathcal{N}_2 \left( \mathbf{0}, \begin{bmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{bmatrix} \right), \quad z \sim P_z$$

and  $(u, v)$  is independent of  $z$ . Therefore, for arbitrary  $i$ , the Continuous Mapping Theorem implies that the sequence of mean-zero random variables

$$\left\{ az_{in}[y_{in} - L(\beta_n \pi_n z_{in})] + bz_{in}v_{in} \right\}_{n=i}^{\infty} \xrightarrow{d} az[(\mathbf{1}(\beta\pi z + u > 0) - L(\beta\pi z))] + bzv.$$

Now, it is shown that

$$\sup_n \mathbb{E}_{\gamma_n} \left[ \left| az_{in}[y_{in} - L(\beta_n \pi_n z_{in})] + bz_{in}v_{in} \right|^{2+\epsilon} \right] < \infty$$

By Minkowski's inequality:

$$\begin{aligned} \mathbb{E}_{\gamma_n} \left[ \left| az_{in}[y_{in} - L(\beta_n \pi_n z_{in})] + bz_{in}v_{in} \right|^{2+\epsilon} \right]^{\frac{1}{2+\epsilon}} &\leq |a| \mathbb{E}_{\gamma_n} \left[ \left| z_{in}[y_{in} - L(\beta_n \pi_n z_{in})] \right|^{2+\epsilon} \right]^{\frac{1}{2+\epsilon}} \\ &\quad + |b| \mathbb{E}_{\gamma_n} \left[ \left| z_{in}v_{in} \right|^{2+\epsilon} \right]^{\frac{1}{2+\epsilon}} \\ &\leq |a| \mathbb{E}_{\gamma_n} \left[ \left| z_{in} \right|^{2+\epsilon} \right]^{\frac{1}{2+\epsilon}} (\text{as } |y_{in} - L(\beta_n \pi_n z_{in})| \leq 1) \\ &\quad + |b| \mathbb{E}_{\gamma_n} \left[ \left| z_{in} \right|^{2+\epsilon} \right]^{\frac{1}{2+\epsilon}} \mathbb{E}_{\gamma_n} \left[ \left| v_{in} \right|^{2+\epsilon} \right]^{\frac{1}{2+\epsilon}} \\ &\quad (\text{since } z_{in} \perp v_{in}) \\ &\leq |a|c_1^{\frac{1}{2+\epsilon}} + |b|c_1^{\frac{1}{2+\epsilon}}c^*(c_2), \text{ for some } c^*(c_2) \in \mathbb{R} \\ &\quad (\text{since } \gamma_n \text{ satisfies } \mathbf{R2} \text{ and } \sigma_v \in [0, c^2]) \end{aligned}$$

The assumptions of Corollary 2 are thus verified. Therefore, under any sequence  $\{\gamma_n\} \in \Gamma^*$  with limit  $\gamma$ :

$$a \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} \left[ y_{in} - L(\beta_n \pi_n z_{in}) \right] + b \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} v_{in} \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_{\gamma}[(az[(\mathbf{1}(\beta\pi z + u > 0) - L(\beta\pi z))] + bzv)^2]),$$

for any  $a, b \in \mathbb{R}$ . The asymptotic variance can be written as  $(a, b)W^*(\pi_0, \gamma)(a, b)'$  with

$$W^*(\pi_0, \gamma) \equiv \mathbb{E}_{\gamma} \left[ \begin{pmatrix} z[\mathbf{1}(\beta\pi z + u > 0) - L(\beta\pi z)] \\ zv \end{pmatrix} \begin{pmatrix} z[\mathbf{1}(\beta\pi z + u > 0) - L(\beta\pi z)] \\ zv \end{pmatrix}' \right].$$

Therefore, assumption Q3 holds with

$$W(\pi_0, \gamma) \equiv \mathbb{E}_{\gamma} \left[ \begin{pmatrix} z[\mathbf{1}(u > 0) - L(0)] \\ zv \end{pmatrix} \begin{pmatrix} z[\mathbf{1}(u > 0) - L(0)] \\ zv \end{pmatrix}' \right].$$

#### E.5. Assumption Q4.1: First estimation error for probit

$$\begin{aligned} \sqrt{n}\Delta_m^{\pi^*}(\pi_n, \pi) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m^*(W_{in}, \beta_n, \pi_n) - m^*(W_{in}, \beta_n, \pi_0) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} [L(\beta_n \pi_n z_{in}) - L(\beta_n \pi_0 z_{in})] \\ &= \frac{1}{n} \sum_{i=1}^n z_{in} \sqrt{n} [L(\beta_n \pi_n z_{in}) - L(\beta_n \pi_0 z_{in})] \end{aligned}$$

The probably limit of the expression above is obtained by combining Claim1 and Claim 2. The idea is to use Corollary 1 to use a Weak Law of Large Numbers for row-wise i.i.d. triangular arrays.

**CLAIM 1:** For any weak sequence  $(\beta_n, \pi_n, \rho_n, \sigma_{vn}, P_{zn}) \rightarrow (0, \pi, \rho, \sigma_v, P_z)$  and  $\sqrt{n}\beta_n \rightarrow b \in \mathbb{R}$

$$z_{in}\sqrt{n}\left(L(\beta_n\pi_n z_{in}) - L(\beta_n\pi_0 z_{in})\right) \xrightarrow{d} bz^2(\pi - \pi_0)\dot{L}(0)$$

REMARK 17: A delta method argument is sufficient to establish the claim. However, the proof takes a different approach to derive a convenient equality that will be exploited afterwards.

PROOF: Note first that for any  $x \in \mathbb{R}$

$$L(x) = \frac{1}{2} + \int_0^x \dot{L}(0) \exp\left(-\frac{1}{2}x^2\right) dx$$

Therefore,

$$\begin{aligned} L(\beta_n\pi_n z_{in}) - L(\beta_n\pi_0 z_{in}) &= \int_0^{\beta_n\pi_n z_{in}} \dot{L}(0) \exp\left(-\frac{1}{2}x^2\right) dx - \int_0^{\beta_n\pi_0 z_{in}} \dot{L}(0) \exp\left(-\frac{1}{2}x^2\right) dx \\ &= \beta_n\pi_n z_{in} \int_0^1 \dot{L}(0) \exp\left(-\frac{1}{2}u^2(\beta_n\pi_n z_{in})^2\right) du \\ &\quad - \beta_n\pi_0 z_{in} \int_0^1 \dot{L}(0) \exp\left(-\frac{1}{2}u^2(\beta_n\pi_0 z_{in})^2\right) du \\ &\quad \text{(where I have used the change of variable } u = x/\beta_n\pi_n z_{in}\text{)} \\ \text{(E.2)} \quad &= \beta_n\pi_n z_{in} G(\beta_n\pi_n z_{in}) - \beta_n\pi_0 z_{in} G(\beta_n\pi_0 z_{in}), \end{aligned}$$

where the function  $G: \mathbb{R} \rightarrow \mathbb{R}$  is defined by:

$$G(a) \equiv \int_0^1 \dot{L}(0) \exp\left(-\frac{1}{2}u^2 a^2\right) du$$

Note now that  $G(a)$  is sequentially continuous at every  $a \in \mathbb{R}$ . To see this, consider any sequence  $a_n \rightarrow a$ . Then, the sequence of functions

$$f_n(u) = \dot{L}(0) \exp\left(-\frac{1}{2}u^2 a_n^2\right) \rightarrow \dot{L}(0) \exp\left(-\frac{1}{2}u^2 a^2\right)$$

for all  $u$ . Since  $\left|\dot{L}(0) \exp\left(-\frac{1}{2}u^2 a_n^2\right)\right| \leq \dot{L}(0)$ , for all  $n \in \mathbb{N}, u \in [0, 1]$ , then  $G(a_n) \rightarrow G(a)$ ; which follows from an application of Lebesgue's dominated convergence theorem (Theorem 16.4 in Billingsley (1995), pg. 209). Consequently, under any weak sequence  $(\beta_n, \pi_n, \rho_n, \sigma_{vn}, P_{zn}) \rightarrow (0, \pi, \rho, \sigma_v, P_z)$

$$z_{in}\sqrt{n}\left(L(\beta_n\pi_n z_{in}) - L(\beta_n\pi_0 z_{in})\right) \xrightarrow{d} bz^2 G(0) - b\pi_0 z G(0) = bz^2(\pi - \pi_0)\dot{L}(0),$$

where  $z \sim P_z$  and  $\sqrt{n}\beta_n \rightarrow b$ .

*Q.E.D.*

Now, a Weak Law of Large Numbers will be applied to the sum:

$$\sqrt{n}\Delta_{m^*}^\pi(\pi_n, \pi) = \frac{1}{n} \sum_{i=1}^n z_{in}\sqrt{n}[L(\beta_n\pi_n z_{in}) - L(\beta_n\pi_0 z_{in})]$$

Corollary 1 gives the desired result under the following condition:

**CLAIM 2:**

$$\sup_{n \geq i} \mathbb{E}_{\gamma_n} \left[ \left| z_{in}\sqrt{n}\left(L(\beta_n\pi_n z_{in}) - L(\beta_n\pi_0 z_{in})\right) \right|^{1+\epsilon/2} \right] < \infty$$

PROOF: Equation E.2 implies

$$\begin{aligned}
\mathbb{E}_{\gamma_n} \left[ \left| z_{in} \sqrt{n} \left( L(\beta_n \pi_n z_{in}) - L(\beta_n \pi_0 z_{in}) \right) \right|^{1+\epsilon/2} \right] &= \sqrt{n} \beta_n \mathbb{E}_{\gamma_n} \left[ \left| z_{in}^2 (\pi_n G(\beta_n \pi_n z_{in}) - \pi_0 G(\beta_n \pi_0 z_{in})) \right|^{1+\epsilon/2} \right] \\
&= \sqrt{n} \beta_n \mathbb{E}_{\gamma_n} \left[ \left| z_{in} \right|^{2+\epsilon} \left| \pi_n G(\beta_n \pi_n z_{in}) - \pi_0 G(\beta_n \pi_0 z_{in}) \right|^{1+\epsilon/2} \right] \\
&\leq \sqrt{n} \beta_n \left| (\pi_n + \pi_0) \dot{L}(0) \right|^{1+\epsilon/2} \mathbb{E}_{\gamma_n} \left[ \left| z_{in} \right|^{2+\epsilon} \right] \\
&\leq \sqrt{n} \beta_n \left| (\pi_n + \pi_0) \dot{L}(0) \right|^{1+\epsilon/2} c_1
\end{aligned}$$

Since the sequence  $\sqrt{n} \beta_n \left| (\pi_n + \pi_0) \dot{L}(0) \right|^{1+\epsilon/2}$  converges, it is bounded. Let  $M$  denote such bound. Therefore:

$$\sup_{n \geq i} \mathbb{E}_{\gamma_n} \left[ \left| z_{in} \sqrt{n} \left( L(\beta_n \pi_n z_{in}) - L(\beta_n \pi_0 z_{in}) \right) \right|^{1+\epsilon/2} \right] \leq M c_1$$

*Q.E.D.*

The last two claims and Corollary 1 imply:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n z_{in} \sqrt{n} [L(\beta_n \pi_n z_{in}) - L(\beta_n \pi_0 z_{in})] &\xrightarrow{d} \mathbb{E}_{\gamma} [b z^2 (\pi - \pi_0) \dot{L}(0)] \\
&= b \mathbb{E}_{\gamma} [z^2 \dot{L}(0) \pi - z^2 \dot{L}(0) \pi_0] \\
&= b \mathbb{E}_{\phi} [z^2] \dot{L}(0) (\pi - \pi_0)
\end{aligned}$$

where the last term equals

$$b \mathbb{E}_{\gamma} [m_{\beta}^*(w_{in}, 0, \pi) - m_{\beta}^*(w_{in}, 0, \pi_0)]$$

as  $m^*(w_{in}, \beta, \pi) = z_{in} (L(\beta \pi z_{in}) - L(0))$ . Therefore, Assumption 3.4.2 holds for the probit model with endogeneity.

#### E.6. Assumption Q4.2: Second estimation error for probit

Just as before, the proof involves a change of variables. However, additional work is required due to the fact that  $\hat{\beta}(\pi_0)$  appears in the sum of interest and a Weak Law of Large Numbers is not directly applicable. Note that

$$\begin{aligned}
\sqrt{n}\Delta_{m^*}^\beta(\beta_n, \widehat{\beta}(\pi_0)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m^*(W_{in}, \beta_n, \pi_0) - m^*(W_{in}, \widehat{\beta}(\pi_0), \pi_0) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{in} [L(\beta_n \pi_0 z_{in}) - L(\widehat{\beta}(\pi_0) \pi_0 z_{in})] \\
&= \frac{1}{n} \sum_{i=1}^n z_{in} \sqrt{n} [L(\beta_n \pi_0 z_{in}) - L(\widehat{\beta}(\pi_0) \pi_0 z_{in})] \\
&= -\frac{1}{n} \sum_{i=1}^n z_{in} \sqrt{n} [L(\widehat{\beta}(\pi_0) \pi_0 z_{in}) - L(\beta_n \pi_0 z_{in})] \\
&= \left( -\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n) \right) \\
&\quad \left( \frac{1}{n} \sum_{i=1}^n \pi_0 z_{in}^2 \int_0^1 \dot{L}(0) \exp\left(-\frac{1}{2}[u(\widehat{\beta}(\pi_0) - \beta) \pi_0 z_{in} + \beta_n \pi_0 z_{in}]^2 du\right) \right)
\end{aligned}
\tag{E.3}$$

(after the change of variable  $u = (x - \beta_n \pi_0 z_{in}) / ((\widehat{\beta}(\pi_0) - \beta_n) \pi_0 z_{in})$ )

Define:

$$\widetilde{G}(a_1, a_2) \equiv \int_0^1 \dot{L}(0) \exp\left(-\frac{1}{2}[ua_1 + a_2]^2\right) du,$$

and note that  $\widetilde{G}(\cdot)$  is sequentially continuous in  $\mathbb{R}^2$ . Consequently, under any weak sequence

$$\widetilde{G}((\widehat{\beta}(\pi_0) - \beta_n) \pi_0 z_{in}, \beta_n \pi_0 z_{in}) \xrightarrow{d} \widetilde{G}(0, 0) = \dot{L}(0).$$

Claim 3 and Claim 4 are used to establish the asymptotic behavior of  $\sqrt{n}\Delta_{m^*}^\beta(\beta_n, \widehat{\beta}(\pi_0))$ . The intuition behind the two claims is as follows. A WLLN is not applicable due to the presence of  $(\widehat{\beta}(\pi) - \beta)$ . Therefore, this random variable is replaced by its probability limit: zero (Claim 3). Finally, one needs to show that difference between the sum in Claim 3 and the original object of interest is  $o_p(1)$  under any sequence (Claim 4).

**CLAIM 3:** *Under any weak sequence:*

$$\frac{1}{n} \sum_{i=1}^n \pi_0 z_{in}^2 \widetilde{G}(0, \beta_n \pi_n z_{in}) \xrightarrow{P} \mathbb{E}_\gamma[\pi_0 z^2 \dot{L}(0)] = \mathbb{E}_\gamma[m_\beta^*(w, 0, \pi_0)]$$

PROOF: Note that  $\{\pi_0 z_{in}^2 \widetilde{G}(0, \beta_n \pi_n z_{in})\}_{i=1}^n, n \in \mathbb{N}$  is a triangular array of row-wise i.i.d. random variables. Since  $\gamma_n \in \Gamma^*$  and  $\widetilde{G}(a_1, a_2)$  is bounded by  $\dot{L}(0)$ ,

$$\sup_{n \geq i} \mathbb{E}_{\gamma_n} \left[ |\pi_0 z_{in}^2 \widetilde{G}(0, \beta_n \pi_n z_{in})|^{1+\epsilon/2} \right] < \infty$$

Furthermore,

$$\pi_0 z_{in}^2 \widetilde{G}(0, \beta_n \pi_n z_{in}) \xrightarrow{d} \pi_0 z^2 \dot{L}(0)$$

The result follows from Corollary 1 and the fact that

$$m^*(w, \beta, \pi) = z_{in} [L(\beta \pi z_{in}) - L(0)].$$

*Q.E.D.*

Now, show that the difference between the sum in Claim 3 and the object of interest is  $o_p(1)$  under any weak sequence.

**CLAIM 4:**

$$\left| \frac{1}{n} \sum_{i=1}^n \pi_0 z_{in}^2 \tilde{G}(\widehat{\beta}(\pi_0) - \beta_n) \pi_0 z_{in}, \beta_n \pi_n z_{in} \right) - \frac{1}{n} \sum_{i=1}^n \pi_0 z_{in}^2 \tilde{G}(0, \beta_n \pi_n z_{in}) \right| \xrightarrow{p} 0$$

**PROOF:** Let  $\widehat{\beta} = \widehat{\beta}(\pi_0)$ . It is sufficient to show that

$$\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}((\widehat{\beta} - \beta_n) \pi_0 z_{in}, \beta_n \pi_n z_{in}) - \tilde{G}(0, \beta_n \pi_n z_{in}) \right| \right| \xrightarrow{p} 0$$

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be the probability space over which the array is defined. For  $\epsilon > 0$  and  $0 < q < 1/4$ , let  $B_0(\epsilon/n^q)$  denote the closed interval of radius  $\epsilon/n^q$  around 0. Abusing notation, consider the event:

$$A_{\epsilon, q}^{n, i} = \left\{ (\widehat{\beta} - \beta_n) \pi_0 z_{in} \in B_0(\epsilon/n^q) \right\}$$

where  $0 < q \leq 1/4$ . Note that

$$\mathbb{P}((A_{\epsilon, q}^{n, i})^c) = \mathbb{P}\left(\left| (\widehat{\beta} - \beta) \pi_0 z_{in} \right| > \frac{\epsilon}{n^q}\right)$$

**RESULT 6:** Suppose that  $\sup_{n \geq i} \mathbb{E}_{\gamma_n}[z_{in}^8] < \infty$ . Then, for each  $i$ :

$$\mathbb{P}\left((A_{\epsilon, q}^{n, i})^c\right) = n^{4q} O\left(\frac{1}{n^2}\right) + \exp\left(-O(n^{1-2q})\right)$$

**PROOF:** The proof is based on the tail bound in Chesneau (2009). For simplicity, suppose that  $\mathbb{E}_{\gamma_n}[z_{in}^2] = 1$ . Consider  $\widehat{\beta} = (n)^{-1} \sum_{i=1}^n z_{in} v_{in}$ . Note that:

$$\mathbb{P}\left(\left| (\widehat{\beta} - \beta) \pi_0 z_{in} \right| > \frac{\epsilon}{n^q}\right) = \mathbb{P}\left(\left| \frac{1}{n} \sum_{j \neq i} z_{jn} v_{jn} \pi_0 z_{in} + \frac{1}{n} z_{in}^2 \pi_0 v_{in} \right| > \frac{\epsilon}{n^q}\right)$$

Let  $Y_{jn} = \frac{1}{n} z_{jn} v_{jn} \pi_0 z_{in}$  and  $Y_{in} = \frac{1}{n} z_{in}^2 \pi_0 v_{in}$ . Note that  $\mathbb{E}_{\gamma_n}[Y_{jn}] = 0$  for all  $j = 1, \dots, n$ . Note also that by the assumption on the moments of  $z_{in}$ , it follows that  $\mathbb{E}_{\gamma_n}[|Y_{in}|^p] < \infty$ , for  $p = 4$ . Set  $Y_{jn}^* = -Y_{jn}$ . Let

$$b_n = \sum_{i=1}^n \mathbb{E}_{\gamma_n}[Y_{in}^2] = \left(\frac{n-1}{n^2}\right) \sigma_{vn}^2 \pi_0^2 + \frac{1}{n^2} \mathbb{E}_{\gamma_n}[z_{in}^4] \sigma_{vn}^2 \pi_0^2$$

Furthermore, let

$$r_{n,4}^* = \sum_{i=1}^n \mathbb{E}_{\gamma_n}[Y_{in}^4] = \frac{n-1}{n^4} \mathbb{E}_{\gamma_n}[z_{in}^4]^2 3 \sigma_{vn}^2 \pi_0^4 + \frac{1}{n^4} \mathbb{E}_{\gamma_n}[z_{in}^8] 3 \sigma_{vn}^2$$

Theorem 2.1 in pg. 2 of Chesneau (2009) implies:

$$\begin{aligned} \mathbb{P}\left(\left| \frac{1}{n} \sum_{j \neq i} z_{jn} v_{jn} \pi_0 z_{in} + \frac{1}{n} z_{in}^2 \pi_0 v_{in} \right| > \frac{\epsilon}{n^q}\right) &= \mathbb{P}\left(\sum_{i=1}^n Y_{in} > \frac{\epsilon}{n^q}\right) + \mathbb{P}\left(\sum_{i=1}^n Y_{in}^* > \frac{\epsilon}{n^q}\right) \\ &\leq 2C \left[ \frac{n^{4q}}{\epsilon^4} \max\left(r_{n,4}^*, b_n^2\right) \right] + 2 \exp\left(-\epsilon^2 O(n^{1-2q})\right) \\ &\leq n^{4q} O\left(\frac{1}{n^2}\right) + \exp\left(-O(n^{1-2q})\right) \end{aligned}$$

*Q.E.D.*

The last result is now used to bound the probability of the event:

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}((\hat{\beta} - \beta_n)\pi_0 z_{in}, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right| > \epsilon \right)$$

which equals

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}((\hat{\beta} - \beta_n)\pi_0 z_{in}, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right| > \epsilon \text{ and } \cap_{i=1}^n A_{\epsilon, q}^{n, i} \right) \\ & + \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}((\hat{\beta} - \beta_n)\pi_0 z_{in}, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right| > \epsilon \text{ and } (\cap_{i=1}^n A_{\epsilon, q}^{n, i})^c \right) \end{aligned}$$

Note that by definition of  $A_{\epsilon, q}^{n, i}$ :

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}((\hat{\beta} - \beta_n)\pi_0 z_{in}, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right| > \epsilon \text{ and } \cap_{i=1}^n A_{\epsilon, q}^{n, i} \right) \\ & \leq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \right|_{\sup_{\delta \in B_0(\epsilon/n^q)} \left| \tilde{G}(\delta, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right| > \epsilon} \right) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}((\hat{\beta} - \beta_n)\pi_0 z_{in}, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right| > \epsilon \text{ and } (\cap_{i=1}^n A_{\epsilon, q}^{n, i})^c \right) \\ & \leq \mathbb{P}\left( (\cap_{i=1}^n A_{\epsilon, q}^{n, i})^c \right) = \mathbb{P}\left( \cup_{i=1}^n (A_{\epsilon, q}^{n, i})^c \right) \leq n \mathbb{P}\left( (A_{\epsilon, q}^{n, i})^c \right) = n^{1+4q} O\left(\frac{1}{n^2}\right) + n \exp\left(-O(n^{1-2q})\right) \end{aligned}$$

Since  $q < 1/4$ ,

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}((\hat{\beta} - \beta_n)\pi_0 z_{in}, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right| > \epsilon \text{ and } (\cap_{i=1}^n A_{\epsilon, q}^{n, i})^c \right)$$

goes to zero as  $n \rightarrow \infty$ . Therefore, it is sufficient to establish the following result:

**RESULT 7:**

$$\frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \right|_{\sup_{\delta \in B_0(\epsilon/n^q)} \left| \tilde{G}(\delta, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \right|} \xrightarrow{p} 0 \text{ as } n \rightarrow \infty$$

PROOF: Note that

$$\tilde{G}(0, \beta_n \pi_0 z_{in}) = \dot{L}(0) \exp\left(-\frac{1}{2}(\beta_n z_{in} \pi_0)^2\right) = \dot{L}(\beta_n z_{in} \pi_0)$$

and using the change of variables  $x = u\delta + \beta_n \pi_0 z_{in}$

$$\tilde{G}(\delta, \beta_n \pi_0 z_{in}) = \frac{1}{\delta} \left[ L(\beta_n \pi_0 z_{in} + \delta) - L(\beta_n \pi_0 z_{in}) \right]$$

Note that as  $\delta \rightarrow 0$ ,  $\tilde{G}(\delta, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \rightarrow 0$  almost surely, by definition of derivative. Therefore, note that

$$\lim_{n \rightarrow \infty} \lim_{\delta \rightarrow 0} \tilde{G}(\delta, \beta_n \pi_0 z_{in}) - \tilde{G}(0, \beta_n \pi_0 z_{in}) \rightarrow 0 \text{ almost surely}$$

Now, simply note that since  $\tilde{G}(\cdot, \beta_n \pi_0 z_{in})$  is bounded. Hence the supremum is finite. In fact:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \sup_{\delta \in B_0(\epsilon/n^q)} \left| \tilde{G}(\delta, \beta_n \pi_n z_{in}) - \tilde{G}(0, \beta_n \pi_n z_{in}) \right| \right. \right| \\ &= \frac{1}{n} \sum_{i=1}^n \left| \pi_0 z_{in}^2 \left| \tilde{G}(\delta^*(\beta_n \pi_0 z_{in}), \beta_n \pi_n z_{in}) - \tilde{G}(0, \beta_n \pi_n z_{in}) \right| \right| \end{aligned}$$

Since  $\delta^*(\beta_n \pi_0 z_{in}) \in B_0(\epsilon/n^q)$ , then  $\delta^*$  converges a.s. to 0. Therefore, for each  $i$ :

$$\left| \pi_0 z_{in}^2 \left| \tilde{G}(\delta^*(\beta_n \pi_0 z_{in}), \beta_n \pi_n z_{in}) - \tilde{G}(0, \beta_n \pi_n z_{in}) \right| \right| \xrightarrow{d} 0$$

The result then follows by Corollary 1.

*Q.E.D.*

Result 6 and Result 7 imply:

$$\left| \frac{1}{n} \sum_{i=1}^n \pi_0 z_{in}^2 \tilde{G}((\hat{\beta}(\pi_0) - \beta_n) \pi_0 z_{in}, \beta_n \pi_n z_{in}) - \frac{1}{n} \sum_{i=1}^n \pi_0 z_{in}^2 \tilde{G}(0, \beta_n \pi_n z_{in}) \right| \xrightarrow{p} 0,$$

which completes the proof of Claim 4.

*Q.E.D.*

Consequently, combining the two claims the desired high-level assumption is verified:

$$\sqrt{n} \Delta_{m^*}^{\beta}(\beta_n, \hat{\beta}(\pi_0)) = - \left( \sqrt{n} (\hat{\beta}(\pi_0) - \beta_n) \right) \mathbb{E}_{\phi}[\pi_0 z^2 \dot{L}(0)]$$

#### E.7. Assumption A4.3

Since:

$$f^*(W_{in}, \beta, \pi) = \beta z_{in}^2,$$

then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n f^*(W_{in}, \beta_n, \pi_n) = \sqrt{n} \beta_n \frac{1}{n} \sum_{i=1}^n z_{in}^2$$

Since  $\gamma^* = (\beta_n, \pi_n, \rho_n, \sigma_{vn}, P_{zn}) \in \Gamma^*$ , then  $\{z_{in}^2\}_{n=i}^{\infty} \xrightarrow{d} z \sim P_z$ , with  $P_{zn} \xrightarrow{d} P_z$ . Once again, **R2** and Corollary 1 imply:

$$\sqrt{n} \beta_n \frac{1}{n} \sum_{i=1}^n z_{in}^2 \xrightarrow{p} b \mathbb{E}_{\gamma} [z^2] = b \mathbb{E}_{\gamma} [f_{\beta}(w, 0, \pi)]$$

#### E.8. Assumption Q5

Note that  $\phi = (\rho, \sigma_v, P_z)$ , where  $P_z$  does not depend on  $\pi$ . Also, since  $(u, v) \sim \mathcal{N}_2(\mathbf{0}, \Omega)$ , where

$$\Omega = \begin{pmatrix} 1 & \rho \sigma_v \\ \rho \sigma_v & \sigma_v^2 \end{pmatrix}$$

does not depend on  $\pi$ , then

$$W(\pi_0, \gamma) = \mathbb{E}_{\phi} [z^2] \mathbb{E}_{\phi} \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix}' \right]$$

does not depend on  $\pi$ .



E.9. *Integrated Likelihood*

PROOF: Let  $x = (x_1, x_2)' \in \mathbb{R}^2$  and  $\gamma = (\gamma_1, \gamma_2)' \sim \mathcal{N}_2(0, \Sigma^0)$ .

$$\begin{aligned}
f_1^*(x_1, x_2; \mu_1^*, \mu_2^*) &= a_1 \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2} \left[ x - \mathbb{E}_\phi[z^2] C^* \gamma \right]' \left[ x - \mathbb{E}_\phi[z^2] C^* \gamma \right] \right) \phi_2(\gamma, \mathbf{0}, \frac{1}{\lambda^2} \Sigma^0) d\gamma \\
&\quad \left( \text{(where I have used the fact } \begin{pmatrix} \dot{L}(0)(\pi(\gamma) - \pi_0) b(\gamma) \\ b(\gamma) \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \right) \\
&= a_2 \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2} \left[ x - \mathbb{E}_\phi[z^2] C^* \gamma \right]' \left[ x - \mathbb{E}_\phi[z^2] C^* \gamma \right] \right) \exp\left(-\frac{1}{2} \lambda^2 \gamma' (\Sigma^0)^{-1} \gamma\right) d\gamma \\
&\quad \text{(By definition of } \phi_2) \\
&= a_2 \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2} \left[ x - \mathbb{E}_\phi[z^2] C^* \gamma \right]' \left[ x - \mathbb{E}_\phi[z^2] C^* \gamma \right] \right) \exp\left(-\frac{1}{2} \lambda^2 \gamma' (C^{*'} C^*) \gamma\right) d\gamma \\
&\quad \text{(since } C^{*'} C^* = (\Sigma^0)^{-1}) \\
&= a_3 \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2} \left[ x - \mathbb{E}_\phi[z^2] \gamma^* \right]' \left[ x - \mathbb{E}_\phi[z^2] \gamma^* \right] \right) \exp\left(-\frac{1}{2} \lambda^2 \gamma^{*'} \gamma^*\right) d\gamma^* \\
&\quad \text{( using the (linear) change of variables } \gamma^* = C^* \gamma) \\
&= a_3 \left[ \int_{\mathbb{R}} \exp\left(-\frac{1}{2} (x_1 - \mathbb{E}_\phi[z^2] \gamma_1^*)^2\right) \exp\left(-\frac{1}{2} \lambda^2 \gamma_1^{*2}\right) d\gamma_1^* \right] \\
&\quad \left[ \int_{\mathbb{R}} \exp\left(-\frac{1}{2} (x_2 - \mathbb{E}_\phi[z] \gamma_2^*)^2\right) \exp\left(-\frac{1}{2} \lambda^2 \gamma_2^{*2}\right) d\gamma_2^* \right] \\
&= a_3 \exp\left(-\frac{1}{2} x_1^2\right) \int_{\mathbb{R}^1} \exp\left(x_1 \mathbb{E}_\phi[z^2] \gamma_1^*\right) \exp\left(-\frac{1}{2} b^2 \gamma_1^{*2}\right) d\gamma_1^* \\
&\quad \exp\left(-\frac{1}{2} x_2^2\right) \int_{\mathbb{R}^{m_2}} \exp\left(x_2 \mathbb{E}_\phi[z^2] \gamma_2^*\right) \exp\left(-\frac{1}{2} b^2 \gamma_2^{*2}\right) d\gamma_2^* \\
&\quad \text{(where } b^2 = \lambda^2 + \mathbb{E}_\phi[z^2]^2) \\
&= c_1 \exp\left(-\frac{1}{2} x_1^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]^2}{2(\lambda^2 + \mathbb{E}_\phi[z^2])} x_1^2\right) \exp\left(-\frac{1}{2} x_2^2\right) \exp\left(\frac{\mathbb{E}_\phi[z^2]}{2(\lambda^2 + \mathbb{E}_\phi[z^2])} x_2^2\right) \\
&\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}$$

*Q.E.D.*

E.10. *Assumption Q6*

PROOF: The proof has three parts:

**PART 1:** We show first that

$$\widehat{S}_n \equiv \frac{1}{n} \sum_{i=1}^n \left[ \begin{pmatrix} y_{in} - L(\widehat{\beta} \pi_0 z_{in}) \\ \widehat{v}_{in} \end{pmatrix} \begin{pmatrix} y_{in} - L(\widehat{\beta} \pi_0 z_{in}) \\ \widehat{v}_{in} \end{pmatrix}' \right]$$

with  $\widehat{v}_{in} = x_{in} - \widehat{\beta} z_{in}$ , converges in probability to:

$$S(\pi_0, \gamma) \equiv \mathbb{E}_\phi \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix}' \right]$$

Let

$$h(W_{in}) \equiv \begin{pmatrix} y_{in} - L(\beta_n \pi_0 z_{in}) \\ v_{in} \end{pmatrix}$$

and

$$k(W_{in}) \equiv \begin{pmatrix} L(\widehat{\beta}\pi_0 z_{in}) - L(\beta_n \pi_0 z_{in}) \\ (\widehat{\beta} - \beta_n) z_{in} \end{pmatrix}$$

And write  $\widehat{S}_n = \frac{1}{n} \sum_{i=1}^n (h(W_{in}) + k_{in}(W_{in}))' (h(W_{in}) + k_{in}(W_{in}))$ .

a) Since  $\phi \in \Phi$ , note first that Corollary 1 implies that for any  $i$

$$h(W_{in}) \xrightarrow{p} \mathbb{E}_\phi \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \right]$$

Therefore,

$$\frac{1}{n} \sum_{i=1}^n h(W_{in})' h(W_{in}) \xrightarrow{p} \mathbb{E}_\phi \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix}' \right]$$

b) Now, it is shown that

$$\frac{1}{n} \sum_{i=1}^n k(W_{in}) \xrightarrow{p} 0$$

Note that  $\frac{1}{n}(\widehat{\beta} - \beta_n) z_{in} \xrightarrow{p} 0$ . In section E.6 we have shown that under any weak sequence:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n L(\widehat{\beta}\pi_0 z_{in}) - L(\beta_n \pi_0 z_{in}) &= (\widehat{\beta} - \beta_n) \sum_{i=1}^n \widetilde{G}((\widehat{\beta} - \beta) \pi_0 z_{in}, \beta_n \pi_0 z_{in}) \\ &= (\widehat{\beta} - \beta_n) \sum_{i=1}^n \widetilde{G}(0, \beta_n \pi_0 z_{in}) + o_{p,\gamma}(1) \\ &\xrightarrow{p} 0 \end{aligned}$$

Therefore, under any weak sequence  $\widehat{S}_n \xrightarrow{p} S(\pi_0, \gamma)$ .

**PART II:** Since  $\phi \in \Phi$ , Corollary 1 implies that:

$$\frac{1}{n} \sum_{i=1}^n z_{in} \xrightarrow{p} \mathbb{E}_\phi[z^2]$$

Hence, under any weak sequence  $\{\gamma_n\} \rightarrow \gamma$

$$\widehat{W}_n = \left( \frac{1}{n} \sum_{i=1}^n z_{in} \right) \widehat{S}_n \xrightarrow{p} \mathbb{E}_\phi[z^2] \mathbb{E}_\phi \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix}' \right] \equiv W(\pi_0, \gamma)$$

**PART III:** Finally, we show that under any weak sequence

$$\pi_0 \frac{1}{n} \sum_{i=1}^n z_{in}^2 L(\widehat{\beta}\pi_0 z_{in}) \xrightarrow{p} \pi_0 \mathbb{E}_\phi[z^2] \dot{L}(0)$$

Write

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n z_{in}^2 L(\widehat{\beta}\pi_0 z_{in}) &= \frac{1}{n} \sum_{i=1}^n z_{in}^2 [L(\widehat{\beta}\pi_0 z_{in}) - L(\beta_n \pi_0 z_{in})] + \frac{1}{n} \sum_{i=1}^n z_{in}^2 L(\beta_n \pi_0 z_{in}) \\
&= \frac{1}{n} \sum_{i=1}^n z_{in}^2 \widetilde{G}(0, \beta_n \pi_0 z_{in}) + \mathbb{E}_\phi[z^2] \dot{L}(0) + o_{p,\gamma}(1) \\
&\quad (\text{where I have used Part II above and Corollary 1})
\end{aligned}$$

Let  $b_0 = [0, 1]$ . Part I, II, III imply that under any weak sequence:

$$\widehat{\Sigma} \equiv \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} \frac{1}{n} \sum_{i=1}^n \left[ \begin{pmatrix} y_{in} - L(\widehat{\beta}\pi_0 z_{in}) \\ \widehat{v}_{in} \end{pmatrix} \begin{pmatrix} y_{in} - L(\widehat{\beta}\pi_0 z_{in}) \\ \widehat{v}_{in} \end{pmatrix}' \right] \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}'$$

converges in probability to

$$\begin{aligned}
\Sigma(\pi_0, \gamma) &\equiv \begin{bmatrix} 1 & -\pi_0 \mathbb{E}_\phi[z^2] \dot{L}(0) \\ 0 & 1 \end{bmatrix} W(\pi_0, \gamma) \begin{bmatrix} 1 & -\pi_0 \mathbb{E}_\phi[z^2] \dot{L}(0) \\ 0 & 1 \end{bmatrix}', \\
W(\pi_0, \gamma) &= \mathbb{E}_\phi[z^2] \mathbb{E}_\phi \left[ \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix} \begin{pmatrix} \mathbf{1}(u > 0) - L(0) \\ v \end{pmatrix}' \right]
\end{aligned}$$

Since  $C^*$  is a continuous transformation of  $\Sigma(\pi_0, \gamma)$ , then the same transformation applied to  $\widehat{\Sigma}$  provides a consistent estimator for  $C^*$ .

*Q.E.D.*

## APPENDIX F: NONLINEAR REGRESSION

Let  $n$  be the sample size. Consider the econometric model:

$$y_{in} = \beta h(X_{in}, \pi) + u_{in} \quad \text{for } i = 1, \dots, n$$

where  $h(X_{in}, \pi) \in \mathbb{R}$  is known up to a finite dimensional parameter  $\pi \in \mathbb{R}^{d\pi}$ . It is assumed that  $\{(X'_{in}, u_{in})'\}_{i=1}^n$  is an i.i.d. collection of random vectors with joint distribution  $P_{xu}$ . The (optimization) parameter space of interest is given by:

$$\Theta = \{(\beta, \pi) \in [-b_1, b_2] \times \Pi^* \subseteq \mathbb{R} \times \mathbb{R}^{d\pi} \mid b_1, b_2 > 0, \Pi^* \text{ a compact subset of } \mathbb{R}^{d\pi}\}$$

Let  $W_{in} = (y_{in}, X'_{in})'$ , with  $y_{in} \in \mathbb{R}$  and  $X'_{in} \in \mathbb{R}^r$ . The i.i.d. data  $W^n$  is given by  $W^n = (W_{1n}, W_{2n}, \dots, W_{nn}) \in \mathbb{R}^{(1+r) \times n}$ . Note that the distribution of  $W_{in}$  is fully specified by the tuple:

$$\gamma = (\beta, \pi, \phi) \quad \text{and} \quad \phi = (P_{ux})$$

where  $P_{ux}$  is a Borel probability measure on  $\mathbb{R}^{1+r}$ . See F.1 for restrictions on the elements  $\phi$ . It is assumed that  $h(\cdot)$  and  $h_\pi(\cdot)$  are both sequentially continuous and that

$$\left| h(x, \pi) - h(x, \pi_0) \right| < M(x) \|\pi - \pi_0\| \quad \text{for all } \pi \in \Pi, \text{ and a.e. } x$$

Two examples of nonlinear regression models are the given in AC12. The first one is the Box-Cox type transformation:

$$(F.1) \quad h(x, \pi) = \frac{|x|^\pi - 1}{\pi}, \quad \pi \in [1, 4].$$

The second one is a smooth transition model:

$$(F.2) \quad h(x, \pi) = \frac{x}{1 + \exp\left(- (x - \pi)\right)}, \quad \pi \in [4, 9]$$

The following exponential model will be introduced in this section and used in the Monte-Carlo exercises in the appendix:

$$(F.3) \quad h(x, \pi) = \exp(\pi x), \quad \pi \in \Pi$$

The least squares sample criterion function is given by:

$$(F.4) \quad Q_n(\beta, \pi) = \frac{1}{n} \sum_{i=1}^n \left( y_{in} - \beta h(X_{in}, \pi) \right)^2$$

F.1. Restrictions on  $\Phi$ 

Let  $P_{ux}$  be a Borel probability measure on  $\mathbb{R}^{1+r}$ . Denote the space of such measures as  $\Delta_B(\mathbb{R}^{1+r})$  and restrict attention to the set:

$$\begin{aligned} \Phi(\pi_0) &\equiv \{P_{xu} \in \Delta_B(\mathbb{R}^{1+r}) \text{ s.t the following restrictions hold} \\ &\mathbf{R1)} \quad \mathbb{E}_{P_{xu}}[u \mid X] = 0 \text{ a.s.} \\ &\mathbf{R2)} \quad \mathbb{E}_{P_{xu}}[|u|^{4+\epsilon}] < c_2 \\ &\mathbf{R3)} \quad \mathbb{E}_{P_{xu}}[|h(X, \pi_0)|^{4+\epsilon}] < c_3 \\ &\mathbf{R4)} \quad \mathbb{E}_{P_{xu}}[|h_\pi(X, \pi_0)|^{4+\epsilon}] < c_4 \\ &\mathbf{R5)} \quad \mathbb{E}_{P_{xu}}[|M(X)|^{4+\epsilon}] < c_5 \\ &\mathbf{R6)} \quad \mathbb{E}_{P_{xu}}[\sup_{\pi} |h(X, \pi)|^{2+\epsilon}] < c_5 \\ &\text{for a fixed } \epsilon > 0. \\ &\} \end{aligned}$$

We will say that a sequence  $\phi_n = (P_{xu,n})$  of elements in  $\Phi(\pi_0)$  converges to  $P_{xu}$  if and only if

$$P_{xun} \xrightarrow{d} P_{xu}.$$

F.2. *Assumption Q1: Derivative of  $Q_n(\beta, \pi)$*

$$D_\beta Q_n(0, \pi_0) = 2 \frac{1}{n} \sum_{i=1}^n y_{in} h(X_{in}, \pi_0)$$

Hence Assumption Q1 is easily verified with:

$$g_1(W^n) = 1, \quad m(W_{in}, \pi_0) \equiv y_{in} h(X_{in}, \pi_0)$$

F.3. *Assumption Q2: Identification Statistic*

Let  $\{\gamma_n = (\beta_n, \pi_0, \phi_n)\}_{n \in \mathbb{N}}$  be an arbitrary sequence in  $\Gamma^*$  with limit  $(\beta, \pi_0, \phi)$ . Consider:

$$\begin{aligned} \widehat{\beta}(\pi_0) &\equiv \left( \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) h_\pi(X_{in}, \pi_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^n y_{in} h_\pi(X_{in}, \pi_0) \\ &= \beta_n + \left( \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) h_\pi(X_{in}, \pi_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^n u_{in} h_\pi(X_{in}, \pi_0) \end{aligned}$$

Note that

$$\{h(X_{in}, \pi_0) h_\pi(X_{in}, \pi_0)\}_{i=1}^n, n \in \mathbb{N}$$

is a triangular array of row-wise i.i.d. random variables. Since  $\gamma_n \in \Gamma^*$ , it follows that  $P_{u_x n} \xrightarrow{d} P_{u_x}$  for some  $P_{u_x}$  in  $\Phi(\theta_0)$ . In addition, for arbitrary  $i$ , restrictions **R3**) and **R4**) in the set  $\Phi(\theta_0)$  implies

$$\sup_{n \geq i} \mathbb{E}_{\gamma_n} \left[ \left| h(X_{in}, \pi_0) h_\pi(X_{in}, \pi_0) \right|^{2+\epsilon/2} \right] \leq \sup_{n \geq i} \mathbb{E}_{\gamma_n} \left[ \left| h(X_{in}, \pi_0) \right|^{4+\epsilon} \right]^{1/2} \mathbb{E}_{\gamma_n} \left[ \left| h_\pi(X_{in}, \pi_0) \right|^{4+\epsilon} \right]^{1/2} \leq c_3 c_4.$$

Therefore, from Corollary 1

$$(F.5) \quad \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) h_\pi(X_{in}, \pi_0) \xrightarrow{P} \mathbb{E}_\gamma [h(X, \pi_0) h_\pi(X, \pi_0)].$$

Consider now the row-wise i.i.d. triangular array given by  $\{u_{in} h_\pi(X_{in}, \pi_0)\}_{i=1}^n, n \in \mathbb{N}$ , where—by assumption— $h_\pi(X_{in}, \pi_0)$  is continuous a.e. Note that for arbitrary  $i$ ,

$$\{u_{in} h_\pi(X_{in}, \pi_0)\}_{n=i}^\infty \xrightarrow{d} u h_\pi(X, \pi_0)$$

where  $(u, X) \sim P_{u_x}$ . Furthermore, note that  $\gamma_n \in \Gamma$  and Cauchy-Schwartz implies

$$\sup_{n \geq i} \mathbb{E}_{\gamma_n} \left[ \left| u_{in} h_\pi(X_{in}, \pi_0) \right|^{2+\epsilon/2} \right] \leq \sup_{n \geq i} \mathbb{E}_{\gamma_n} \left[ \left| u_{in} \right|^{4+\epsilon} \right]^{1/2} \mathbb{E}_{\gamma_n} \left[ \left| h_\pi(X_{in}, \pi_0) \right|^{4+\epsilon} \right]^{1/2} \leq c_2 c_4.$$

where the last inequality follows by the definition of  $\Gamma$ . Corollary 2 implies:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h_\pi(X_{in}, \pi_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{E}_\gamma [u^2 h_\pi^2(X, \pi_0)]).$$

Hence, for an arbitrary sequence  $\{\gamma_n = (\beta_n, \pi_0, \phi_n)\}_{n \in \mathbb{N}} \in \Gamma^*$  with limit  $(\beta, \pi_0, \phi)$ :

$$\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n)$$

is asymptotically normal.

Hence assumption Q2 is satisfied with:

$$g_2(W^n) = \left( \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) h_\pi(X_{in}, \pi_0) \right)^{-1}, \quad f(W_{in}, \pi_0) \equiv y_{in} h_\pi(X_{in}, \pi_0),$$

$$a(\pi_0, \gamma) \equiv \mathbb{E}_\gamma [h(X_{in}, \pi_0) h_\pi(X_{in}, \pi_0)]^{-1}.$$

F.4. *Assumption Q3:  $m^*(W_{in}, \beta, \pi)$  and  $f^*(W_{in}, \beta, \pi)$*

In the nonlinear regression model the data  $W_{in} = (y_{in}, X'_{in})' \in \mathbb{R}^{r+1}$ . Consider the real valued functions:

$$m^*(W_{in}, \beta, \pi) = \beta h(X_{in}, \pi) h(X_{in}, \pi_0) \quad \text{and} \quad f^*(W_{in}, \beta, \pi) = \beta h(X_{in}, \pi) h_\pi(X_{in}, \pi_0)$$

Note that:

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n m(W_{in}, \pi_0) - m^*(W_{in}, \beta_n, \pi_n) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n f(W_{in}, \pi_0) - f^*(W_{in}, \beta, \pi_n) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h(X_{in}, \pi_0) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h_\pi(X_{in}, \pi_0) \end{pmatrix}.$$

It is now shown that for arbitrary  $a, b \in \mathbb{R}$ , the linear combination

$$(F.6) \quad a \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h(X_{in}, \pi_0) + b \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h_\pi(X_{in}, \pi_0)$$

follows a Central Limit Theorem under arbitrary sequences of true parameter values  $\gamma_n \in \Gamma^*$ .

To establish the asymptotic normality of F.6, the assumptions of Corollary 2 are now verified. Note first that:

$$\mathbb{E}_{\gamma_n} \left[ u_{in} h(X_{in}, \pi_0) \right] = 0 \quad \text{and} \quad \mathbb{E}_{\gamma_n} \left[ u_{in} h_\pi(X_{in}, \pi_0) \right] = 0,$$

as  $\mathbb{E}_{\gamma_n} [u_{in} | X_{in}] = 0$  a.s. Note also that,

$$\gamma_n = (\beta_n, \pi_n, P_{u_x n})$$

converges to

$$\gamma = (\beta, \pi, P_{u_x}) \in \Gamma,$$

which implies—using the continuity of  $h(\cdot, \pi_0)$  and  $h_\pi(\cdot, \pi_0)$ —that the sequence of mean-zero random variables

$$\left\{ a u_{in} h(X_{in}, \pi_0) + b u_{in} h_\pi(X_{in}, \pi_0) \right\}_{n=i}^\infty \xrightarrow{d} a u h(X, \pi_0) + b u h_\pi(X, \pi_0), \quad (u, x) \sim P_{u_x}$$

Now, it is shown that

$$\sup_n \mathbb{E}_{\gamma_n} \left[ \left| a u_{in} h(X_{in}, \pi_0) + b u_{in} h_\pi(X_{in}, \pi_0) \right|^{2+(\epsilon/2)} \right] < \infty$$

By Minkowski's inequality:

$$\begin{aligned} \mathbb{E}_{\gamma_n} \left[ \left| a u_{in} h(X_{in}, \pi_0) + b u_{in} h_\pi(X_{in}, \pi_0) \right|^{2+(\epsilon/2)} \right]^{\frac{1}{2+(\epsilon/2)}} &\leq |a| \mathbb{E}_{\gamma_n} \left[ \left| u_{in} h(X_{in}, \pi_0) \right|^{2+(\epsilon/2)} \right]^{\frac{1}{2+(\epsilon/2)}} \\ &\quad + |b| \mathbb{E}_{\gamma_n} \left[ \left| u_{in} h_\pi(X_{in}, \pi_0) \right|^{2+\epsilon} \right]^{\frac{1}{2+(\epsilon/2)}} \\ &\leq |a| \mathbb{E}_{\gamma_n} \left[ \left| u_{in} \right|^{4+\epsilon} \right]^{\frac{1}{4+\epsilon}} \mathbb{E}_{\gamma_n} \left[ \left| h(X_{in}, \pi_0) \right|^{4+\epsilon} \right]^{\frac{1}{4+\epsilon}} \\ &\quad + |b| \mathbb{E}_{\gamma_n} \left[ \left| u_{in} \right|^{4+\epsilon} \right]^{\frac{1}{4+\epsilon}} \mathbb{E}_{\gamma_n} \left[ \left| h_\pi(X_{in}, \pi_0) \right|^{4+\epsilon} \right]^{\frac{1}{4+\epsilon}} \\ &\quad \text{(by Cauchy-Scharwz)} \\ &\leq |a| (c_2 c_3)^{\frac{1}{4+\epsilon}} + |b| (c_2 c_4)^{\frac{1}{4+\epsilon}} \\ &\quad \text{(since } \gamma_n \text{ satisfies } \mathbf{R2}, \mathbf{R3}, \mathbf{R4}) \end{aligned}$$

The assumptions of Corollary 2 are thus verified. Therefore, under any sequence  $\{\gamma_n\} \in \Gamma^*$  with limit  $\gamma$ :

$$a \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h(X_{in}, \pi_0) + b \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h_\pi(X_{in}, \pi_0) \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_\gamma[(a u_{in} h(X_{in}, \pi_0) + b u_{in} h_\pi(X_{in}, \pi_0))^2]),$$

for any  $a, b \in \mathbb{R}$ . The asymptotic variance can be written as  $(a, b)W^*(\pi_0, \gamma)(a, b)'$  with

$$W^*(\pi_0, \gamma) \equiv \mathbb{E}_\gamma \left[ \begin{pmatrix} uh(X, \pi_0) \\ uh_\pi(X, \pi_0) \end{pmatrix} \begin{pmatrix} uh(X, \pi_0) \\ uh_\pi(X, \pi_0) \end{pmatrix}' \right].$$

Therefore, Assumption Q3 is satisfied with

$$m^*(W_{in}, \beta, \pi) \equiv \beta h(X_{in}, \pi_0) h(X_{in}, \pi) \quad \text{and} \quad f^*(W_{in}, \beta, \pi) = \beta h_\pi(X_{in}, \pi_0) h(X_{in}, \pi)$$

Hence,

$$m(W_{in}, \beta, \pi) - m^*(W_{in}, \beta, \pi) = [y_{in} - \beta h(X_{in}, \pi)] h(X_{in}, \pi_0),$$

$$f(W_{in}, \beta, \pi) - f^*(W_{in}, \beta, \pi) = [y_{in} - \beta h(X_{in}, \pi)] h_\pi(X_{in}, \pi_0).$$

#### F.5. Assumption Q4.1: First estimation error

Remember

$$m^*(W_{in}, \beta, \pi) = \beta h(X_{in}, \pi) h(X_{in}, \pi)$$

$$\begin{aligned} \sqrt{n} \Delta_{m^*}^\pi(\pi_n, \pi) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m^*(W_{in}, \beta_n, \pi_n) - m^*(W_{in}, \beta_n, \pi_0) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta_n h(X_{in}, \pi_n) h(X_{in}, \pi_0) - \beta_n h(X_{in}, \pi_0) h(X_{in}, \pi_0) \\ &= \sqrt{n} \beta_n \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) [h(X_{in}, \pi_n) - h(X_{in}, \pi_0)] \end{aligned}$$

Note that  $\sqrt{n} \beta_n \rightarrow b$  as  $\{\gamma_n\} \in \Gamma^*(\gamma, 0)$ . The asymptotic behavior of

$$\frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) [h(X_{in}, \pi_n) - h(X_{in}, \pi_0)]$$

is established using Corollary 1. Note first that the continuity of  $h$  implies

$$\{h(X_{in}, \pi_0) [h(X_{in}, \pi_n) - h(X_{in}, \pi_0)]\}_{n=i}^\infty \xrightarrow{d} h(X, \pi_0) [h(X, \pi) - h(X, \pi_0)]$$

where  $x \sim P_x$  and  $P_x$  is the marginal distribution of  $P_{u,x} \in \Phi(\pi_0)$ . Note that:

$$\begin{aligned} &\mathbb{E}_{\gamma_n} \left[ |h(X_{in}, \pi_0)|^{2+\epsilon/2} |h(X_{in}, \pi_n) - h(X_{in}, \pi_0)|^{2+\epsilon/2} \right] \\ &\leq \mathbb{E}_{\gamma_n} \left[ |h(X_{in}, \pi_0)|^{4+\epsilon} \right]^{1/2} \mathbb{E}_{\gamma_n} \left[ |h(X_{in}, \pi_n) - h(X_{in}, \pi_0)|^{4+\epsilon} \right]^{1/2} \\ &= \mathbb{E}_{\gamma_n} \left[ |h(X_{in}, \pi_0)|^{4+\epsilon} \right]^{1/2} \mathbb{E}_{\gamma_n} \left[ |M(X_{in})|^{4+\epsilon} \right]^{1/2} d(|\pi_n - \pi_0|)^{1/2} \\ &\leq (c_3 c_5)^{1/2} c \end{aligned}$$

(where I have used **R3**, **R5**) and  $\pi_n \rightarrow \pi$ )

The assumptions of Corollary 1 are verified and:

$$\sqrt{n}\Delta_{m^*}^{\pi}(\pi_n, \pi) \xrightarrow{P} b\mathbb{E}_{\gamma}[h(X, \pi_0)[h(X, \pi) - h(X, \pi_0)]]$$

F.6. *Second Estimation Error*

$$\begin{aligned} \sqrt{n}\Delta_{m^*}^{\beta}(\beta_n, \widehat{\beta}(\pi_0)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m^*(W_{in}, \beta_n, \pi_0) - m^*(W_{in}, \widehat{\beta}(\pi_0), \pi_0) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta_n h(X_{in}, \pi_0) h(X_{in}, \pi_0) - \widehat{\beta}(\pi_0) h(X_{in}, \pi_0) h(X_{in}, \pi_0) \\ &= -\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n) \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0)^2 \\ &= -\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n) \mathbb{E}_{\gamma}[h(X, \pi_0)^2] \\ &\quad - \sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n) \left[ \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0)^2 - \mathbb{E}_{\gamma}[h(X, \pi_0)^2] \right] \end{aligned}$$

By **R3**) and Corollary 1

$$\left[ \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0)^2 - \mathbb{E}_{\gamma}[h(X, \pi_0)^2] \right] \xrightarrow{P} 0$$

And note that

$$\begin{aligned} \sqrt{n}\Delta_{m^*}^{\beta}(\beta_n, \widehat{\beta}(\pi_0)) &= \sqrt{n}\beta_n \left( \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) h_{\pi}(X_{in}, \pi_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_n) h_{\pi}(X_{in}, \pi_0) \\ &\quad + \left( \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0) h_{\pi}(X_{in}, \pi_0) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h_{\pi}(X_{in}, \pi_0) \\ &\quad - \sqrt{n}\beta_n \\ &= b \mathbb{E}_{\gamma}[h(X, \pi_n) h(X, \pi_0)] \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_n) h_{\pi}(X_{in}, \pi_0) \\ &\quad + \mathbb{E}_{\gamma}[h(X, \pi_n) h(X, \pi_0)] \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in} h_{\pi}(X_{in}, \pi_0) \\ &\quad + b + o_{p, \gamma_n}(1) \\ &\quad \text{(where I have used equation F.5)} \\ &= b \mathbb{E}_{\gamma}[h(X, \pi_n) h(X, \pi_0)] \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_n) h_{\pi}(X_{in}, \pi_0) + O_{p, \gamma_n}(1) \\ &\quad \text{(where I have used equation F.6)} \end{aligned}$$

Note that:

$$\{h(X_{in}, \pi_n) h_{\pi}(X_{in}, \pi_0)\} \xrightarrow{d} h(X, \pi) h_{\pi}(X, \pi_0)$$



Furthermore:

$$\begin{aligned}
\mathbb{E}_{\gamma_n} [|h(X, \pi_n)h_\pi(X, \pi_0)|^{1+\epsilon/2}] &\leq \mathbb{E}_{\gamma_n} [|h(X, \pi_n)|^{2+\epsilon/2}]^{1/2} \mathbb{E}_{\gamma_n} [|h_\pi(X, \pi_0)|^{2+\epsilon}]^{1/2} \\
&\leq \mathbb{E}_{\gamma_n} [|h(X, \pi_n)|^{2+\epsilon}]^{1/2} c^* \\
&\quad \text{(where I have used **R4**)} \\
&\leq \mathbb{E}_{\gamma_n} [\sup_{\pi} |h(X, \pi)|^{2+\epsilon}]^{1/2} c^* \\
&\leq c_6^{1/2} c^* \\
&\quad \text{(where I have used **R6**)}
\end{aligned}$$

Corollary 1 implies:

$$\sqrt{n}\Delta_{m^*}^{\beta}(\beta_n, \widehat{\beta}(\pi_0)) = -\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n)\mathbb{E}_{\gamma}[h(X, \pi_0)^2] + o_{p, \gamma_n}(1)$$

Since

$$m_{\beta}(W_{\in}, \beta, \pi) = h(X_{in}, \pi)h(X_{in}, \pi_0)$$

the result follows.

Assumption Q4.2 is satisfied and the second estimation error behaves as:

$$\sqrt{n}\Delta_{m^*}^{\beta}(\beta_n, \widehat{\beta}(\pi_0)) = -\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n)\mathbb{E}_{\gamma}[h(X, \pi_0)^2] + o_{p, \gamma_n}(1)$$

Note, however, that under weak sequences that do not impose the null hypothesis:

$$\begin{aligned}
\sqrt{n}(\widehat{\beta}(\pi_0) - \beta_n) &= \left( \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_0)h_{\pi}(X_{in}, \pi_0) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in}h_{\pi}(X_{in}, \pi_0) \\
&+ \sqrt{n}\beta_n \left[ \left( \mathbb{E}_{\gamma}[h(X_{in}, \pi_0)h_{\pi}(X_{in}, \pi_0)] \right)^{-1} \mathbb{E}_{\gamma}[h(X_{in}, \pi)h_{\pi}(X_{in}, \pi_0)] - 1 \right] + o_{p, \gamma_n}(1).
\end{aligned}$$

Hence, there is additional bias when  $\pi \neq \pi_0$ . This will impact the centrality parameter of the asymptotic distribution for  $\beta$ -score-star

#### F.7. Assumption AQ4.3

Finally,

$$f^*(W_{in}, \beta, \pi) = \beta h(X_{in}, \pi)h_{\pi}(X_{in}, \pi_0),$$

then

$$\frac{1}{\sqrt{n}}f^*(W_{in}, \beta_n, \pi_n) = \sqrt{n}\beta_n \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_n)h_{\pi}(X_{in}, \pi_0),$$

Since  $\gamma^* = (\beta_n, \pi_n, P_{u|x_n}) \in \Gamma^*$ , then **R4**, **R6** and Corollary 1 imply

$$\sqrt{n}\beta_n \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi_n)h_{\pi}(X_{in}, \pi_0) \xrightarrow{p} b\mathbb{E}_{\gamma}[h(X, \pi)h_{\pi}(X, \pi_0)] = b\mathbb{E}_{\gamma}[f^*(w, 0, \pi)]$$

Assumption Q4.3 is satisfied and:

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n f^*(W_{in}, \beta_n, \pi_n) &= \sqrt{n}\beta_n \frac{1}{n} \sum_{i=1}^n h(X_{in}, \pi)h_{\pi}(X_{in}, \pi_0) \\
&\xrightarrow{p} b\mathbb{E}_{\gamma}[h(X_{in}, \pi)h_{\pi}(X_{in}, \pi_0)]
\end{aligned}$$

F.8. *Statistical Model for Nonlinear regression*

In the nonlinear regression model:

$$a(\pi_0, \gamma) = \mathbb{E}_\gamma[h(x, \pi_0)h_\pi(x, \pi_0)]^{-1}$$

and

$$\mathbb{E}_\gamma[f_\beta^*(w, 0, \pi)] = \mathbb{E}_\gamma[h(x, \pi)h_\pi(x, \pi_0)]$$

so,

$$a^*(\pi_0, \gamma) = \mathbb{E}_\gamma[h(x, \pi)h_\pi(x, \pi_0)]\mathbb{E}_\gamma[h(x, \pi_0)h_\pi(x, \pi_0)]^{-1}$$

Therefore, the statistical model in Proposition 1 is given by:

$$(F.7) \quad \begin{pmatrix} D_\beta^* \\ \beta^* \end{pmatrix} \sim \mathcal{N}_2 \left( \begin{pmatrix} \mathbb{E}_\gamma[h(x, \pi_0)h(x, \pi)] - a^*(\pi_0, \gamma)\mathbb{E}_\gamma[h(x, \pi_0)^2] \\ \mathbb{E}_\gamma[h(x, \pi)h_\pi(x, \pi_0)] \end{pmatrix} b, \Sigma(\pi_0, \gamma) \right),$$

where  $\Sigma(\pi_0, \gamma)$  is given by:

$$A(\pi_0, \gamma) \mathbb{E}_\gamma \left[ \begin{pmatrix} uh(X, \pi_0) \\ uh_\pi(X, \pi_0) \end{pmatrix} \begin{pmatrix} uh(X, \pi_0) \\ uh_\pi(X, \pi_0) \end{pmatrix}' \right] A(\pi_0, \gamma)'$$

and

$$A(\pi_0, \gamma) \equiv \begin{bmatrix} 1 & -\mathbb{E}_\gamma[h(x, \pi_0)^2]\mathbb{E}_\gamma[h(x, \pi_0)h_\pi(x, \pi_0)]^{-1} \\ 0 & 1 \end{bmatrix}.$$

Once again, if the distribution of  $(u, x)$  does not depend on  $\pi$ , the matrix  $\Sigma(\pi_0, \gamma)$  does not depend on  $\pi$ . This verifies Assumption Q5.

F.9. *Boundedly-complete, boundary sufficient statistic for Nonlinear regression*

Just as before, let

$$\Sigma^0 = \Sigma(\pi_0, (0, \pi, \phi)) \equiv \begin{pmatrix} \sigma_D^2 & \sigma_{D\beta} \\ \sigma_{D\beta} & \sigma_\beta^2 \end{pmatrix}.$$

So that

$$C^* = \begin{pmatrix} 1/\sigma_D & 0 \\ -\rho/\sqrt{1-\rho^2}\sigma_D & 1/\sqrt{1-\rho^2}\sigma_\beta \end{pmatrix}$$

is used to rotate  $(D_\beta^*, \beta^*)$ . Again, the boundary-sufficient statistic for the this model is:

$$X_2 = \left[ \beta^*/\sigma_\beta - \rho D_\beta^*/\sigma_D \right] / \sqrt{1-\rho^2}.$$

and

$$\begin{aligned} \mu_1^*(\pi_0, b, \pi, \phi) &= \left( \mathbb{E}_\phi[h(x, \pi_0)h(x, \pi)] - a^*(\pi_0, \phi)\mathbb{E}_\gamma[h(x, \pi_0)^2] \right) b/\sigma_D, \\ \mu_2^*(\pi_0, b, \pi, \phi) &= b \left[ \mathbb{E}_\phi[h(x, \pi)h_\pi(x, \pi_0)]/\sigma_\beta \right. \\ &\quad \left. - \rho \left( \mathbb{E}_\phi[h(x, \pi_0)h(x, \pi)] - a^*(\pi_0, \phi)\mathbb{E}_\gamma[h(x, \pi_0)^2] \right) / \sigma_D \right] / \sqrt{1-\rho^2}. \end{aligned}$$

and  $a^*(\pi_0, \gamma) = \mathbb{E}_\phi[h(x, \pi)h_\pi(x, \pi_0)]\mathbb{E}_\phi[h(x, \pi_0)h_\pi(x, \pi_0)]^{-1}$ . Note that if  $\rho < 1, \sigma_\beta > 0$  and

$$\mathbb{E}_\phi[h(x, \pi_0)h_\pi(x, \pi_0)] \neq 0$$

then  $X_2$  is boundedly complete, as the function  $\mu_2^*(\pi_0, \cdot, \pi_0, \phi)$  is onto on  $\mathbb{R}$ .

F.10. ECS test for the limiting statistical model of nonlinear regression and results for general implementation of ECS

From Proposition 2 it follows that the limiting statistical model associated to the nonlinear regression model is given by:

$$\begin{aligned} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} &\equiv \begin{pmatrix} D_\beta^*/\sigma_D \\ [\beta^*/\sigma_\beta - \rho D_\beta^*/\sigma_D]/\sqrt{1-\rho^2} \end{pmatrix} \\ &\sim \mathcal{N}_2 \left( \begin{array}{c} \mu_1^*(\pi_0, b, \pi; \phi) \\ \mu_2^*(\pi_0, b, \pi; \phi) \end{array}, \mathbb{I}_2 \right) \end{aligned}$$

where

$$\begin{aligned} \mu_1^*(\pi_0, b, \pi, \phi) &= \left( \mathbb{E}_\phi[h(x, \pi_0)h(x, \pi)] - a^*(\pi_0, \phi)\mathbb{E}_\gamma[h(x, \pi_0)^2] \right) b/\sigma_D, \\ \mu_2^*(\pi_0, b, \pi, \phi) &= b \left[ \mathbb{E}_\phi[h(x, \pi)h_\pi(x, \pi_0)]/\sigma_\beta \right. \\ &\quad \left. - \rho \left( \mathbb{E}_\phi[h(x, \pi_0)h(x, \pi)] - a^*(\pi_0, \phi)\mathbb{E}_\gamma[h(x, \pi_0)^2] \right) / \sigma_D \right] / \sqrt{1-\rho^2}. \end{aligned}$$

and  $a^*(\pi_0, \gamma) = \mathbb{E}_\phi[h(x, \pi)h_\pi(x, \pi_0)]\mathbb{E}_\phi[h(x, \pi_0)h_\pi(x, \pi_0)]^{-1}$

**Integrated Likelihood:** In this and other examples it might be desirable to compute the integrated likelihood

$$\int_{\mathbb{R}^{d_\beta} \times \Pi} \exp\left(-\frac{1}{2}\|x_1 - \mu_1^*(\pi_0, b, \pi; \phi)\|^2\right) \exp\left(-\frac{1}{2}\|x_2 - \mu_2^*(\pi_0, b, \pi; \phi)\|^2\right) dP_1(b, \pi).$$

by means of numerical integration. In such cases, it is necessary to evaluate the functions  $\mu_j^*(\pi_0, b, \pi; \phi)$ ,  $j = 1, 2$ , which depend on the unknown parameter  $\phi$ . Consider the following assumption:

**ASSUMPTION Q7:** There exists functions  $\widehat{\mu}_j^*(\pi_0, b, \pi)$  such that under any sequence  $\phi_n \rightarrow \phi$ ,

$$\widehat{\mu}_{j,n}^*(\pi_0, b, \pi) \xrightarrow{P} \mu_j^*(\pi_0, b, \pi; \phi)$$

for almost every  $(b, \pi) \in \mathbb{R}^{d_\beta} \times \Pi$ ,  $j = 1, 2$ .

Under Assumption 7 the following integrated likelihood for ECS tests is available:

$$f_1^*(x_1, x_2; \widehat{\mu}_n^*) \equiv \int_{\mathbb{R}^{d_\beta} \times \Pi} \exp\left(-\frac{1}{2}\|x_1 - \widehat{\mu}_{1,n}^*(\pi_0, b, \pi; \phi)\|^2\right) \exp\left(-\frac{1}{2}\|x_2 - \widehat{\mu}_{2,n}^*(\pi_0, b, \pi; \phi)\|^2\right) dP_1(b, \pi).$$

EXAMPLE—NONLINEAR REGRESSION: In the nonlinear regression model Assumption 7 is satisfied with:

$$\begin{aligned} \widehat{\mu}_1^*(\pi_0, b, \pi) &= \left( \frac{1}{n} \sum_{i=1}^n [h(x_{in}, \pi_0)h(x_{in}, \pi)] - \widehat{a}^*(\pi_0) \frac{1}{n} \sum_{i=1}^n [h(x_{in}, \pi_0)^2] \right) b/\widehat{\sigma}_D, \\ \widehat{\mu}_2^*(\pi_0, b, \pi) &= b \left[ \frac{1}{n} \sum_{i=1}^n [h(x_{in}, \pi)h_\pi(x_{in}, \pi_0)]/\widehat{\sigma}_\beta \right. \\ &\quad \left. - \widehat{\rho} \left( \frac{1}{n} \sum_{i=1}^n [h(x_{in}, \pi_0)h(x_{in}, \pi)] - \widehat{a}^*(\pi_0) \frac{1}{n} \sum_{i=1}^n [h(x_{in}, \pi_0)^2] \right) / \widehat{\sigma}_D \right] / \sqrt{1-\widehat{\rho}^2}. \end{aligned}$$

$\widehat{a}^*(\pi_0) = \frac{1}{n} \sum_{i=1}^n [h(x_{in}, \pi)h_\pi(x_{in}, \pi_0)] \frac{1}{n} \sum_{i=1}^n [h(x_{in}, \pi_0)h_\pi(x_{in}, \pi_0)]^{-1}$  and  $(\widehat{\sigma}_D, \widehat{\sigma}_\beta, \widehat{\rho}) \xrightarrow{P} (\sigma_D, \sigma_\beta, \rho)$ .

Note that Lebesgue's dominated theorem (for convergence in probability) implies that  $f_1^*(x_1, x_2; \widehat{\mu}_n^*) \xrightarrow{P} f_1^*(x_1, x_2; \widehat{\mu}^*)$  pointwise in  $(x_1, x_2)$ .

**Critical Value Function:** Let

$$z(x_1, x_2; \mu^*) \equiv \frac{f_1^*(x_1, x_2; \mu_1^*, \mu_2^*)}{f_{\text{Bd}}(x_1 | x_2)}.$$

For a fixed  $x_2$ , let  $c(x_2, \alpha; \mu^*)$  be 1- $\alpha$  quantile of the  $z(X_1, x_2; \mu^*)$ , where  $X_1 \sim \mathcal{N}_{d_\beta}(\mathbf{0}, \mathbb{I}_{d_\beta})$ . Likewise, let  $c(x_2, \alpha; \widehat{\mu}_n)$  be the 1- $\alpha$  quantile of  $z(X_1, x_2; \widehat{\mu}_n)$ . To denote the critical value, let  $G(\mu, x_2)$  denote the distribution of  $z(X_1, x_2; \mu)$  where  $\mu : \mathbb{R}^{d_\beta} \times \Pi \rightarrow \mathbb{R}^{2d_\beta}$ . Then the critical value is given by:

$$c(x_2, \alpha; \mu) = \inf_t \{G(\mu, x_2)(t) \leq 1 - \alpha\}.$$

We will show that the test

$$\phi_{\text{ECS}}(x_1, x_2; \widehat{\mu}_n) \equiv \begin{cases} 1 & \text{if } z(x_1, x_2; \widehat{\mu}_n) - c(x_2, \alpha; \widehat{\mu}_n) > 0. \\ 0 & \text{i.o.c.} \end{cases}$$

evaluated at sample analogues is asymptotically similar and achieves the same WAP of the ECS test that treats  $\mu^*$  as known.

*Notation:* Let  $X_n \equiv (\widehat{x}_1, \widehat{x}_2)$  be the sample analogues of  $X \equiv (x_1, x_2)$ . Just as before, let  $P(b, \pi; \mu^*(b, \pi; \phi^*))$  denote the multivariate normal measure with independent components and with mean vector  $\mu^*(b, \pi; \phi^*)$ . Let  $M$  be the space of continuous functions  $m : \mathbb{R}^{d_\beta} \times \Pi \rightarrow \mathbb{R}^{2d_\beta}$  endowed with the topology of pointwise convergence. Let  $P_n(b, \pi, \phi_n)$  denote the distribution of  $(\widehat{x}_1, \widehat{x}_2, \widehat{\mu}_n^*)$ , which is a function of the data  $W^n$ .

**RESULT 8:** *Suppose Assumptions Q1-Q7 hold. Fix  $\phi^* \in \Phi$  and let  $\mu^*(b, \pi; \phi^*)$  be the pointwise limit of  $\widehat{\mu}_n^*(b, \pi)$ . Let*

$$A \equiv \{(x_1, x_2, \mu) \in \mathbb{R}^{2d_\beta} \times M \mid z(x_1, x_2, \mu) - c(x_2, \alpha, \mu) > 0\}$$

*Let BDA denote the boundary of the set A. Suppose that for any  $(b, \pi)$ ,  $P(b, \pi; \mu^*(b, \pi; \phi^*))\{\text{BDA}\} = 0$ . Then, under any weak sequence  $(\beta_n, \pi_0, \phi_n)$  with limit  $(b, \pi_0, \phi^*)$ :*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^{2d_\beta} \times M} \phi_{\text{ECS}}(x_1, x_2; \mu) dP_n(b, \pi_0; \phi_n)(x_1, x_2, \mu) = \alpha$$

*Furthermore, under any sequence  $(\beta_n, \pi_0, \phi_n)$  with limit  $(b, \pi, \phi^*)$  and given weights  $P_1(b, \pi)$*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^{d_\beta} \times \Pi} \int_{\mathbb{R}^{2d_\beta} \times M} \phi_{\text{ECS}}(x_1, x_2; \mu) dP_n(b, \pi; \phi_n)(x_1, x_2, \mu) dP_1(b, \pi) = \text{WAP}(\phi_{\text{ECS}}(x; \mu^*))$$

**PROOF:** Note that by Assumption Q1-Q7 it follows that under any weak sequence  $(\beta_n, \pi_0, \phi_n)$  with limit  $(b, \pi_0, \phi^*)$ ,  $(\widehat{x}_1, \widehat{x}_2, \widehat{\mu}_n) \xrightarrow{d} (x_1, x_2, \widehat{\mu}^*(b, \pi_0, \phi^*))$ . Since  $P(b, \pi_0; \mu^*(b, \pi_0; \phi^*))\{\text{BDA}\} = 0$ , it follows—by the Portmanteu Theorem—that

$$\begin{aligned} \int_{\mathbb{R}^{2d_\beta} \times M} \phi_{\text{ECS}}(x_1, x_2; \mu) dP_n(b, \pi_0; \phi_n)(x_1, x_2, \mu) &= P_n(b, \pi_0, \phi_n)\{A\} \\ &\rightarrow P(b, \pi_0, \mu^*(b, \pi; \phi^*))\{A\} \\ &= \int_{\mathbb{R}^{2d_\beta} \times M} \phi_{\text{ECS}}(x_1, x_2; \mu) dP(b, \pi_0; \mu)(x_1, x_2, \mu^*(b, \pi_0, \phi^*)) \\ &= \alpha \end{aligned}$$

This establishes the asymptotic similarity of  $\phi_{\text{ECS}}(x_1, x_2; \widehat{\mu}_n^*)$  evaluated at sample analogues. For the WAP, note again that under any sequence  $(\beta_n, \pi, \phi)$  with limit  $(\beta_n, \pi, \phi^*)$  the Portmanteu Theorem implies

$$\begin{aligned}
\int_{\mathbb{R}^{2d\beta \times M}} \phi_{\text{ECS}}(x_1, x_2; \mu) dP_n(b, \pi; \phi^n)(x_1, x_2, \mu) &= P_n(b, \pi, \phi^n)\{A\} \\
&\rightarrow P(b, \pi, \mu^*(b, \pi; \phi^*))\{A\} \\
&= \int_{\mathbb{R}^{2d\beta \times M}} \phi_{\text{ECS}}(x_1, x_2; \mu) dP(b, \pi; \mu)(x_1, x_2, \mu^*(b, \pi, \phi^*))
\end{aligned}$$

pointwise in  $(b, \pi)$ . The dominated convergence theorem implies that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^{d\beta \times \Pi}} \int_{\mathbb{R}^{2d\beta \times M}} \phi_{\text{ECS}}(x_1, x_2; \mu) dP_n(b, \pi; \phi^n)(x_1, x_2, \mu) dP_1(b, \pi)$$

converges to

$$\int_{\mathbb{R}^{d\beta \times \Pi}} \int_{\mathbb{R}^{2d\beta \times M}} \phi_{\text{ECS}}(x_1, x_2; \mu) dP(b, \pi; \phi^*)(x_1, x_2, \mu^*(b, \pi; \phi^*)) dP_1(b, \pi)$$

which equals  $\text{WAP}(\phi_{\text{ECS}}(x; \mu^*))$ .

*Q.E.D.*

## APPENDIX G: LIMITING STATISTICAL MODEL FOR WEAKLY-IDENTIFIED GMM

In this section, I shall derive 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,  $\phi$ . For the sake of exposition, 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 weakly-identified GMM models. The concepts and main results in this section are illustrated using a weakly identified IV model with non-homoskedastic and/or serially correlated errors

**a) ECONOMETRIC MODEL:** Let  $x_{tT}$  be an  $\mathbb{R}^d$ -valued random variable. The econometrician observes the data set  $\{x_{tT}\}_{t=1}^T$ , the  $t$ th-row of a row-wise independent triangular array of random variables, whose unknown distribution depends on a scalar parameter of interest  $\theta \in \mathbb{R}$  and an additional nuisance parameter  $\phi$ . The distribution of  $x_{tT}$  is completely specified by  $\gamma = (\theta, \phi)$ .

There is a known  $\mathbb{R}^m$ -valued function  $h(x_t, \theta)$  that identifies the true parameter  $\theta^*$  through the following moment condition:

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

where  $\Phi$  is the set of possible values for the nuisance parameter  $\phi$ . In a slight abuse of notation let  $\gamma^*$  denote any element of the form  $(\theta^*, \phi)$ .

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

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

The testing problem of interest is

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

**EXAMPLE (GMM-IV):** Let  $x_t \equiv (y_{tT}, Y_{tT}, Z_{tT})$ , where  $y_{tT}$  is the outcome variable;  $Y_{tT}$  is a single endogenous regressor, and  $Z_{tT}$  is a vector of  $k \times 1$  instruments. Consider the function

$$h(y_{tT}, Y_{tT}, Z_{tT}, \theta) = Z_{tT}(y_{tT} - \theta Y_{tT}).$$

Note that

$$\mathbb{E}_{\gamma^*}[h(y_{tT}, Y_{tT}, Z_{tT}, \theta_0)] = (\theta^* - \theta_0)\mathbb{E}_{\gamma^*}[Z_{tT}Y_{tT}] = (\theta^* - \theta_0)\mathbb{E}_{\gamma^*}[Z_{tT}Z'_{tT}]\Pi$$

and

$$\mathbb{E}_{\gamma^*}[\dot{h}(x_{tT}, \theta)] = -\mathbb{E}_{\gamma^*}[Z_{tT}Z'_{tT}]\Pi.$$

**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  $\gamma^*$ ) weak convergence assumptions for the sample moment condition and its derivative:

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

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  $t$ , regardless of the value of the nuisance parameter  $\delta$ . Consider the following regularity conditions for  $C_t$  and its derivative  $\dot{C}_t(\theta^*, \theta, \delta) \equiv \partial C_t(\theta^*, \theta, \delta)/\partial \theta$ :

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

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

EXAMPLE (GMM-IV): Using the local-to-zero assumption of [Staiger and Stock \(1997\)](#),

$$\mathbb{E}_{\theta^*} [h(y_{tT}, Y_{tT}, Z_{tT}, \theta_0)] = (\theta^* - \theta_0) \mathbb{E}[Z_{tT} Z'_{tT}] \Pi = (\theta^* - \theta_0) \mathbb{E}[Z_{tT} Z'_{tT}] \delta / \sqrt{T}$$

and

$$\mathbb{E}_{\theta^*} [\dot{h}(y_{tT}, Y_{tT}, Z_{tT}, \theta_0)] = -\mathbb{E}[Z_{tT} Z'_{tT}] \Pi = -\mathbb{E}[Z_{tT} Z'_{tT}] \delta / \sqrt{T}.$$

Therefore,

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

and

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

which in the i.i.d. case does not depend on  $tt$ . Under standard regularity conditions for the second moments of  $Z_t$ ,

$$C(\theta^*, \theta_0, \delta) = (\theta^* - \theta_0) \delta \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[Z_{tT} Z'_{tT}] < \infty,$$

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

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

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

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

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

where

$$Q \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[Z_{tT} Z'_{tT}].$$

Since  $Q$  is assumed known and nonsingular in the limiting experiment it is possible to redefine  $\tilde{\delta}$  as  $Q\delta$ . In a slight abuse of notation  $\tilde{\delta}$  is relabeled as  $\delta$ . The specific form of the matrix  $\Omega(\theta_0, \phi)$  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

$$\Omega(\theta_0, \phi) = \begin{pmatrix} \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}_\phi \left[ Z_{tT} Z'_{tT} (y_{tT} - \theta_0 Y_{tT})^2 \right] & \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}_\phi \left[ Z_{tT} Z'_{tT} (y_{tT} - \theta_0 Y_{tT}) v_{2tT} \right] \\ \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}_\phi \left[ Z_{tT} Z'_{tT} (y_{tT} - \theta_0 Y_{tT}) v_{2tT} \right] & \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}_\phi \left[ Z_{tT} Z'_{tT} v_{2tT}^2 \right] \end{pmatrix}.$$

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 standardize the limiting experiment described above. Let  $[\Omega(\theta_0, \phi)]_m$  denote the upper left  $m \times m$  block of the matrix  $\Omega(\theta_0, \phi)$ ; that is, the asymptotic variance of the sample moment condition. In section ?? of the Appendix, I show there is a  $2m$  square matrix of the form:

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

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

$$\begin{aligned} \begin{pmatrix} m_T(\theta_0) \\ d_T(\theta_0) \end{pmatrix} &\equiv D(\theta_0, \phi) \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T h(x_{tT}, \theta_0) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \dot{h}(x_{tT}, \theta_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} m(\theta_0) \\ d(\theta_0) \end{pmatrix} \\ &\sim \mathcal{N}_{2m} \left( \begin{pmatrix} [\Omega(\theta_0, \phi)]_m^{-1/2} C(\theta^*, \theta_0, \delta) \\ d_1 C(\theta^*, \theta_0, \delta) + d_2 \dot{C}(\theta^*, \theta_0, \delta) \end{pmatrix}, \mathbb{I}_{2m} \right). \end{aligned}$$

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  $\theta^*$  and the nuisance vector  $\delta$ . 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) WEIGHTS 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  $\theta_0$ , consider the mapping  $C^* : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{2m}$  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



$C^* : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^{2k}$  is given by

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

The mapping  $C^*$  is continuous in  $(\theta^*, \delta)$ . 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 9:** *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:*

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

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

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

PROOF: 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  $T \equiv C^* : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{2m}$  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  $f : \mathbb{R}^{2m} \rightarrow \mathbb{R}$  by:

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

Let  $\mu$  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

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

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

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

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

*Q.E.D.*

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

$$\left( \frac{1}{T} \sum_{t=1}^T h(x_{tT}, \theta_0) \right)' [\Omega(\theta_0, \phi)]_m^{-1} \left( \frac{1}{T} \sum_{t=1}^T h(x_{tT}, \theta_0) \right) > \chi_{m, 1-\alpha}^2,$$

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

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

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

**EXAMPLE (GMM-IV):** The condition of Result 9 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:

$$\left( \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tT}(y_{tT} - \theta_0 Y_{tT}) \right)' \left( \frac{1}{T} \sum_{t=1}^T Z_{tT} Z_{tT}' (y_{tT} - \theta_0 Y_{tT})^2 \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tT}(y_{tT} - \theta_0 Y_{tT}) \right) > \chi_{1, 1-\alpha}^2,$$

which coincides with the robust version of the AR test derived in section 5.1. When  $k > 1$ , there are no weights over  $(\theta^*, \delta)$  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 10:** Let  $p_1$  be a full-support weight over  $\mathbb{R}^{n+1}$  and suppose that assumption RGMM1 holds. Then the  $\alpha$ -ECS test statistic for the problem  $\mathbf{H}_0 : \theta^* = \theta_0$  vs.  $\mathbf{H}_1 : \theta^* \neq \theta_0$  rejects the null hypothesis if

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

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

**PROOF:**

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

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

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

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

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

*Q.E.D.*

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

$$z(\widehat{m}(\theta_0), \widehat{d}(\theta_0)) = \int_{\mathbb{R}^{k+1}} \left[ \exp \left( \left( \begin{array}{c} \widehat{m}(\theta_0) \\ \widehat{d}(\theta_0) \end{array} \right)' \widehat{D}(\theta_0) \left( \begin{array}{c} (\theta^* - \theta_0)\delta \\ \delta \end{array} \right) \right) \right. \\ \left. \exp \left( -\frac{1}{2} \left( \begin{array}{c} (\theta^* - \theta_0)\delta \\ \delta \end{array} \right)' \widehat{\Omega}(\theta_0)^{-1} \left( \begin{array}{c} (\theta^* - \theta_0)\delta \\ \delta \end{array} \right) \right) p_1(\theta^*, \delta) \right] d\theta^* d\delta,$$

where

$$\widehat{D}(\theta_0) = \begin{pmatrix} \left[ \frac{1}{T} \sum_{t=1}^T Z_{tT} Z'_{tT} (y_{tT} - \theta_0 Y_{tT})^2 \right]^{-1/2} & \mathbf{0} \\ \widehat{d}_1 & \widehat{d}_2 \end{pmatrix}, \quad \widehat{\Omega}(\theta_0)^{-1} = \widehat{D}(\theta_0)' \widehat{D}(\theta_0),$$

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:

$$\widehat{d}(\theta_0) = \widehat{d}_1 \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tT} (y_{tT} - \theta_0 Y_{tT}) + \widehat{d}_2 \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tT} Y_{tT},$$

and

$$\widehat{m}(\theta_0) = \left( \frac{1}{T} \sum_{t=1}^T Z_{tT} Z'_{tT} (y_{tT} - \theta_0 Y_{tT})^2 \right)^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tT} (y_{tT} - \theta_0 Y_{tT}).$$

The critical value  $c(\widehat{d}(\theta_0))$  is obtained by fixing  $\widehat{d}(\theta_0)$  and computing the  $1-\alpha$  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  $z(\widehat{m}(\theta_0), \widehat{d}(\theta_0)) > c(\widehat{d}(\theta_0))$ .

## APPENDIX H: WLLN AND CLT FOR ROW-WISE INDEPENDENT TRIANGULAR ARRAYS

The following well-known theorems are the main tools used to verify high-level assumptions for the two running examples. This section presents the proof of a simple corollary for each theorem.

**THEOREM 2:** (*Durrett (2010), pg. 59; Weak Law of Large Numbers for row-wise independent triangular arrays*) Let  $\{X_{in}\}_{i=1}^n$ ,  $n \in \mathbb{N}$  be a row-wise independent triangular array of real-valued random variables defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ . Suppose that as  $n \rightarrow \infty$

1.  $\sum_{i=1}^n \mathbb{P}(|X_{in}| > n) \rightarrow 0$ , and
2.  $n^{-2} \sum_{i=1}^n \mathbb{E}_{\mathbb{P}}[X_{in}^2 \mathbf{1}_{|X_{in}| < n}] \rightarrow 0$

then

$$\frac{1}{n} \sum_{i=1}^n X_{in} - \mathbb{E}_{\mathbb{P}}[X_{in} \mathbf{1}_{|X_{in}| \leq n}] \xrightarrow{P} 0$$

The following Corollary simplifies the application of Theorem 1 by providing a simple sufficient condition for 1) and 2) above. The sufficient condition also guarantees that  $\mathbb{E}_{\mathbb{P}}[X_{in} \mathbf{1}_{|X_{in}| \leq n}]$  can be replaced by  $\mathbb{E}_{\mathbb{P}}[X]$ , where  $X_{in} \xrightarrow{d} X$ .

**COROLLARY 1:** Let  $\{X_{in}\}_{i=1}^n$ ,  $n \in \mathbb{N}$  be a row-wise i.i.d. triangular array of real-valued random variables defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  and suppose that for arbitrary  $i \in \mathbb{N}$ , the sequence  $\{X_{in}\}_{n=i}^{\infty} \xrightarrow{d} X$ . If

$$S = \sup_{n \geq 1} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{1+\delta}] < \infty, \quad \text{for some } \delta > 0$$

then

$$\frac{1}{n} \sum_{i=1}^n X_{in} \xrightarrow{P} \mathbb{E}_{\mathbb{P}}[X]$$

**PROOF:** Let  $F_n$  be the c.d.f of  $X_{in}$ . Since  $\{X_{in}\}_{i=1}^n$ ,  $n \in \mathbb{N}$  is a row-wise i.i.d. triangular array of real-valued random variables, then

$$\sum_{i=1}^n \mathbb{P}(|X_{in}| > n) = n \mathbb{P}(|X_{in}| > n) = \int_{\mathbb{R}} n \mathbf{1}_{|x| > n} dF_n(x) \leq \int_{\mathbb{R}} |x| \mathbf{1}_{|x| > n} dF_n(x) = \mathbb{E}_{\mathbb{P}}[|X_{in}| \mathbf{1}_{|X_{in}| > n}].$$

Note that for every  $n \in \mathbb{N}$

$$\mathbb{E}_{\mathbb{P}}[|X_{in}| \mathbf{1}_{|X_{in}| > n}] \leq \frac{1}{n^\delta} \mathbb{E}_{\mathbb{P}}[|X_{in}|^\delta \mathbf{1}_{|X_{in}| > n}] \leq \frac{1}{n^\delta} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{1+\delta}] \leq \frac{1}{n^\delta} S.$$

Hence, Assumption 1 of Theorem 2 is verified.

To verify Assumption 2, we consider the following two cases:  $\delta \geq 1$  and  $0 < \delta < 1$ . If  $\delta \geq 1$ , Assumption 2 follows immediately. Suppose now  $0 < \delta < 1$ . Again, using the row-wise i.i.d. assumption and the fact that

for any nonnegative random variable  $Y$ ,  $E(Y^p) = \int_0^\infty pY^{p-1}P(Y > y)dy$ :

$$\begin{aligned}
\frac{1}{n^2} \sum_{i=1}^n \mathbb{E}_{\mathbb{P}}[|X_{in}|^2 \mathbf{1}_{|X_{in}| < n}] &= \frac{1}{n} \mathbb{E}_{\mathbb{P}}[|X_{in} \mathbf{1}_{|X_{in}| < n}|^2] = \frac{1}{n} \int_0^\infty 2|x \mathbf{1}_{|x| < n}| \mathbb{P}(|X_{in} \mathbf{1}_{|X_{in}| < n}| > x) dx \\
&= \frac{1}{n} \int_0^n 2|x| \mathbb{P}(x < |X_{in}| < n) dx \leq \frac{1}{n} \int_0^n 2|x| \mathbb{P}(|X_{in}| > x) dx \\
&\leq \frac{2}{n} \int_0^n \mathbb{E}_{\mathbb{P}}[|X_{in}| \mathbf{1}_{|X_{in}| > x}] dx \leq \frac{2}{n} \int_0^n \frac{1}{x^\delta} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{1+\delta} \mathbf{1}_{|X_{in}| > x}] dx \\
&= \frac{2}{n} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{1+\delta}] \int_0^n \frac{1}{x^\delta} dx \\
&= \frac{2}{n} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{1+\delta}] \left[ \frac{x^{1-\delta}}{1-\delta} \right]_0^n \quad \text{integrability follows from } 0 < \delta < 1 \\
&= \frac{2}{n} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{1+\delta}] \left[ \frac{n^{1-\delta}}{1-\delta} \right] \\
&= \frac{2}{(1-\delta)n^\delta} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{1+\delta}] \\
&\leq \frac{2}{(1-\delta)n^\delta} S.
\end{aligned}$$

Hence, Assumption 2 of Theorem 2 is verified. Therefore,

$$\frac{1}{n} \sum_{i=1}^n X_{in} - \mathbb{E}_{\mathbb{P}}[X_{in} \mathbf{1}_{|X_{in}| \leq n}] \xrightarrow{P} 0.$$

Finally, since  $\{X_{in}\}_{n=i}^\infty \xrightarrow{d} X$  it follows that  $\{X_{in} \mathbf{1}_{|X_{in}| \leq n}\}_{i=n}^\infty \xrightarrow{d} X$ . Since  $S < \infty$ , the sequence  $\{X_{in} \mathbf{1}_{|X_{in}| \leq n}\}_{i=n}^\infty$  is uniformly integrable as defined in pg. 31 [Billingsley \(1995\)](#). Therefore,

$$\mathbb{E}_{\mathbb{P}}[X_{in} \mathbf{1}_{|X_{in}| \leq n}] \rightarrow \mathbb{E}_{\mathbb{P}}[X]$$

*Q.E.D.*

**THEOREM 3:** (*Durrett (2010), pg. 129; Lindeberg-Feller Central Limit Theorem for row-wise independent triangular arrays*) Let  $\{X_{in}\}_{i=1}^n$ ,  $n \in \mathbb{N}$  be a row-wise independent triangular array of zero-mean real-valued random variables defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $\mathbb{E}_{\mathbb{P}}[X_{in}] = 0$ . Suppose that as  $n \rightarrow \infty$

1.  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathbb{P}}[X_{in}^2] \rightarrow \sigma^2 > 0$
2.  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathbb{P}}[|X_{in}|^2 \mathbf{1}_{|X_{in}| > \sqrt{nm}}] \rightarrow 0$ , for all  $m > 0$ .

Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{in} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

**COROLLARY 2:** Let  $\{X_{in}\}_{i=1}^n$ ,  $n \in \mathbb{N}$  be a row-wise i.i.d. triangular array of real-valued random variables defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  and suppose that for arbitrary  $i \in \mathbb{N}$ , the sequence  $\{X_{in}\}_{n=i}^\infty \xrightarrow{d} X$ . If

$$S = \sup_{n \geq i} \mathbb{E}_{\mathbb{P}}[|X_{in}|^{2+\delta}] < \infty, \quad \text{for some } \delta > 0$$

then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{in} \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_{\mathbb{P}}[X^2])$$

PROOF: Since the triangular array is row-wise i.i.d. then  $n^{-1} \sum_{i=1}^n \mathbb{E}_{\mathbb{P}}[X_{in}^2] = \mathbb{E}_{\mathbb{P}}[X_{in}^2]$ , for an arbitrary  $i$ . Furthermore,  $\{X_{in}^2\}_{n=i}^\infty \xrightarrow{d} X$  and the sequence is—by assumption—uniformly integrable. Therefore,

Theorem 3.5 in Billingsley (1995) implies:

$$\mathbb{E}_{\mathbb{P}}[X_{in}^2] \rightarrow \mathbb{E}_{\mathbb{P}}[X^2] \equiv \sigma^2,$$

which verifies Assumption 1 of Theorem 3. Now, note that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathbb{P}}[|X_{in}|^2 \mathbf{1}_{|X_{in}| > m}] &= \mathbb{E}_{\mathbb{P}}[X_{in}^2 \mathbf{1}_{|X_{in}| > \sqrt{nm}}] \\ &\leq \frac{1}{m^\delta \sqrt{n}^\delta} \mathbb{E}_{\mathbb{P}}[X_{in}^2 |X|^\delta \mathbf{1}_{|X_{in}| > \sqrt{nm}}] \\ &\leq \frac{1}{m^\delta \sqrt{n}^\delta} \mathbb{E}_{\mathbb{P}}[|X_{in}| 2 + \delta] \\ &\leq \frac{1}{m^\delta \sqrt{n}^\delta} S. \end{aligned}$$

Therefore,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{in}^2 \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_{\mathbb{P}}[X^2])$$

*Q.E.D.*

REMARK 18: The following fact is used extensively throughout the appendix. The distribution of the  $i$ -th component of the  $n$ -th row of the triangular array is fully parameterized by  $\gamma_n \equiv (\beta_n, \pi_n, \phi_n)$ . Let  $F_{\gamma_n}$  denote the distribution of  $W_{in}$  when the true parameter is  $\gamma_n$ . Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be the probability space over which the triangular array is defined. By the change of variable formula in Billingsley (1999) pg. 216, for any measurable function  $g$ :

$$\mathbb{E}_{\mathbb{P}}[g(W_{in})] = \int_{\Omega} g(W_{in}(\omega)) d\mathbb{P}(\omega) = \int_W g(w) dF_{\gamma_n}(w) \equiv \mathbb{E}_{\gamma_n}[g(W)].$$

In light of this result, the index  $\mathbb{P}$  and  $\gamma$  (or  $\gamma_n$ ) will be used interchangeably whenever appropriate.