Testability and methods of moments in nonparametric and semiparametric models

Jean-Marie Dufour* Frédéric Jouneau-Sion† Olivier Torrès‡

Abstract

Bahadur and Savage (1956) (hereafter BS) showed that any valid test procedure about the mean of a sequence of i.i.d. random variables has power not larger than its size when the distribution is not specified. We refer to this situation as *non-testabilty*. We propose extensions of BS's result to multivariate models with dependence and/or heterogeneity and where the parameter of interest does not necessarily coincide with the expectation. We show for instance that the covariance or the linear correlation matrix parameters are *non-testable parameters*. When non-testability obtains, asymptotic consistent test procedures necessarily have level 1 in the limit, regardless of the claimed level. Moreover, the convergence of the type 1 risk to the nominal level α is arbitrarily slow. These properties are true even when "corrections" for size distortion (*e.g.*,bootstrapping) are made. Non-testability problems are then investigated in econometric models, where a distinction between exogenous and endogenous variables is made, and where the parameters of interest are defined by moment conditions. We propose a valid and somewhere powerful test for the slope coefficients of a linear regression model when the error terms are identically distributed. However, we also show that the slope parameters are typically non-testable when the assumption of identically distributed errors is relaxed (for instance if the error terms are assumed independent and homoskedastic). We show how to extend these results to nonlinear regression and nonparametric regression models.

Key words: exact inference, semiparametric regression models, non-testability

Journal of Economic Literature classification: C10, C12, C14.

*William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com

†GREMARS, Université Lille 3 and C.O.R.E., Université Catholique de Louvain. Mailing address: GREMARS, Université Charles-de-Gaulle Lille 3, BP 149, 59653 Villeneuve d'Ascq cedex, France. TEL: (33) 3 20 41 68 34; e-mail: frederic.jouneau@univ-lille3.fr.

‡GREMARS, Université Lille 3. Mailing address: GREMARS, Université Charles-de-Gaulle Lille 3, BP 149, 59653 Villeneuve d'Ascq cedex, France . TEL: (33) 3 20 41 67 35; e-mail: olivier.torres@univ-lille3.fr.

1 Introduction

Providing inference procedures that are robust to various specifications of the DGP (data generating process) may undoubtedly be regarded as a major goal in econometrics. No less than five chapters of the two last issues of the *Handbook of Econometrics* deal with this issue [see Aït-Sahalia et al. (2001), Powell (1994), Andrews (1994), Arellano and Honoré (2001), Matzkin (1994)]. Procedures have been proposed to handle semiparametric specifications for limited dependent endogenous variables, time series and panel models. The rapidly growing development of computing capacities has undoubtedly played a large role in these achievements.

Nevertheless, several recent papers cast doubts on the validity of some of the techniques proposed: see for instance Dufour (1997), Gleser and Hwang (1987). More recently, Pötscher (2002) pointed at a class of "ill-posed" problems where reliable and meaningful inference procedures do not exist.

A pioneering result in this context is Bahadur and Savage (1956) (hereafter BS). In a semiparametric setting, BS proves that any valid test shows power not larger than its size, when the parameter of interest is an expectation. Moreover in this model, a.s. bounded confidence regions have necessarily zero coverage probability and the loss of point estimators cannot be controlled.

A typical result obtained by Bahadur and Savage (1956) is the following. Assume $X_1, \ldots, X_n$ are i.i.d. random variables whose common expectation is either 0 or 1, and whose all higher order moments exist. Then the power of any α-level test of $H_0 : \mathsf{E}(X_1) = 0$ against $H_1 : \mathsf{E}(X_1) = 1$ is necessarily less than α (see example 2 below). Moreover, if C is a confidence set for $\mathsf{E}(X_1)$ such that $\mathsf{P}_j(k \notin C) = 1$ when $j \neq k$, where P_j is a probability for X_1 with expectation j, $j, k = 0, 1$, then the confidence level of C cannot be $1 - \alpha$, for all $\alpha \in]0, 1[$. In other words, C coincide with the parameter space with probability 1.

Although the results by BS are not restricted to testing problems, it has been suggested to refer to this situation as "non-testability" [Dufour (2003)]. Tibshirani and Wasserman (1988) generalized the results of Bahadur and Savage. More recently Romano (2004) examined the size and power properties of the t-test in a semiparametric setting.

Despite their theoretical importance, BS's results —as well as those obtained by Tibshirani and Wasserman (1988) and Romano (2004) — deal with finite random samples (i.i.d. observations). Now, most semiparametric inference methods are developed in the non-i.i.d. case and are justified on asymptotic grounds. As such, the results obtained by BS and their followers do not directly apply in this context. Therefore, it could be claimed that using covariates and/or relying on asymptotic approximations could be a way out of BS's non-testability results.

This paper investigates non-testability issues in the context of statistical models more relevant to the econometrician.

First, we provide a formal definition of (non)-testability of a testing problem. This definition is naturally extended to a parameter. Roughly speaking, a testing problem is non-testable is the power of any test cannot exceed its level. A parameter θ is thus said non-testable whenever any hypothesis of the form $\theta = \theta_0$ is non-testable against the alternative $\theta \neq \theta_0$. We stress that non-testability of a property of a testing problem and not that of a particular test or inference method. We therefore study how this property is affected by various transformations of the testing problem (enlargement of the null and/or alternative hypothesis, parameter mappings, etc). In particular, we show that any transformation of a non-testable parameter is non-testable.

Second, as an extension of BS's initial result, we provide sufficient conditions for non-testability in more general settings. In particular, our result applies to possibly multidimensional parameters, not necessarily related to the expectation of a distribution, and dependence and/or heterogeneity are allowed. Interestingly, we relate this to results obtained by Pötscher (2002) and show that when non-testability holds, there is no small enough neighbourhood in the alternative hypothesis over which the type II risk can be bounded. Based on this result, we provide examples of non testable parameters. Specifically, we show that no meaningful test procedure exists for the covariance or the linear correlation matrix parameters. The variance and the bounds of the support of a random variable are partially non-testable parameters, a weaker form of non-testability.

Next we consider a possible departure from the original context of BS's result by letting the sample size be arbitrarily large. We investigate the asymptotic properties of test procedures for non-testable problems. We show that non-testability (either partial or not) implies that any asymptotic consistent test procedure necessarily has level 1 in the limit, regardless of its claimed level. This holds even when "corrections" (for instance bootstrapping) are made in order to reduce the discrepancies between the actual and the claimed level. As a consequence, we prove that the convergence towards α of the null rejection probabilities of an asymptotic consistent test is arbitrarily slow when the problem under test is non-testable.

Finally, we examine testability issues that are more specific to econometric models, where the parameters of interest are defined in terms of moment conditions. We first consider a basic semiparametric linear regression model. Surprisingly, although the expectation of the dependent variable is non-testable (due to BS's original result), a valid α-level and somewhere powerful test may be provided for the slope parameters when the errors terms are assumed identically distributed. Thus non-testability concerns the intercept parameter only. This shows that using covariates may indeed be a way out of BS' impossibility results. However, when the assumption of identically distributed error terms is relaxed, testability of the slopes is typically lost. In particular, the parameters of a linear regression model with independent and homoskedastic error terms are non-testable. We next show how to extend this problem to nonlinear or nonparametric regression models. In this last case, neither the regression function nor the hypothesis of significance of a given regressor are testable.

The paper is organized as follows. In section 2, we provide the formal definitions of the concepts of testability and non-testability, and we derive basic properties related to these concepts that are free of a particular specification of the underlying statistical model. Section 3 presents some extensions of the main result of Bahadur and Savage (1956). Section 4 is devoted to the study of the asymptotic behavior of tests in presence of non testability. In section 5, we provide various examples of non-testability. We investigate testability of the parameters of linear regression models in section 6 and extensions to more general models are considered in section 7. Section 8 concludes. All proofs are given in the appendix section.

2 Testability and non-testability

Before investigating issues related to the results presented in this paper, we provide the definitions of size, level and power of a test used in the paper. These definitions are those of Lehmann (1986). Let $\mathscr{P}$ be a family of probability distributions for a n-tuple $(X_1, \ldots, X_n)$ of random vectors whose distribution is determined by a probability measure $\mathsf{P} \in \mathscr{P}$. We consider the problem of testing

$$H_0 : \mathsf{P} \in \mathscr{P}_0 \qquad \text{against} \qquad H_1 : \mathsf{P} \in \mathscr{P}_1 \,, \tag{2.1}$$

where $\mathscr{P}_0$ and $\mathscr{P}_1$ are two subsets of $\mathscr{P}$. We denote this problem by the ordered pair (H_0, H_1). A test φ for (H_0, H_1) is a (possibly random) function of $(X_1, \ldots, X_n)$ to $\{0, 1\}$. We shall write φ_n for $\varphi(X_1, \ldots, X_n)$. The equality $\varphi = 1$ is interpreted as the action of "rejecting" H_0 (against H_1) and $\varphi = 0$ is the action of "accepting" H_0.

The *size* of φ is the number $\sup_{\mathsf{P}\in\mathscr{P}_0} \mathsf{P}(\varphi_n = 1)$. The test φ has *level* α whenever its size is less than or equal to α. Moreover, φ is *similar* at level α if $\mathsf{P}(\varphi_n = 1) = \alpha$, $\forall \mathsf{P} \in \mathscr{P}_0$. The *power* of φ is the function $\mathsf{P} \in \mathscr{P}_1 \longmapsto \mathsf{P}(\varphi_n = 1)$. Finally, a test φ is *biased* at level α, if (1) φ has level α, and (2) for some $\mathsf{P} \in \mathscr{P}_1$ the power of φ is strictly less than α. Note also that $\mathsf{P}(\varphi_n = 1) = \mathsf{E}_{\mathsf{P}}(\varphi_n)$, where E_{P} is the expectation associated with P.

Whatever $\alpha \in [0, 1]$, it is always possible to obtain a test with size α. This test is denoted $\varphi^*(\alpha)$ and is defined by

$$\varphi^*(\alpha) = \mathrm{I}(U_{[0,1]} \leq \alpha), \tag{2.2}$$

where $U_{[0,1]}$ is uniform on $[0, 1]$ and $\mathrm{I}(A)$ is the indicator function of the event A. It is therefore clear that the main issue in any testing problem is whether a test better than $\varphi^*(\alpha)$ exists. If not, $\varphi^*(\alpha)$ is UMP at level α. In other words, it is optimal not to use the data for deciding H_0 or H_1. The hypotheses under test are then called *non-testable*, in the sense that no data can provide information that helps discriminating between the null and the alternative. We provide a formal definition of this concept.

2.1 Definitions

Definition 2.1 (TESTABILITY AND NON-TESTABILITY). *Let* (H_0, H_1) *be the testing problem in* (2.1) *and let* $\alpha \in [0, 1]$.

(1) H_0 *is* testable at level α *against* H_1 *iff there exists a test* φ *for* (H_0, H_1) *such that:* (i) $\mathsf{P}(\varphi = 1) \leq \alpha$ *for all* $\mathsf{P} \in \mathscr{P}_0$, *and* (ii) $\mathsf{P}(\varphi = 1) > \alpha$ *for at least one* $\mathsf{P} \in \mathscr{P}_1$.

(2) H_0 *is* non-testable at level α *against* H_1 *iff* H_0 *is not testable at level* α *against* H_1.

(3) H_0 *is* testable *against* H_1 *iff* H_0 *is testable at level* α *against* H_1 *for some* $\alpha \in [0, 1]$.

(4) H_0 *is* non-testable *against* H_1 *iff* H_0 *is non-testable at level* α *against* H_1 *for all* $\alpha \in [0, 1]$.

In parts (1) and (2) of the above definition, we formally allow $\alpha = 1$; but for $\alpha = 1$, it is clear that H_0 cannot be *testable at level* α against any hypothesis H_1 and the property that H_0 *is non-testable at level* 1 against H_1 necessarily holds irrespective of the pair

(H_0, H_1) considered. From the above definition, it is also easy to see that H_0 is *non-testable at level* α against H_1 iff

$$\mathsf{P}(\varphi_n = 1) \leq \alpha, \forall \mathsf{P} \in \mathscr{P}_0 \Rightarrow \mathsf{P}(\varphi_n = 1) \leq \alpha, \forall \mathsf{P} \in \mathscr{P}_1, \tag{2.3}$$

i.e., if the power of any level-α test of H_0 against H_1 cannot exceed the level α. In such cases, a test of the form $\varphi^*(\alpha)$ in (2.2) which does not depend on the data is UMP at level α. Similarly, H_0 is *non-testable* against H_1 iff the power of any test does not exceed its level, irrespective of the level $\alpha \in [0, 1]$. When H_0 is non-testable (at level α) against H_1, we also say that the problem (H_0, H_1) is *non-testable* (*at level* α).

In certain situations, it is useful to consider a weaker notion of non-testability, we call partial non-testability.

Definition 2.2 (PARTIAL NON-TESTABILITY). *Let* (H_0, H_1) *be the testing problem in* (2.1) *and let* $\alpha \in [0, 1]$.

(1) H_0 *is* partially non-testable at level α *against* H_1 *iff for any* α*-level test* φ *of* H_0*, there exists a non empty set* $\hat{\mathscr{P}}_1 \subset \mathscr{P}_1$ *such that* $\mathsf{P}(\varphi = 1) \leq \alpha, \forall \mathsf{P} \in \hat{\mathscr{P}}_1$.

(2) H_0 *is* partially non-testable *against* H_1 *iff* H_0 *is partially non-testable at level* α *against* H_1 *for all* $\alpha \in [0, 1]$.

When H_0 is partially non-testable (at level α), against H_1, we also say that *the problem* $H = (H_0, H_1)$ *is partially non-testable (at level* α*).* Partial non-testability of H simply means that the power of *any* test of H_0 against H_1 collapses for some alternatives. The subset $\hat{\mathscr{P}}_1$ over which the power of a test does not exceed its level may depend on the level and on the test used. Obviously, non-testability implies partial non-testability with (for instance) $\hat{\mathscr{P}}_1 = \mathscr{P}_1$ for all α and all test φ.

Non-testability (or partial non-testability) of a given problem directly follows from the specification of the statistical model $\mathscr{P}$ in which it is formulated. Non-testability (or partial non-testability) is then a property of the model itself.

Testability and (partial) non-testability may be extended to a parameter in a natural way. Let $\theta : \mathscr{P} \to \Theta$ be a parameter defining mapping on $\mathscr{P}$ —which defines the parameter θ— and $\overline{\theta} \in \Theta$ denote the image of the true DGP $\overline{\mathsf{P}}$: $\overline{\theta} \equiv \theta(\overline{\mathsf{P}})$. When introducing a testing problem about the parameter θ, we will denote a null hypothesis as $H_0 : \overline{\theta} \in \Theta_0$, which is a shortcut for

$$H_0 : \overline{\mathsf{P}} \in \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) \in \Theta_0\}. \tag{2.4}$$

In particular, let us write:

$$H_0(\theta_0) : \overline{\mathsf{P}} \in \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) = \theta_0\}, \qquad H_1(\theta_0) : \overline{\mathsf{P}} \in \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) \neq \theta_0\}. \quad (2.5)$$

In statistical applications, the mapping θ defines the parameter of interest. Notice that fully nonparametric models are allowed by choosing the mapping θ as the identity mapping on $\mathscr{P}$, *i.e.*, $\theta(\mathsf{P}) = \mathsf{P}$.

Definition 2.3 (TESTABLE PARAMETER). *Consider a family of testing problems indexed by $\theta_0 \in \Theta$ as defined in (2.5) and let Θ^* be a non empty subset of Θ.*

(1) *The parameter θ is testable at level α on Θ^* iff $H_0(\theta_0)$ is testable at level α against $H_1(\theta_0)$ for any $\theta_0 \in \Theta^*$.*

(2) *θ is testable on Θ^* iff $H_0(\theta_0)$ is testable against $H_1(\theta_0)$ for any $\theta_0 \in \Theta^*$.*

(3) *The parameter θ is (partially) non-testable on Θ^* iff $H_0(\theta_0)$ is (partially) non-testable against $H_1(\theta_0)$ for any $\theta_0 \in \Theta^*$.*

(4) *The parameter θ is (partially) non-testable iff it is (partially) non-testable on $\theta(\mathscr{P})$.*

Since θ is a *mapping*, the parameter is necessarily identified in the usual sense where: $\mathscr{P}(\theta_1) \cap \mathscr{P}(\theta_2) \neq \varnothing \Rightarrow \theta_1 = \theta_2$. To account for non-identifiability, we need consider θ as a *correspondence* $\mathscr{P} \to \mathcal{P}(\Theta)$, where $\mathcal{P}(\Theta)$ is the set of all subsets of Θ. Now assume that we face a non-identified parametric problem *i.e*, there exists two distinct values θ_1, θ_2 in Θ such that $\mathscr{P}(\theta_1) \cap \mathscr{P}(\theta_2) \neq \varnothing$. Consider $H_0 : \overline{\theta} \subset \Theta_0$ and $H_1 : \overline{\theta} \not\subset \Theta_0$. When $\theta_1 \in \Theta_0$ and $\theta_2 \notin \Theta_0$ H_0 is partially non-testable against H_1. Unless explicitly mentioned, we address testability issues that arise when the parameter is identified and θ is therefore a mapping.

2.2 Testability and confidence regions

It is well known that confidence regions and testing problems are closely related. We have the following property.

Proposition 2.4 (NON-INFORMATIVE PROPERTY OF CONFIDENCE REGIONS ON NON-TESTABLE PARAMETERS).*Let C be a confidence region with level $1-\alpha$ for the parameter θ taking values in Θ. If θ is non-testable at level α on Θ^*, then $\mathsf{P}(\theta_0 \in C) \geq 1-\alpha$ for all admissible value of $\theta_0 \in \Theta^*$ and all $\mathsf{P} \in \mathscr{P}$.*

This shows that non-testability entails that valid confidence regions are totally uncontrollable. In particular, if some metric on Θ is available, we obtain results similar to Dufour (1997). For instance, if Θ is an unbounded set and the confidence set C is almost surely bounded, confidence regions on non-testable parameters have zero coverage probability.

2.3 Invariance properties

We examine how testability of a testing problem is preserved when the problem is transformed through modifications of the null hypothesis or the alternative hypothesis.

2.3.1 Modifications of the testing problem

Let $\mathscr{P}$ be a statistical model and let $\tilde{\mathscr{P}}_0$ and $\tilde{\mathscr{P}}_1$ be defined as $\tilde{\mathscr{P}}_k = \mathscr{P}_k \cap \tilde{\mathscr{P}} \neq \varnothing$, $k = 0, 1$ for some subset $\tilde{\mathscr{P}}$ of $\mathscr{P}$ and $\mathscr{P}_0 \cup \mathscr{P}_1 = \mathscr{P}$. Clearly, $\tilde{\mathscr{P}}_k \subseteq \mathscr{P}_k$, $k = 0, 1$, so that $\tilde{\mathscr{P}}_k$ is more restrictive than $\mathscr{P}_k$. Let us also define

$$H_k : \overline{\mathsf{P}} \in \mathscr{P}_k, \qquad \tilde{H}_k : \overline{\mathsf{P}} \in \tilde{\mathscr{P}}_k, \tag{2.6}$$

$k = 0, 1.$

Proposition 2.5 (MONOTONICITY OF PARTIAL NON-TESTABILITY). *If $(\tilde{H}_0, \tilde{H}_1)$ is partially non-testable then (H_0, H_1) is partially non-testable.*

One may think that if a problem is non-testable, it remains non-testable in a larger model. Proposition 2.5 shows this is true for partially non-testable problems. However it may not hold for non-testable ones as shown by the following example.

EXAMPLE 1. Let $\tilde{\mathscr{P}}$ be the set of all continuous distributions on $\mathbb{R}$ with finite expectation. The null hypothesis is $\tilde{H}_0 : \mathsf{E}_{\overline{\mathsf{P}}}(X) = 0$ and $\tilde{H}_1 : \mathsf{E}_{\overline{\mathsf{P}}}(X) \neq 0$. According to Bahadur and Savage (1956), the problem $\tilde{H} = (\tilde{H}_0, \tilde{H}_1)$ is non-testable. Now let $\mathscr{P} = \tilde{\mathscr{P}} \cup \{\delta_{\{1\}}\}$ where $\delta_{\{1\}}$ is the Dirac mass at point 1. Let $H_0 : \mathsf{E}_{\overline{\mathsf{P}}}(X) = 0$ and $H_1 : \mathsf{E}_{\overline{\mathsf{P}}}(X) \neq 0$. Let $0 < \alpha < 1$ and consider the following test procedure "if the first observation of the sample is 1 reject the null otherwise draw U in the uniform distribution on $[0, 1]$ and reject the null whenever U is smaller than α". Under the null, the distribution is continuous, hence the event $X_1 = 1$ has probability zero. Thus the procedure has level α. Now under the alternative $\delta_{\{1\}}$ the rejection probability is $1 > \alpha$. $\square$

As for non-testability, we have the following result.

Proposition 2.6 (MONOTONICITY OF NON-TESTABILITY).*For each* $\alpha \in]0, 1[$ *if* $(\tilde{H}_0, H_1)$ *is non-testable at level* α*, then* $(H_0, \tilde{H}_1)$ *is non-testable at level* α*.*

The last result of this section shows that when enlarging a testable problem, a somewhere powerful α-level test of the "small" problem cannot be valid for the "large" one, if this large problem is non-testable.

Proposition 2.7 *Suppose* $\tilde{H}_0$ *is testable at level* α *against* $\tilde{H}_1$*, while* H_0 *is not testable ate level* α *against* H_1*. If* φ *is not dominated by* $\varphi^*(\alpha)$ *for testing* $(\tilde{H}_0, \tilde{H}_1)$ *then* φ *is not a test with level* α *for testing* (H_0, H_1)*.*

2.3.2 Parameter transformations

We consider the problem of testability of functions of a parameter.

Proposition 2.8 (TRANSFORMATIONS OF NON-TESTABLE PARAMETERS). *Let* $\mathscr{P}$ *be a statistical model and* $\theta : \mathscr{P} \to \Theta$ *be a parameter s.t.* $\theta(\mathsf{P})$ *is defined* $\forall\, \mathsf{P} \in \mathscr{P}$*. Let* $g : \Theta \to \Lambda$ *be any mapping and define* $\lambda : \mathscr{P} \to \Lambda$ *by* $\lambda = g \circ \theta$*. If* θ *is non-testable then* λ *is non-testable.*

However, if θ is partially non-testable only, some transformations of this parameter may be testable (see Proposition 6.2 below for an example).

As a related result, we now consider the case $\Theta = \Gamma \times \Psi$ so that any $\theta \in \Theta$ can be decomposed as $\theta = (\gamma, \psi) \in \Gamma \times \Psi$. We may define the sub-parameter mappings $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ as the projections of θ on Γ and Ψ, respectively. The true values of θ, γ and ψ are $\overline{\theta} = \theta(\overline{\mathsf{P}}) = (\overline{\gamma}, \overline{\psi})$, with $\overline{\gamma} \equiv \boldsymbol{\gamma}(\overline{\mathsf{P}})$ and $\overline{\psi} \equiv \boldsymbol{\psi}(\overline{\mathsf{P}})$.

Proposition 2.9 (MARGINILIZATION OF NON-TESTABLE PARAMETERS). *If* θ *is non-testable, then* $\boldsymbol{\gamma}$ *is non-testable.*

Proposition 2.10 (SUBPARAMETERS OF NON-TESTABLE PARAMETERS). *If* $\forall \gamma_0 \in \Gamma$, $\forall \psi_0 \in \Psi$, $H_0^{\psi_0}(\gamma_0) : \{\overline{\gamma} = \gamma_0, \overline{\psi} = \psi_0\}$ *is non-testable against* $H_1^{\psi_0}(\gamma_0) : \{\overline{\gamma} \neq \gamma_0, \overline{\psi} = \psi_0\}$*, then*

(1) $\boldsymbol{\gamma}$ is non-testable;

(2) θ is partially non-testable.

3 Bahadur-Savage result and its extensions

As we already mentioned Bahadur and Savage (1956) provide an example of a non-testable parameter. Bahadur and Savage (1956) consider a n-tuple of i.i.d. real random variables. They establish in this setting a general result which implies the non-testability of the expectation. The purpose of this section is to extend BS's result to models more relevant to econometrics, by considering models where the i.i.d. assumption is not required: dependence and/or heterogeneity is allowed. Moreover, probability distributions are possibly multivariate and parameters are not necessarily defined as an expectation.

3.1 A generalized BS-type theorem

The main result is given by Theorem 3.1 below. It formally proves BS's comments on possible extensions of their Theorem 1 (and the accompanying corollaries) to parameters other than the expectation.

Theorem 3.1 (GENERALIZED BAHADUR-SAVAGE THEOREM).*For a real random vector X of $\mathbb{R}^d$, consider a family of probability distributions $\mathscr{P}$. For some set Θ, let $\theta : \mathscr{P} \to \Theta$ be a mapping such that $\theta(\mathsf{P})$ is defined for all $\mathsf{P} \in \mathscr{P}$, and let $\mathscr{P}(\theta) \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) = \theta\}$. Suppose the following conditions are satisfied:*

(BSE1) *for every $\theta \in \Theta$, there exists a $\mathsf{P} \in \mathscr{P}$ such that $\theta(\mathsf{P}) = \theta$;*

(BSE2) *for all $\pi \in \,]0, 1[$, $\theta_1 \in \Theta$, $\theta_2 \in \Theta$ and $\mathsf{P}_1 \in \mathscr{P}(\theta_1)$, there exists a probability distribution $\tilde{\mathsf{P}}$ such that $\pi\mathsf{P}_1 + (1-\pi)\tilde{\mathsf{P}} \in \mathscr{P}(\theta_2)$.*

For any positive integer ν, $\mathcal{B}_\nu$ denotes the set of all Borel functions defined on $\mathbb{R}^\nu$, taking values in $[0, 1]$. Then for any integer $N \geq 1$, for any $\theta \in \Theta$ and any function $f \in \mathcal{B}_{Nd}$,

$$\sup_{\mathsf{P}\in\mathscr{P}(\theta)} \mathsf{E}_\mathsf{P}(f_N) = \sup_{\mathsf{P}\in\mathscr{P}} \mathsf{E}_\mathsf{P}(f_N) \text{ and } \inf_{\mathsf{P}\in\mathscr{P}(\theta)} \mathsf{E}_\mathsf{P}(f_N) = \inf_{\mathsf{P}\in\mathscr{P}} \mathsf{E}_\mathsf{P}(f_N)\,,$$

where for $f \in \mathcal{B}_{Nd}$ and $\mathsf{P} \in \mathscr{P}$, $f_N \equiv f(X_1, \dots, X_N)$ and:

$$\mathsf{E}_\mathsf{P}(f_N) \equiv \int_{\mathbb{R}^{Nd}} f(x_1, \dots, x_N) \mathsf{d}\mathsf{P}^{\otimes N}(x_1, \dots, x_N).$$

In most applications of Theorem 3.1, we will have $N = 1$ and $X = (X_1, \dots, X_n)$ where X_i is a real random q-vector, so that $d = qn$. Notice Theorem 3.1 also applies to fully nonparametric models if we set $\theta(\mathsf{P}) = \mathsf{P}$.

Corollary 3.2 (PROPERTIES OF UMP AND SIMILAR TESTS FOR NON-TESTABLE HYPOTHESES). *Let $\mathscr{P}$ be a statistical model for a n-tuple $(X_1, \dots, X_n)$ of real random q-vectors, with true distribution $\overline{\mathsf{P}} \in \mathscr{P}$. Let θ be the mapping of Theorem 3.1. Consider the problem of testing $H_0 : \theta(\overline{\mathsf{P}}) \in \Theta_0$ against $H_1 : \theta(\overline{\mathsf{P}}) \in \Theta_1$, where Θ_0 and Θ_1 are disjoint non empty subsets of Θ with $\Theta_0 \cup \Theta_1 = \Theta$. If BSE1 and BSE2 hold, then for any $\alpha \in]0, 1[$ and any α-level test φ,*

(1) *the test $\varphi^*(\alpha)$ defined in* (2.2) *is UMP for testing H_0 against H_1;*

(2) *if φ is similar at level α, then $\mathsf{E}_\mathsf{P}(\varphi_n) = \alpha, \forall \mathsf{P} \in \mathscr{P}$.*

(3) *if φ is not similar then φ is biased at level α.*

As part (1) of Corollary 3.2 holds whatever the choices for Θ_0 and Θ_1, BSE1 and BSE2 are sufficient conditions for the non-testability of θ.

3.2 Refined results

A careful look at the proof of Theorem 3.1 shows if BSE1 and BSE2 hold, for any $\theta_0 \in \Theta$, any $\mathsf{Q} \notin \mathscr{P}(\theta_0)$, we may find $\mathsf{P} \in \mathscr{P}(\theta_0)$ such that $\forall N \geq 1$, $\sup_{g \in \mathcal{B}_{Nq}} |\mathsf{E}_\mathsf{P}(g_N) - \mathsf{E}_\mathsf{Q}(g_N)|$ is arbitrarily small. In the context of testing $H_0(\theta_0)$ against $H_1(\theta_0)$ with a sample $(X_1, \dots, X_n) \in \mathbb{R}^{qn}$, this translates into

$$\forall \epsilon > 0, \forall \mathsf{Q} \notin \mathscr{P}(\theta_0), \exists \mathsf{P} \in \mathscr{P}(\theta_0) \text{ such that } |\mathsf{E}_\mathsf{P}(\varphi_n) - \mathsf{E}_\mathsf{Q}(\varphi_n)| < \epsilon, \forall \varphi \in \mathcal{B}_{nq}. \quad (3.1)$$

In other words, under BSE1 and BSE2, for any test φ the distance between the type-1 risk and the power must be arbitrarily small under *any* alternative. This means that when bounding the type-1 risk (or the size) of a test, one also necessarily bounds its power. This leads to part (1) of Corollary 3.2.

This conclusion can be refined by noting that under BSE1 and BSE2, (3.1) holds whatever the choice of $\theta_1 \neq \theta_0$ and any P_1 with $\theta(\mathsf{P}_1) = \theta_1$. In particular, if Θ is a metric space with metric ρ, we have (3.1) with $\rho(\theta_0, \theta_1)$ arbitrarily large. As $D(\mathsf{P}_0, \mathsf{P}_1) \equiv \sup_{\varphi \in \mathcal{B}_{nd}} |\mathsf{E}(\varphi_n) - \mathsf{E}(\varphi_n)|$ defines a pseudo-distance on $\mathscr{P}$, we may interpret (3.1) in terms of continuity of the parameter mapping θ. Indeed we have under BSE1 and BSE2:

$$\forall \mathsf{P}_1 \in \mathscr{P}(\theta_1), \forall 0 < \epsilon \leq \mathrm{diam}(\Theta), \forall \delta > 0, \exists \mathsf{P}_0 \in \mathscr{P}(\theta_0) \text{ such that}$$
$$D(\mathsf{P}_1, \mathsf{P}_0) < \delta \text{ and } \rho\big(\theta(\mathsf{P}_1), \theta(\mathsf{P}_0)\big) \geq \epsilon,$$

where $\mathrm{diam}(\Theta)$ denotes the diameter of Θ. This implies that (*i*) the mapping θ is discontinuous at every $\mathsf{P}_1 \in \mathscr{P}(\theta_1)$ and (*ii*) discontinuity jumps are of arbitrarily large amplitude.

This result is similar to those obtained by Pötscher (2002) in the context of point estimation. Pötscher (2002, Corollary 2.2) shows that when the parameter is a discontinuous function of probability distributions, the minimum risk of any estimator of the parameter is strictly positive in every neighborhood of the true value. It is possible to derive a similar result for tests when the parameter of interest is non-testable.

Proposition 3.3 (PROBABLITY OF TYPE II ERROR FOR A NON-TESTABLE HYPOTHESIS). *If θ is a non-testable parameter, then for any $\alpha \in \,]0,1[$, any $\theta_0 \in \Theta$ and any $\mathsf{P}_1 \notin \mathscr{P}(\theta_0)$, we have*

$$\inf_{\varphi \in \Phi_\alpha} \inf_{\epsilon > 0} \sup_{\mathsf{P} \in \mathscr{V}_\epsilon(\mathsf{P}_1)} \big[1 - \mathsf{E}_{\mathsf{P}}(\varphi_n)\big] \geq 1 - \alpha$$

where $\mathscr{V}_\epsilon(\mathsf{P}_1) \equiv \{\mathsf{P} \in \mathscr{P} : D(\mathsf{P}, \mathsf{P}_1) \leq \epsilon\}$ and Φ_α is the set of all tests φ such that $\sup_{\mathsf{P} \in \mathscr{P}(\theta_0)} \mathsf{E}(\varphi_n) \leq \alpha$.

Suppose we are interested in testing $H_0 : \overline{\theta} = \theta_0$ against $H_1 : \overline{\theta} \neq \theta_0$. The above proposition shows that for any distribution P_1 in the alternative there is no small enough neighborhood of P_1 such that the type 2 risk of an α-level test is less than $1 - \alpha$.

A further refined result is obtained by noting that BSE2 is actually not a necessary condition for obtaining 1 of Corollary 3.2. Define $\mathscr{P}(\Theta_k) \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) \in \Theta_k\}$, $k = 0, 1$. Result 1 of Corollary 3.2 follows as soon as $\mathscr{P}(\Theta_0)$ is dense in $\mathscr{P}(\Theta_1)$ w.r.t. D:

(BSE3) $\quad \forall \epsilon \in \,]0,1], \forall \mathsf{P}_1 \in \mathscr{P}(\Theta_1), \exists \mathsf{P}_0 \in \mathscr{P}(\Theta_0)$, s.t. $D(\mathsf{P}_0, \mathsf{P}_1) < \epsilon$;

[see Tibshirani and Wasserman (1988, definition of section 2) or Romano (2004, condition A)]. Using Lemma B.2 (see the Appendix), it is easily seen that BSE3 holds whenever

(BSE4) $\quad \forall \pi \in]0,1[, \forall \mathsf{P}_1 \in \mathscr{P}(\Theta_1), \exists \tilde{\mathsf{P}}$ s.t. $\pi \mathsf{P}_1 + (1-\pi)\tilde{\mathsf{P}} \in \mathscr{P}(\Theta_0)$

holds, where $\tilde{\mathsf{P}}$ is some probability distribution. Condition BSE4 is weaker than BSE2. It requires that contaminating appropriately any distribution in the alternative hypothesis yields a distribution in the null. In our paper, many results similar to (1) of Corollary 3.2 will be shown to follow directly from BSE4.

Results related to Theorem 3.1 and Corollary 3.2 have been obtained by Tibshirani and Wasserman (1988), Devroye and Lugosi (2002), Romano (2004) and Forchini (2005). Gleser and Hwang (1987), Blough (1992), Pfanzagl (1998), Dufour (1997) and Faust (1999) derived similar results for set and point estimation. Some of these references also provide a result similar to the following one.

Proposition 3.4 (CONTINUITY OF NON-TESTABILITY).*Let $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ define a statistical model for a random vector X, with Θ a set on which some mode of convergence may be defined. Consider the testing problem $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$. Let Θ_1^* be the set of all elements of Θ_1 defined by*

$$\theta_1 \in \Theta_1^* \iff (\exists\{\theta_{0,n} : n \geq 1\} \subset \Theta_0, \lim_{n\to\infty} \theta_{0,n} = \theta_1).$$

Introduce the following assumption:

(BSE5) $\{\theta_n : n \geq 1\} \subset \Theta$ *and* $\lim_{n\to\infty} \theta_n = \theta \in \Theta$ *implies* $\mathsf{P}_{\theta_n} \xrightarrow{w} \mathsf{P}_\theta$, $n \to \infty$, *where* $\xrightarrow{w}$ *denotes weak convergence.*

Let φ be an α-level test of H_0 against H_1 and assume $\Theta_1^* \neq \varnothing$. Under assumption BSE5, if for some $\theta_1 \in \Theta_1^*$ we have $\mathsf{P}_{\theta_1}\big(\partial\{X : \varphi(X) = 1\}\big) = 0$, then $\inf_{\theta\in\Theta_1} \mathsf{P}_\theta(\varphi = 1) \leq \alpha$. Moreover, if this holds for any $\theta_1 \in \Theta_1^*$ and $\Theta_1^* = \Theta_1$, then $\sup_{\theta\in\Theta_1} \mathsf{P}_\theta(\varphi = 1) \leq \alpha$.

3.3 Non-testability and the "size" of models

It is certainly true that non-testability does not arise in sufficiently "small" statistical models. This is for instance true if the set of possible probability distributions is reduced to only two elements, for in that case, we may invoke the Neyman-Pearson lemma. It may thus be thought that non-testability arises when the problem under test is "too large", in the sense that the model is not constrained enough. The following examples show that what is a "constrained enough" model is unclear.

EXAMPLE 2. Consider a model for a n-tuple $X_1, \ldots, X_n$ of i.i.d. real random variables. In this model it is assumed that $\mathsf{E}(X_1)$ exists and is either θ_0 or θ_1, so that $\Theta = \{\theta_0, \theta_1\}$. The family $\mathscr{P}$ consists of all such probability distributions for $X_1, \ldots, X_n$. Consider testing $H_0 : \overline{\theta} = \theta_0$ against $H_1 : \overline{\theta} = \theta_1$. Choose $\epsilon \in]0, 1[$, $\pi \in]0, 1[$ such that $1 - \pi^n < \epsilon$, and $\mathsf{P}_1 \in \mathscr{P}$ such that $\theta(\mathsf{P}_1) = \theta_1$. Now define $\tilde{\theta} \equiv \frac{\theta_0 - \pi\theta_1}{1-\pi}$ and

$$\tilde{X}_i = U_i X_i + (1 - U_i)\tilde{\theta}, \quad i = 1, \dots, n, \tag{3.2}$$

where $X_1, \dots X_n$ are i.i.d. P_1 and $U_1, \dots, U_n$ i.i.d. Bernoulli $\mathscr{B}(\pi)$, with $X_1, \dots, X_n$, $U_1, \dots, U_n$ independent. We easily check that $\mathsf{E}(\tilde{X}_i) = \theta_0$. In other words, if P_0 denotes the distribution of $\tilde{X}_1$, P_0 satisfies the null hypothesis. From equation (3.2) and Lemma B.3 (see the Appendix), we get $\left|\mathsf{E}\big(g(X_1, \dots, X_n)\big) - \mathsf{E}\big(g(\tilde{X}_1, \dots, \tilde{X}_n)\big)\right| < \epsilon$ for all measurable function $g : \mathbb{R}^n \longrightarrow [0, 1]$. If φ is a test of H_0 against H_1, the above inequality establishes that the probability of rejecting the null under P_0 (*i.e.*, when the null is true) can differ from the probability of rejecting it under P_1 (*i.e.*, when the null is false) by at most ϵ. As this holds for any positive ϵ, if φ has level α, we get part (1) of Corollary 3.2. □

EXAMPLE 3. Denote $A = \{0, 1\}$ and let $\mathscr{P}$ be the set of all probabilities with support A, *i.e.*, $\mathsf{P} \in \mathscr{P} \iff \mathsf{P}(A) = 1$. Let $\overline{\mathsf{P}} \in \mathscr{P}$ be the true probability distribution from which we observe an i.i.d. sample $X_1, \dots, X_n$. For some reason, one wishes to test $H_0 : \overline{\mathsf{P}} \in \mathscr{P}_0$ against $H_1 : \overline{\mathsf{P}} \in \mathscr{P} \setminus \mathscr{P}_0$, where $\mathscr{P}_0 \equiv \{\mathsf{P} \in \mathscr{P} : \mathsf{P}(\{0\}) > 0\}$. H_0 is non-testable against H_1, as we now prove. Take $\epsilon \in]0, \infty[$ and $\pi \in]0, 1[$ such that $1 - \pi^n < \epsilon$. Choose $\mathsf{P}_1 \in \mathscr{P} \setminus \mathscr{P}_0$ and define $\mathsf{P}_0 = \pi\mathsf{P}_1 + (1 - \pi)\delta_{\{0\}}$, where $\delta_{\{0\}}$ is the Dirac mass at 0. It is easily checked that $\mathsf{P}_0 \in \mathscr{P}_0$. Using Lemma B.2, (see the Appendix) we may conclude, as in the previous example, that H_0 is non-testable against H_1.[1] □

In the opposite direction, hypotheses about the parameters of some very "large" models turn out to be testable. Consider for instance a model for a n-tuple of independent (non necessarily identically distributed) random variables with common median. This typically qualifies as a very large model. However, it is well known that the median is a testable parameter in this model [see Lehmann and D'Abrera (1998)].

4 Non-testability and asymptotic approximations

We now consider consequences of BS-type results for testing procedures based on asymptotic approximations. It is frequently claimed that, although a given testing prob-

[1] In this example, if the null and alternative hypotheses are exchanged, $\mathsf{P}(\{0\}) = 0$ is testable against $\mathsf{P}(\{0\}) > 0$.

lem (H_0, H_1) may be non-testable, there could exist valid and consistent test procedures for this problem. In the terminology of Section 2, testability of (H_0, H_1) may be achieved for a sufficiently large sample size in the following sense.

Definition 4.1 (ASYMPTOTIC TESTABILITY). *Consider* $\{\mathscr{P}_n : n \geq 1\}$ *a sequence of statistical models, where for any* n, $\mathscr{P}_n$ *is a set of probability distributions for a* n*-tuple* $(X_1, \ldots, X_n)$ *of real random* q*-vectors. Let also* $\overline{\mathsf{P}}_n \in \mathscr{P}_n$ *be the true distribution of* $(X_1, \ldots, X_n)$, $n = 1, 2, \ldots$ *A sequence of null hypotheses* $H_{0,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{0,n}$ *is* asymptotically testable *against the sequence of alternatives* $H_{1,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{1,n}$ *when for any given* $\alpha \in]0, 1[$ *there exists a sequence* $\{\varphi_n : n \geq 1\}$ *of tests such that*

AT1. $\limsup_{n\to+\infty} \sup_{\mathsf{P}_n \in \mathscr{P}_{0,n}} \mathsf{E}_{\mathsf{P}_n}(\varphi_n) \leq \alpha,$

AT2. $\liminf_{n\to+\infty} \mathsf{E}_{\mathsf{P}_n}(\varphi_n) > \alpha,$ *for any sequence* $\{\mathsf{P}_n : n \geq 1\}$ *with* $\mathsf{P}_n \in \mathscr{P}_{1,n}$ $\forall n.$

Condition AT1 insures that φ_n is a *uniformly* asymptotically α-level test [see Romano (2004)], which is obviously stronger than the *pointwise* control on type-1 error, but allows for approximation of the *size* in finite samples. A test for which AT1 holds has level α in the limit.

Obviously, if a consistent α-level test [see Lehmann (1986, p.478)] exists, then AT1 and AT2 hold. Notice that in that case, the type 2 risk converges to 0 for *any* sequence of alternatives. We call a test with this property *everywhere consistent*. If along *some* sequence of alternatives, the $\limsup$ of the type 2 risk of a test converges to 0, we say this test is *somewhere consistent*. Finally a test is *nowhere consistent* if the $\limsup$ of its type 2 risk is strictly positive for any sequence of alternatives.

Asymptotic implications of (non-)testability cannot be obtained as an immediate consequence of BS-type results. Indeed it is assumed in Bahadur and Savage (1956) that the sample is a.s. finite. However, it is easy to see that non-testability also entails domination of asymptotic procedures by the uniform test $\varphi^*(\alpha)$. We have the following result, which shows that for a (partially) non-testable problem AT1 and AT2 cannot hold simultaneously.

Proposition 4.2 (IMPOSSIBILITY OF AN ASYMPTOTICALLY VALID TEST WITH NON-TRIVIAL POWER FOR A NON-TESTABLE HYPOTHESIS). *Consider a sequence of statistical models* $\mathscr{P}_n$ *and a corresponding sequence of testing problems defined by* $H_{0,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{0,n}$ *and* $H_{1,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{1,n}$ *with* $\mathscr{P}_n = \mathscr{P}_{0,n} \cup \mathscr{P}_{1,n}$.

(1) *Assume for any* n, $H_{0,n}$ *is partially non-testable against* $H_{1,n}$. *Any test* φ_n *with level* $\alpha < 1$ *in the limit cannot be everywhere consistent.*

(2) *Assume for any* n, $H_{0,n}$ *is non-testable against* $H_{1,n}$. *Any somewhere consistent test* φ_n *has level 1 in the limit.*

One particular case of Proposition 4.2 is the original result of BS on the testability of the expectation. Indeed, let $\mathscr{P}_1$ be the set of all probability distributions on $\mathbb{R}$ with a finite expectation. For any sample size $n \geq 1$, define $\mathscr{P}_n$ the set of all n-fold product probabilities $\mathsf{P}^{\otimes n}$ such that P is in $\mathscr{P}_1$. For any $\mathsf{P}_n \in \mathscr{P}_n$, let $\theta_n(\mathsf{P}_n)$ denote the expectation of P_n. Fix a real number θ_0, and consider testing $H_{0,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{0,n}$ against $H_{1,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{1,n}$, where $\mathscr{P}_{0,n} \equiv \{\mathsf{P}_n \in \mathscr{P}_n : \theta_n(\mathsf{P}_n) = \theta_0 \iota_n\}$, $\mathscr{P}_{1,n} \equiv \mathscr{P}_n \setminus \mathscr{P}_{0,n}$, and $\iota_n \equiv (1, \ldots, 1)' \in \mathbb{R}^n$. If we choose $\overline{\mathsf{P}} \in \mathscr{P}_1$ such that $\overline{\mathsf{P}}_n = \overline{\mathsf{P}}^{\otimes n}$, it is clear that this problem amounts to testing $\theta_1(\overline{\mathsf{P}}) = \theta_0$ against $\theta_1(\overline{\mathsf{P}}) \neq \theta_0$ from an i.i.d. random sample of size n of $X \sim \overline{\mathsf{P}}$. Bahadur and Savage (1956) show that for any sample size n, this problem is non-testable. Therefore, any consistent test φ must have size 1 in the limit. To obtain a "uniformly" asymptotically α-level *and* consistent t-test, Romano (2004, see equation (9)) imposes some kind of uniform integrability condition on the standardized random variable $(X - \mathsf{E}(X))/\sigma(X)$.

Consistent tests are typically derived invoking a central limit theorem showing that for some $\alpha < 1$, a given sequence of tests $\{\varphi_n : n \geq 1\}$ has a type 1 risk converging to α for any sequence of null distributions, while its power converges to 1 for some sequences of alternatives. The following Corollary of Proposition 4.2 shows that for such tests, the approximation of the size must be arbitrarily bad when non-testability problems arise.

Corollary 4.3 (IMPOSSIBILITY OF UNIFORM SIZE CONVERGENCE). *Consider* $\{H_n = (H_{0,n}, H_{1,n}) : n \geq 1\}$, *the sequence of testing problems of Proposition 4.2, and let* $\{\varphi_n : n \geq 1\}$ *be an associated sequence of tests satisfying*

$$\lim_{n\to\infty} \mathsf{E}_{\mathsf{P}_{0,n}}(\varphi_n) = \alpha < 1\,, \tag{4.1}$$

for any sequence $\{\mathsf{P}_{0,n} : n \geq 1\} \subset \mathscr{P}_{0,\infty}$, *where* $\mathscr{P}_{0,\infty} \equiv \bigtimes_{n=1}^{\infty} \mathscr{P}_{0,n}$.

(1) *If* $\{\varphi_n : n \geq 1\}$ *is somewhere consistent and* H_n *is non-testable for all* n, *then the convergence in* (4.1) *cannot be uniform on* $\mathscr{P}_{0,\infty}$.

(2) *If* $\{\varphi_n : n \geq 1\}$ *is everywhere consistent and* H_n *is partially non-testable for all* n, *then the convergence in* (4.1) *cannot be uniform on* $\mathscr{P}_{0,\infty}$.

Corollary 4.3 shows the pointwise convergence of the type 1 risk to α is arbitrarily slow over $\mathscr{P}_{0,\infty}$. As a consequence, no matter how large the sample size n, the limit α of the sequence $\{\mathsf{E}_{\mathsf{P}_{0,n}}(\varphi_n) : n \geq 1\}$ cannot be an approximation of the size of φ_n. Indeed, we may always find a $\mathsf{P}_{0,n} \in \mathscr{P}_{0,n}$ such that $\mathsf{E}_{\mathsf{P}_{0,n}}(\varphi_n)$ is arbitrarily close to 1, provided n is large enough. As a matter of fact, techniques used in the proof of Theorem 3.1 point out sequences of null distributions for which the type 1 risk is as large as one wishes.

Many corrections have been proposed in order to reduce discrepancies between the actual size of a test and the desired level. Bootstrap and Bartlett's corrections are the most well-known. The theoretical arguments for preferring these "corrected" methods is often based on asymptotic approximations. For instance, one shows that the actual size of a "corrected" test converges more quickly towards α than its "usual" counterpart (see, *e.g.* Davidson and MacKinnon (1999)). Under (partial) non-testability, Proposition 4.2 and Corollary 4.3 show that if both the "corrected" and "usual" procedures are (everywhere) consistent, the asymptotic level must be 1, regardless of the claimed level and the method of correction. In other words, these corrections fail and are ineffective in the limit, despite their asymptotic justification.

Note that obtaining a central limit theorem typically requires imposing further conditions on the model. For instance, in an i.i.d. sampling model conditions BSE1 and BSE2 are not sufficient to guarantee convergence of the sample mean to its theoretical value. One would typically impose the existence of moments of order $2 + \delta$ for some strictly positive δ. But this does not affect the non-testability of the expectation, as conditions BSE1 and BSE2 are still fulfilled. Actually, imposing the existence of moments of any order would not contradict any of the conditions BSE1 and BSE2. Therefore, non-testability of the expectation remains. As a consequence, Proposition 4.2 and its Corollary 4.3 apply. More generally, for a given parameter θ, as long as further conditions imposed on the model with the hope of getting rid of the non-testability, are compatible with BSE1 and BSE2 for any sample size, the problem remains.

There exist semiparametric models in which meaningful inference is indeed possible. Of course, at least one of the conditions BSE1 and BSE2 must be violated. We already mentioned Romano (2004) who imposes a kind of uniform integrability condition to derive a consistent uniformly asymptotically α-level t-test. Another instance is Anderson (1967) where it is assumed that the support of the distribution is bounded by some known values [see also Romano and Wolf (2000) for a more general treatment]. This assumption violates BSE2. It could also be argued that in the model $Y_i = \theta + \varepsilon_i$, where the ε_is are i.i.d. with a distribution symmetrical about 0, inference on the median θ of the Y_is amounts to inference on their expectation (provided this expectation exists). Several valid and powerful procedures have been proposed to test θ in this context. But it is easy to see that imposing a symmetrical distribution for the Y_is violates the convexity condition BSE2 (take for instance a mixture of two uniform distributions with disjoint supports). Now, when the symmetry condition is relaxed, the median does not necessarily coincide with the expectation. While the former is testable, the latter is not.

5 Examples of non-testability: covariances and regression coefficients

As shown by Bahadur and Savage (1956), the expectation is non-testable in a simple i.i.d. sampling model. In Theorem 3.1, we showed this could also be the case for parameters other than the expectation. In this section, we investigate the testability of various usual parameters.

5.1 Covariance and regression

The following Proposition 5.1 shows that the covariance is non-testable parameter.

Proposition 5.1 (COVARIANCE NON-TESTABILITY).*Let $\mathscr{P}$ be the family of all probability distributions* P *for a couple $Z = (X', Y')'$ of random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, for which $\mathsf{V}_\mathsf{P}(Z)$ exists and is finite. The parameter $\theta(\mathsf{P}) = \mathsf{Cov}_\mathsf{P}(X, Y)$ is non-testable.*
If $\mathscr{P}$ is restricted to $\mathscr{P}^* \equiv \{\mathsf{P} \in \mathscr{P} : \mathsf{V}_\mathsf{P}(Z)$ is invertible$\}$, $\mathsf{Cov}_\mathsf{P}(X, Y)$ remains non-testable.

From Proposition 5.1, it is easy to show that the coefficients of the linear regression are non-testable.

Proposition 5.2 (NON-TESTABILITY OF LINEAR REGRESSION COEFFICIENTS).*In the model $\mathscr{P}^*$ of Proposition 5.1, for a given $\mathsf{P} \in \mathscr{P}^*$, define $\mathsf{EL}_\mathsf{P}(Y|X)$, the linear regres-*

sion of Y on X:

$$\mathsf{EL_P}(Y|X) = \mathsf{E_P}(Y) + \mathsf{Cov_P}(Y, X)\mathsf{V_P}(X)^{-1}\big[X - \mathsf{E_P}(X)\big].$$

The parameter $\theta(\mathsf{P}) \equiv \mathsf{Cov_P}(Y, X)\mathsf{V_P}(X)^{-1}$ *is non-testable.*

5.2 Variance

Clearly conditions BSE1 and BSE2 (or BSE1, and BSE6 to BSE8 of Lemma B.1) imply non-testability of a given parameter. However, although these conditions may not hold, *some* testing problems about this parameter may remain non-testable. This may entail the partial non-testability of this parameter as shown by Proposition 5.3 for the variance.

Proposition 5.3 (VARIANCE PARTIAL NON-TESTABILITY). *Let $\mathscr{P}$ be the family of all probability distributions* P *for a real random variable X for which the parameter* $\theta(\mathsf{P}) \equiv \mathsf{V_P}(X)$ *exists and is finite. Then*

(1) *for any strictly positive real number θ^*, $H_0(\theta^*) : \theta(\overline{\mathsf{P}}) \geq \theta^*$ is non-testable against $H_1(\theta^*) : \theta(\overline{\mathsf{P}}) < \theta^*$;*

(2) *the variance is partially non-testable.*

In the model of Proposition 5.3, condition BSE7 (see Lemma B.1 in the Appendix) fails, because we may find two variances θ_1, θ_2 and a $\lambda \in]1, +\infty[$ for which $\theta_1 < \frac{\lambda-1}{\lambda}\theta_2$. In such a case, $\lambda\theta_1 + (1 - \lambda)\theta_2$ is strictly negative and cannot be a variance. As a consequence, although one may not test $\mathsf{V}(X) \geq 1$ against $\mathsf{V}(X) < 1$, say, a powerful test of $\mathsf{V}(X) \leq 1$ against $\mathsf{V}(X) > 1$ exists.[2]

It is easy to see that Propositions 5.1 to 5.3 hold even if it is further assumed that P is the common distribution of n independent couples $(X_1, Y_1), \ldots, (X_n, Y_n)$

BS's results as well as Propositions 5.1 to 5.3 show that (partial) non-testability arises for "moment" parameters. Our next result show the problem is not restricted to such parameters.

[2]Such a test may be based on a Markov-type inequality involving the empirical and theoretical variances.

5.3 Support of a distribution

Proposition 5.4 (DISTRIBUTION SUPPORT PARTIAL NON-TESTABILITY). *Let $\mathscr{P}$ be the family of all probability distributions on $\mathbb{R}$. For $\mathsf{P} \in \mathscr{P}$ let $\theta(\mathsf{P})$ be defined by $\theta(\mathsf{P}) \equiv \sup\{x : \mathsf{P}(X \leq x) < 1\}$. For any $\theta \in \mathbb{R}$ define $\mathscr{P}(\theta) = \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) \geq \theta\}$. For any $\theta^* \in \mathbb{R}$, $H_0 : \overline{\mathsf{P}} \in \mathscr{P}(\theta^*)$ is non-testable against $H_1 : \overline{\mathsf{P}} \in \mathscr{P} \setminus \mathscr{P}(\theta^*)$ and θ is partially non-testable.*

Notice when permuting the null and alternative hypotheses we can propose a somewhere powerful valid test for the support upper bound. Indeed the following procedure "reject $H_0 : \overline{\mathsf{P}} \in \mathscr{P} \setminus \mathscr{P}(\theta^*)$ whenever at least one observation is larger that θ^* and otherwise draw U as $\mathcal{U}_{[0,1]}$ and reject H_0 if we observe $U \leq \alpha$" has level α. Moreover we can easily find a $\mathsf{P} \in \mathscr{P}(\theta^*)$ such that the power $\mathsf{P}(X_i > \theta^*$ for some $i = 1, \ldots, n)$ is strictly larger than α.

Also notice that partial non-testability obtains for the lower bound of a distribution when permuting the null and alternative hypotheses of Proposition 5.4.

6 Testability issues in a semiparametric linear regression model

Although Bahadur and Savage (1956) result is established in the context of pure i.i.d. sampling statistical models, we mentioned in Section 3 that the i.i.d. assumption is not necessary for deriving a similar result (see Corollary 3.2). We may then be interested in investigating non-testability issues in models where the form of the heterogeneity is given. In many econometric models, the heterogeneity in the distribution of an endogenous variable is "explained" by a set of covariates.

In this Section, we deal with one of the most simple of such models, namely the linear regression model. Results will be extended to more general models in Section 7.

From the previous sections we know that (*i*) the expectation is a non-testable parameter and (*ii*) in general models, the linear regression coefficient is non-testable (see Proposition 5.2). From this, we could expect the parameter of a linear regression model to be non-testable. Surprisingly, in a linear regression model which could be viewed as a natural extension of the model considered in Bahadur and Savage (1956), some hypotheses about this parameter are testable.

6.1 A semiparametric linear regression model with identically distributed error terms

Let $X \equiv (X_1', \ldots, X_n')'$ be a real random nK-vector and $Y = (Y_1, \ldots, Y_n)'$ a real random n vector, and let $\mathsf{E}_\mathsf{P}(Y_i|X)$ denotes the expectation of Y_i conditionally on X, when $(X', Y')' \sim \mathsf{P}$. Our model is the set $\mathscr{P}$ of all probability distributions P for $(X', Y')'$ such that the following two conditions hold.

C1. $\forall i = 1, \ldots, n$, $\mathsf{E}_\mathsf{P}(Y_i|X) = \mu + \beta' X_i$ for a unique $(\mu, \beta) \in \mathbb{R} \times \mathbb{R}^K$ a.s.

C2. Conditionally on X, the error terms $\varepsilon_i \equiv Y_i - \mathsf{E}_\mathsf{P}(Y_i|X)$, $i = 1, \ldots, n$, are identically distributed.

For $\mu \in \mathbb{R}$ and $\beta \in \mathbb{R}^K$, we also denote $\theta \equiv (\mu, \beta')' \in \mathbb{R}^{K+1}$. The parameter of interest is defined by $\theta : \mathscr{P} \to \mathbb{R}^{K+1}$ with $\theta(\mathsf{P}) = \theta \iff \mathsf{E}_\mathsf{P}(Y_i|X) = \mu + \beta' X_i$, $i = 1, \ldots, n$. We also define the sub-parameters μ and β as the projection of θ on $\mathbb{R}$ and $\mathbb{R}^K$, respectively. Finally, $\overline{\mathsf{P}}$ denotes the true distribution of $(X', Y')'$ and $\overline{\theta}$ is defined as $\overline{\theta} = \theta(\overline{\mathsf{P}})$. Similarly, $\overline{\mu} = \mu(\overline{\mathsf{P}})$ and $\overline{\beta} = \beta(\overline{\mathsf{P}})$, so that $\overline{\theta} = (\overline{\mu}, \overline{\beta}')'$.

6.1.1 Non-testability of the intercept parameter and partial non-testability of the whole parameter

Our first result shows that in this model, a BS-type result applies to the intercept parameter μ.

Proposition 6.1 (INTERCEPT NON-TESTABILITY IN LINEAR REGRESSIONS WITH IDENTICALLY DISTRIBUTED ERRORS). *In the model defined by C1 and C2, the following holds true:*

(1) *the parameter* μ *is non-testable;*

(2) *the parameter* θ *is partially non-testable;*

(3) *for all* $\alpha \in [0, 1[$, *no consistent asymptotically* α*-level test exists for* θ.

The above Proposition does not contradict results obtained in the literature on non-parametric methods for the linear model by, *e.g.*, Adichie (1967), Bickel (1971) and Jureckova (1971). An overview is given by Puri and Sen (1985). They propose tests about θ and μ under conditions C1 and i.i.d. error terms, conditionally on X (which is evidently stronger than C2).[3] A careful look at Puri and Sen (1985) reveals that a symmetry assumption is added every time μ is fixed under the null.[4] It is interesting to note this book contains no reference to Bahadur and Savage (1956). We already mentioned that the symmetry assumption allows to escape BS impossibility results (see the last paragraph of section 4 above). More interestingly, this symmetry assumption is *not* imposed for testing problems on β only. Puri and Sen (1985) provide no reason why symmetry is needed when μ is fixed under the null and may be avoided when it is a nuisance parameter. This will become clearer as we now show that β turns out to be testable under C1 and C2 only and that symmetry is not necessary for existence of valid and somewhere powerful test procedures about β.

6.1.2 Testability of the slopes

The result of this section shows that the slope parameter β is testable.

Proposition 6.2 (SUFFICIENT CONDITION FOR SLOPE TESTABILITY IN LINEAR REGRESSIONS WITH IDENTICALLY DISTRIBUTED ERRORS). *In the model defined by C1 and C2, choose $\beta_0 \in R^K$ and consider testing $H_0(\beta_0) : \overline{\beta} = \beta_0$ against $H_1(\beta_0) : \overline{\beta} \neq \beta_0$. Let $n_{X,0} = \max_{\beta \in \mathbb{R}^K \setminus \{\beta_0\}} \#\{(\beta - \beta_0)'X_1, \ldots, (\beta - \beta_0)'X_n\}$. $H_0(\beta_0)$ is testable against $H_1(\beta_0)$ at any level $\alpha \in [\frac{1}{n_{X,0}}, 1[$. The parameter β is testable.*

Notice that condition C1 ensures $n_{X,0} \geq 2$. Proposition 6.2 is demonstrated by fixing β_0 and exhibiting an α-level test of $H_0(\beta_0)$ with power arbitrarily close to 1 for some distribution compatible with $H_1(\beta_0)$. Incidentally, it appears that no such test has been proposed in the context of models defined by C1 and C2.

The next section shows that, in a semiparametric linear regression model, β becomes non-testable if the assumption of identically distributed error terms is relaxed.

6.2 Semiparametric linear models with non-identically distributed error terms

In this section we examine non-testability issues in linear regression models when the condition C2 is no longer imposed. The model we consider now is the set $\mathscr{P}$ of all prob-

[3] It is easy to check that the proof of part (1) of Proposition 6.1 establishes that μ remains non-testable when we further assume independent error terms.

[4] See Puri and Sen (1985, p.137 equation (5.2.26), p.146 first line, p.186 first line, and p.240 Theorem 7.2.1; see also p.11 third line of the second paragraph).

ability distributions P for $(X', Y')'$ which satisfies condition C1.We have the following result:

Proposition 6.3 (SLOPE NON-TESTABILITY IN LINEAR REGRESSIONS WITH NON-IDENTICALLY DISTRIBUTED ERRORS).*In the model defined by C1, the parameter* β *is non-testable.*

In view of Propositions 6.2 and 6.3, a natural issue to address is the search of restrictions that can be imposed on $\mathscr{P}$ to achieve the testability of β. Proposition 6.3 above shows that if only the identifiability of β and the linearity of the conditional expectation are assumed, the slope coefficient is non-testable. As shown by Proposition 6.2, imposing identically distributed error terms is one way to recover the testability of β. Our next result shows that, if other possibilities exist, they cannot be obtained by a combination of homoskedasticity, uncorrelatedness and independence of the error terms.

Proposition 6.4 (SLOPE NON-TESTABILITY IN LINEAR REGRESSIONS WITH NON-IDENTICALLY DISTRIBUTED ERRORS UNDER STRONGER ASSUMPTIONS).*In the model* $\mathscr{P}$ *defined above,* β *remains non-testable even if it is further assumed that the error terms are homoskedastic and/or uncorrelated and/or independent, conditionally on* X.

As a particular case of Proposition 6.4, the parameters of the usual textbook linear regression model with (conditionally) uncorrelated and homoskedastic error terms are non-testable (see point 3 in the proof of Proposition 6.4).

6.3 Linear regression model with i.i.d. observations and semiparametric limited dependent variable models

Finally, we mention imposing an i.i.d. condition on the observable variables or on the error terms are very different assumptions. We know from Proposition 6.2 that in a semiparametric linear regression model, the slopes are testable when it is assumed that the error terms are i.i.d.[5] Consider instead a semiparametric linear regression model where we assume the following condition:

C3. $(X_1', Y_1)', \ldots, (X_n', Y_n)'$ are i.i.d.

We then have the following property.

[5] In Proposition 6.2, it is only assumed that errors are identically distributed conditionally on the regressors. Actually, the proof of this proposition shows that the slopes remain testable if we assume that, conditionally on X, the error terms are also independent.

Proposition 6.5 (NON-TESTABILITY OF LINEAR REGRESSIONS FOR I.I.D. OBSERVATIONS). *In the model defined by C1 and C3 above, the parameter* β *is non-testable.*

The condition C3 is common in nonparametric settings (see section 7.2 below). It also appears in papers dealing with semi-parametric inference in single index models [see Lee (1996) for an overview]. In many cases, the underlying latent structure fulfills conditions C1, C3 and the following conditions.

C4. the K-th component of X_1 has an everywhere positive Lebesgue density conditional on the other components;

C5. $\overline{\beta}$ belongs to the unit sphere of $\mathbb{R}^K$ and the K-th coordinate of $\overline{\beta}$ is non zero.

The reason for condition C3 is that investigation of the asymptotic behaviors of the proposed procedures relies on U-statistics introduced by Serfling (1980) and Lehmann and D'Abrera (1998). Conditions C4 and C5 are motivated by identification issues arising in these models.

It is easy to see the proof of Proposition 6.5 remains valid when condition C4 on the distribution of the regressors is further imposed. Moreover, one can also check that C5 plays no role in this result. An immediate corollary to this result is the non-testability of the slope coefficients in single index models based on conditions C1 and C3 to C5 only [see for instance Sherman (1993)]. Indeed, inference in single index models may be viewed as inference on the latent model, where the available procedures depend on some function of the endogenous variable. For instance, if the latent model is defined by C1 and C3 to C5, and if only X and $Z_i = \text{sign}(Y_i)$, $i = 1, \ldots, n$, are observed, then the family of feasible tests on the slope $\overline{\beta}$ is a subset of all tests on $\overline{\beta}$ that would be available if Y_i, $i = 1, \ldots, n$, where observed.

7 Non-testability in models defined by moment conditions

We consider now general models where the parameter of interest is defined through estimating equations taking the form of moment conditions. Let $Z = (X', Y')' = (X_1', \ldots, X_n', Y_1', \ldots, Y_n')'$ be a n-tuple of real random vectors, with $X_i \in \mathscr{X} \subseteq \mathbb{R}^{p_X}$ and $Y_i \in \mathscr{Y} \subseteq \mathbb{R}^{p_Y}$, $i = 1, \ldots, n$. Let $\mathscr{P}$ be a set of probability distributions for Z, Θ be a set and Q be a mapping defined on $\mathscr{P} \times \Theta$ taking values in $\mathbb{R}^q$ such that for any

$\mathsf{P} \in \mathscr{P}$ there exists a $\theta \in \Theta$ such that $Q(\mathsf{P}, \theta) = 0$. This defines a correspondence $\uptheta$: $\mathscr{P} \to \mathcal{P}(\Theta)$ which we call the parameter of the model $\mathscr{P}$, where $\mathcal{P}(\Theta)$ denotes the set of all subsets of Θ. We will assume this parameter is identified in the sense that for any $\mathsf{P} \in \mathscr{P}$, the equality $Q(\mathsf{P}, \theta) = 0$ has a unique solution in Θ. With such an assumption, $\uptheta$ is a mapping from $\mathscr{P} \to \Theta$ defined by the equivalence $Q(\mathsf{P}, \theta_0) = 0 \iff \uptheta(\mathsf{P}) = \theta_0$ and the equality $Q(\mathsf{P}, \theta) = 0$ is called an estimating equation [see Godambe (1960)].

In many statistical models, this estimating equation takes the form of a moment condition. In the simplest case we have $Q(\mathsf{P}, \theta) = \big(\mathsf{E}_{\mathsf{P}}(h(Z_1, \theta)), \ldots, \mathsf{E}_{\mathsf{P}}(h(Z_n, \theta))\big)'$ for some function h, and we define:

$$\mathscr{P}(\theta_0) \equiv \{\mathsf{P} : \uptheta(\mathsf{P}) = \theta_0\} = \{\mathsf{P} : \mathsf{E}_{\mathsf{P}}(h(Z_i, \theta_0)) = 0, \, \forall\, i = 1, \ldots, n\}.$$

Hence any test about $\uptheta$ is equivalent to test an hypothesis about the expectation of h. As an immediate consequence, for any given θ_0 the following two testing problems $H = (H_0 : \uptheta(\overline{\mathsf{P}}) = \theta_0; H_1 : \uptheta(\overline{\mathsf{P}}) \neq \theta_0)$ and $\tilde{H} = (\tilde{H}_0 : \mathsf{E}_{\mathsf{P}}(h(Z_i, \theta_0)) = 0, \forall\, i; \tilde{H}_1 : \mathsf{E}_{\mathsf{P}}(h(Z_i, \theta_0)) \neq 0, \text{ for some } i)$ are the same. A test about $\uptheta$ amounts to a test about an expectation. If BS result applies to the expectation of the variables $h(Z_1, \theta), \ldots, h(Z_n, \theta)$, for any θ, the parameter of interest is non-testable.

A usual practice in GMM framework [see Hansen (1982)] is to avoid parametric assumptions on $h(Z_1, \theta), \ldots, h(Z_n, \theta)$. Thus non-testability problems are bound to arise. However notice that, by the above arguments, if $\tilde{H}$ turns out to be testable, then so is H. In such a case, the parameter of interest is testable. This could be achieved for instance, if h is chosen so that $h(Z_1, \theta), \ldots, h(Z_n, \theta)$ are symmetrical random variables when $\theta = \theta_0$ while they are not for another value of the parameter. Another way is to choose h with bounded range, the bounds being known. We may then apply Anderson (1967)'s technique. Notice that the availability of this technique combined with Theorem 3.1, yields a result on the impossibility of characterizing the expectation of a random variable by some moment conditions.

Proposition 7.1 (NON-TESTABILITY OF STRUCTURAL COEFFICIENTS DEFINED BY NON-TESTABLE MOMENTS).*The expectation of a n-tuple of i.i.d. real random variables $(Z_1, \ldots, Z_n)$ cannot be characterized by the moment conditions*

$$\mathsf{E}_{\mathsf{P}}\big(h(Z_i, \theta_0)\big) = 0, \forall i = 1, \ldots, n \iff \mathsf{E}_{\mathsf{P}}(Z_i) = \theta_0, \forall i = 1, \ldots, n\,,$$

where h is a function with a known bounded range.

Another route to testability is to work with conditional moment conditions with a particular form of heterogeneity. Notice this is why the slope parameter of a linear regression model is testable (see Proposition 6.2) while the regression coefficients are not (see Proposition 5.2).

7.1 A semiparametric nonlinear regression model

Let g be some known mapping from $\mathscr{X} \times \Theta$ to $\mathbb{R}$, where $\mathscr{X} \subseteq \mathbb{R}^p$. Our model is now the set $\mathscr{P}$ of all probability distributions P for $Z = (X_1', \ldots, X_n', Y_1, \ldots, Y_n)' \in \mathscr{X}^n \times \mathscr{Y}^n$, where $\mathscr{Y} \subseteq \mathbb{R}$, such that the following conditions hold:

C6. $\mathsf{E}_\mathsf{P}(Y_i - g(X_i, \theta)|X) = 0$, $i = 1, \ldots, n$, for some unique $\theta \in \mathbb{R}^K$ a.s.

The model is therefore defined by the moment conditions C6 above and the parameter is defined by $\theta(\mathsf{P}) = \theta \iff \mathsf{E}_\mathsf{P}(Y_i|X) = g(X_i, \theta)$, $i = 1, \ldots, n$.

Proposition 7.2 (NON-TESTABILITY IN SEMIPARAMETRIC NONLINEAR REGRESSIONS). *In the model defined by condition C6, the parameter* θ *is non-testable.*

As the model defined here is a nonlinear version of that of Section 6.2 where we have Proposition 6.3, this result should come as no surprise. We also know from section 6 that putting more constraints on the model is one way to make θ testable. We add to C6 the following condition:

C7. Conditionally on X, the error terms $Y_i - \mathsf{E}_\mathsf{P}(Y_i|X)$, $i = 1, \ldots, n$, are i.i.d.

Our following results show that under C6 and C7, some hypotheses about θ are testable and some are not. We first define the binary relation $\mathcal{R}$ on $\mathbb{R}^K \times \mathbb{R}^K$ by

$$\theta_1 \mathcal{R} \theta_2 \iff \exists \mu \in \mathbb{R} \text{ s.t. } g(x, \theta_1) = g(x, \theta_2) + \mu, \forall x \tag{7.1}$$

It is easy to check that $\mathcal{R}$ is reflexive, symmetric and transitive. Therefore, $\mathcal{R}$ is an equivalence relation and for any $\theta^* \in \mathbb{R}^K$ we define its equivalence class $\mathscr{E}(\theta^*)$ as $\mathscr{E}(\theta^*) = \{\theta \in \mathbb{R}^K : \theta^* \mathcal{R} \theta\}$.

Proposition 7.3 (PARTIAL NON-TESTABILITY IN SEMIPARAMETRIC NONLINEAR REGRESSIONS WITH I.I.D. ERROR TERMS). *Consider* θ_0 *such that* $\mathscr{E}(\theta_0) \neq \{\theta_0\}$. *Then in the model defined by C6 and C7,* $H_0(\theta_0) : \overline{\theta} = \theta_0$ *is partially non-testable against* $H_1(\theta_1) : \overline{\theta} = \theta_1$, *for any* $\theta_0 \neq \theta_1 \in \mathscr{E}(\theta_0)$.

Partial non-testability of $H_0 : \overline{\theta} = \theta_0$ comes from alternatives where $\overline{\theta}$ belong to the same equivalence class as θ_0. Under *any* of these alternatives, $g(X, \overline{\theta}) - g(X, \theta_0)$ remains constant when X varies. Therefore, when we restrict the model to this equivalence class, it is parameterized only by the parameter $\mu \equiv g(X, \theta) - g(X, \theta_0)$, $\theta \in \mathscr{E}(\theta_0)$. In this restricted model, the testing problem translates into hypotheses about the value of the *constant* conditional expectation $\mu = \mathsf{E}(Y - g(X, \theta_0)|X)$. Proposition 6.1 then applies.

Consider the particular case where the conditional expectation has an "intercept" term, *i.e.*, $g(x, \theta) = \mu + h(x, \beta)$ for some $\mu \in \mathbb{R}$ and $\beta \in \mathbb{R}^{K-1}$. Condition C6 together with the definition of the binary relation $\mathcal{R}$ imply that $(\mu_1, \beta_1')'\mathcal{R}(\mu_2, \beta_2')'$ iff $\beta_1 = \beta_2$. In this case, Proposition 7.3 implies the partial non-testability of the whole parameter $\theta = (\mu, \beta')'$. The presence of an "intercept" in a nonlinear model makes the whole parameter partially non-testable.

We know non-testability problems originate from equivalent classes $\mathscr{E}_\ell \subseteq \Theta$, $\ell \in L$, generated the relation $\mathcal{R}$ defined in (7.1). The family of all equivalence classes defines a partition of the model $\bigcup_{\ell \in L} \mathscr{P}_\ell$ where $\mathscr{P}_\ell = \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) \in \mathscr{E}_\ell\}$. According to the above remarks and using a device as in the proof of Proposition 6.2, one might expect that it is possible to construct an α-level test, with power somewhere above α that discriminate equivalence classes of θ.We show this is indeed possible under conditions C6, and condition C8 below, which is obviously weaker that C7:

C8. Conditionally on X, the error terms $Y_i - \mathsf{E}_{\mathsf{P}}(Y_i|X)$, $i = 1, \ldots, n$, are identically distributed.

Proposition 7.4 (SUFFICIENT CONDITION FOR TESTABILITY IN SEMIPARAMETRIC NONLINEAR REGRESSIONS). *Consider the model defined by conditions C6 and C8. For some $\ell \in L$, one wishes to test $H_0(\ell) : \overline{\mathsf{P}} \in \mathscr{P}_\ell$ against $H_1(\ell) : \overline{\mathsf{P}} \notin \mathscr{P}_\ell$. Let $n_{X,\ell} = \max_{\theta \notin \mathscr{E}_\ell} \#\{g(X_1, \theta) - g(X_1, \theta_\ell), \ldots, g(X_n, \theta) - g(X_n, \theta_\ell)\}$ where θ_ℓ is any element of $\mathscr{E}_\ell$. H_0 is testable against H_1 at any level $1 > \alpha \geq \frac{1}{n_{X,\ell}}$.*

It seems difficult to derive further results on testability of certain hypotheses about general functions of θ without specifying the regression function g. Consider the following example

$$g(X, \theta) = \lambda \exp(\gamma X)$$

where $\theta = (\lambda, \gamma)' \in \mathbb{R}^2$. It easily checked that $\mathscr{E}\big((\gamma, \lambda)'\big) = \{(\gamma, \lambda)'\}$ whenever $\lambda\gamma \neq 0$ and that otherwise $\mathscr{E}\big((0, \gamma)'\big) = \{0\} \times \mathbb{R}$ and $\mathscr{E}\big((\lambda, 0)'\big) = \mathbb{R} \times \{0\}$. For given λ_0 and

γ_0, consider the testing problem $H_0 : (\overline{\lambda}, \overline{\gamma})' = (\lambda_0, \gamma_0)'$ against $H_1 : (\overline{\lambda}, \overline{\gamma})' \neq (\gamma_0, \lambda_0)'$.

When $\lambda_0 \neq 0$ and $\gamma_0 = 0$, for any given $\mu \in \mathbb{R}$, we can find $\theta_1 = (\mu + \lambda_0, 0)' \in \mathscr{E}\big((\gamma_0, \lambda_0)'\big)$ such that $g(X, \theta_1) = \mu + g(X, \theta_0)$. In other words, Proposition 7.3 applies: $H_0 : (\overline{\lambda}, \overline{\gamma})' = (\lambda_0, 0)'$ is partially non-testable against $H_1 : (\overline{\lambda}, \overline{\gamma})' \neq (\lambda_0, 0)'$. This holds for any $\lambda_0 \neq 0$.

Next consider the case where $\lambda_0\gamma_0 \neq 0$. As $\mathscr{E}\big((\gamma_0, \lambda_0)'\big) = \{(\gamma_0, \lambda_0)'\}$, the null H_0 is obviously equivalent to $H_0 : \overline{\theta} \in \mathscr{E}\big((\gamma_0, \lambda_0)'\big)$. From Proposition 7.4, we know H_0 is testable.

7.2 Nonparametric regression models

In the previous sections, we addressed testability issues in regression models where the regression function in known up to a finite dimensional parameter θ. We know investigate the same issues when the regression function is totally unknown.

Formally, the statistical model we consider is a family $\mathscr{P}$ of probability distributions for a real random $(nK + n)$-vector $(X', Y')' = (X_1', \ldots, X_n', Y_1, \ldots, Y_n)$. Let Θ be the set of all Borel mappings from $\mathbb{R}^K$ into $\mathbb{R}$. The family $\mathscr{P}$ is defined by the following condition. A probability distribution P on $\mathbb{R}^{nK+n}$ is an element of $\mathscr{P}$ if and only if $(X', Y')' \sim \mathsf{P}$ implies condition C3 and

C9. $\forall i = 1, \ldots, n$, $\mathsf{E}_{\mathsf{P}}(Y_i|X)$ exists.

As usual, for $\mathsf{P} \in \mathscr{P}$ and $\theta \in \Theta$, we write $\uptheta(\mathsf{P}) = \theta$ iff $\mathsf{E}_{\mathsf{P}}(Y_i|X) = \theta(X_i)$, $i = 1, \ldots, n$. Condition C3 is a usual maintained assumption in nonparametric regression estimation [see for instance Härdle and Linton (1994)]. We are interested in testing $H_0(\theta_0) : \uptheta(\overline{\mathsf{P}}) = \theta_0$ against $H_1(\theta_0) : \uptheta(\overline{\mathsf{P}}) \neq \theta_0$, where θ_0 is some known element of Θ.

Proposition 7.5 (NON-TESTABILITY OF HYPOTHESIS PAIRS IN NONPARAMETRIC REGRESSIONS). *Consider the family $\mathscr{P}$ defined by C3 and C9. Define $\mathscr{P}(\theta_0, \theta_1) \equiv \{\mathsf{P} \in \mathscr{P} : \uptheta(\mathsf{P}) = \theta_0$ or $\uptheta(\mathsf{P}) = \theta_1\}$, where θ_0 and θ_1 are arbitrary distinct elements of Θ. Then $H_0 : \uptheta(\overline{\mathsf{P}}) = \theta_0$ is not testable against $H_1 : \uptheta(\overline{\mathsf{P}}) = \theta_1$ in $\mathscr{P}(\theta_0, \theta_1)$.*

We therefore have the following immediate corollary.

Corollary 7.6 (PARAMETER NON-TESTABILITY IN NONPARAMETRIC REGRESSIONS). *In the model $\mathscr{P}$ defined by C3 and C9, the parameter $\uptheta$ is non-testable.*

Although the assumption C3 is common, it is obviously not necessary for obtaining the non-testability result of Proposition 7.5 and its Corollary.

In the nonparametric regression literature, the issue of "significance testing" has received a lot of attention [see *e.g.*, Matzkin (1994), Fan and Li (1996), Lavergne and Vuong (2000) and Aït-Sahalia et al. (2001)]. The problem arises when the regressors are separated as $X_i = (W_i', Z_i')'$, where W_i and Z_i are random real K_W- and K_Z-vectors, respectively, with $K_W + K_Z = K$, $i = 1, \ldots, n$. One wishes to test $H_0(W) : \overline{\mathsf{P}} \in \mathscr{P}_W$ against $H_1(W) : \overline{\mathsf{P}} \notin \mathscr{P}_W$, where $\mathscr{P}_W \equiv \{\mathsf{P} \in \mathscr{P} : \mathsf{E}_{\mathsf{P}}(Y_i|X) = \mathsf{E}_{\mathsf{P}}(Y_i|W), \mathsf{P}\text{-a.s.}\}$. We have the following result.

Proposition 7.7 (NON-TESTABILITY OF PARTIAL HYPOTHESIS IN NONPARAMETRIC REGRESSIONS). *In model $\mathscr{P}$ defined by C3 and C9, $H_0(W)$ is non-testable against $H_1(W)$.*

Various results may be derived along the same lines. For instance, shape restrictions on $\mathsf{E}_{\mathsf{P}}(Y_1|X)$ (as a function of X), such as monotonicity or concavity, may be imposed. As this kind of restrictions arise naturally in economic contexts, constrained nonparametric testing has been extensively studied (see *e.g.*, Matzkin (1994)). Such restrictions cannot entail testability (see for instance the proof of Proposition 7.5).

8 Conclusion

As first recognized by Bahadur and Savage (1956) some pairs of hypotheses (H_0, H_1) are non-testable in the sense that a test procedure that makes no use of the data is UMP.

This paper investigates this result extends to models and settings of more relevance to the econometrician. We establish that when non-testability prevails, asymptotic approximations based on central limit theorem provide misleading insights. In particular we show that consistency and control of the level are incompatible. Moreover, pointwise convergence of the type 1 risk of somewhere consistent test is arbitrarily slow. This applies in particular to corrections (such as those proposed by Bartlett or bootstrap) made in order to lower the gap between targeted and actual level of the test.

In the case of a parameter, we link non-testability to a lack of continuity of the mapping between the set of probability distributions and the set of admissible values of the parameter. We show that several parameters linked to moments conditions are non-testable.

Non-testability of a parameter is a property of both the mapping defining the parameter and the set of probability distributions on which it is defined. To recover testability of the parameter, the econometrician may then either change the mapping (by considering for instance inference about the median rather than the expectation) or change the domain of the mapping (for instance by imposing more conditions on the set of distributions).

In the context of regression models, results are twofold. When errors are assumed identically distributed, parts of the parameter (e.g. the slope coefficients in a linear regression model) are testable whereas testability is typically lost for heterogeneous error terms. In particular the slope parameter of a linear regression model is non-testable if it is only assumed that the error terms are independent and homoskedastic.

Although many of these results apply to cross-section, times series and panel data, it is clear that further reasearch is needed to assert the full range of (non-)testability concepts in econometric models. This is in particular the case in times series models where the same parameter may be linked to first- and second-order moments conditions.

ACKNOWLEDGEMENTS: The authors thank seminar participants at the 2004 European Meeting of the Econometric Society (Madrid, 2004), the Maastricht School of Economics and Organizations, the Workshop on "New Trouble for Standard Regression Analysis" (Universität Regensburg, 2005), McGill University, and the CIREQ Conference on Generalized Method of Moments (2007), for several useful comments and discussions. This work was supported by the Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de Finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on Mathematics of Information Technology and Complex Systems (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), the Fonds de recherche sur la nature et les technologies (Québec), and a Killam Fellowship (Canada Council for the Arts).

Appendix

A Proofs

Proof of Proposition 2.4 Let C be a $1-\alpha$ level confidence region for θ, *i.e.*, $\mathsf{P}(\theta(\mathsf{P}) \in C) \geq 1-\alpha$, for all $\mathsf{P} \in \mathscr{P}$. Choose any $\theta_0 \in \Theta^*$ and consider testing $H_0 : \overline{\theta} = \theta_0$ against $H_1 : \overline{\theta} \neq \theta_0$. The test $\varphi_n = \mathrm{I}(\theta_0 \notin C)$ has level α. As θ is non-testable at level α on Θ^*, for any P we must have $\mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \alpha$, or equivalently, $\mathsf{P}(\theta_0 \notin C) \leq \alpha$.
QED

Proof of Proposition 2.5 Consider an α-level test φ for (H_0, H_1). The procedure φ has also level α for $(\tilde{H}_0, \tilde{H}_1)$. Now since $(\tilde{H}_0, \tilde{H}_1)$ is partially non-testable, there exists $\hat{\mathscr{P}}_1 \subseteq \tilde{\mathscr{P}}_1 \subseteq \mathscr{P}_1$ such that $\sup_{\mathsf{P}\in\hat{\mathscr{P}}_1} \mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \alpha$. QED

Proof of Proposition 2.6 Assume $(H_0, \tilde{H}_1)$ is testable at level α and $(\tilde{H}_0, H_1)$ is non-testable. Consider then an α-level test φ for $(H_0, \tilde{H}_1)$. As $\tilde{\mathscr{P}}_0 \subseteq \mathscr{P}_0$, φ is α-level for $(\tilde{H}_0, H_1)$. Now as we assumed $(H_0, \tilde{H}_1)$ testable, we may find $\tilde{\mathsf{P}}_1 \in \tilde{\mathscr{P}}_1$ such that $\mathsf{E}_{\tilde{\mathsf{P}}_1}(\varphi_n) > \alpha$. But $\tilde{\mathscr{P}}_1 \subseteq \mathscr{P}_1$ implies $\tilde{\mathsf{P}}_1 \in \mathscr{P}_1$. Hence $(\tilde{H}_0, H_1)$ is testable, a contradiction. QED

Proof of Proposition 2.7 Let φ be an α-level test for $(\tilde{H}_0, \tilde{H}_1)$. If $\sup_{\mathsf{P}\in\mathscr{P}_0} \mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \alpha$, then non-testability of (H_0, H_1) implies $\sup_{\mathsf{P}\in\tilde{\mathscr{P}}_1} \mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \sup_{\mathsf{P}\in\mathscr{P}_1} \mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \alpha$. Thus φ is dominated by $\varphi^*(\alpha)$ for testing $(\tilde{H}_0, \tilde{H}_1)$. QED

Proof of Proposition 2.8 Let λ_0 be any element of Λ and consider the following testing problem $H_0(\lambda_0) : \lambda(\overline{\mathsf{P}}) = \lambda_0$ against $H_1(\lambda_0) : \lambda(\overline{\mathsf{P}}) \neq \lambda_0$. This amounts to testing $H_0 : \theta(\overline{\mathsf{P}}) \in g^{-1}(\lambda_0)$ against $H_1 : \theta(\overline{\mathsf{P}}) \notin g^{-1}(\lambda_0)$ which by the non-testability of θ is a non-testable problem. QED

Proof of Proposition 2.9 Fix $\alpha \in]0,1[$. Assume $H_0(\gamma_0) : \overline{\gamma} = \gamma_0$ is testable against $H_1(\gamma_0) : \overline{\gamma} \neq \gamma_0$. Then there exists $\psi_1 \in \Psi$ and $\mathsf{P}_1 \in \mathscr{P}$ with $\gamma(\mathsf{P}_1) = \gamma_1 \neq \gamma_0$ and $\psi(\mathsf{P}_1) = \psi_1$ such that $\mathsf{E}_{\mathsf{P}_1}(\varphi_n) > \alpha$ for some α-level test φ of $H_0(\gamma_0)$. This test is also α-level for testing $H_0^{\psi_1}(\gamma_0) : \overline{\theta} = (\gamma_0, \psi_1)$ against $H_1^{\psi_1}(\gamma_0) : \overline{\theta} = (\gamma_1, \psi_1)$ As θ is non-testable, for all distribution $\tilde{\mathsf{P}}_1$ with $\theta(\tilde{\mathsf{P}}_1) \neq (\gamma_0, \psi_1)$, we have $\mathsf{E}_{\tilde{\mathsf{P}}_1}(\varphi_n) \leq \alpha$. But P_1 is such a distribution. We thus have a contradiction. As this is true for any $\alpha \in]0,1[$ and any α-level test φ, $H_0(\gamma_0) : \overline{\gamma} = \gamma_0$ is non-testable against $H_1(\gamma_0) : \overline{\gamma} \neq \gamma_0$. This is true for any $\gamma_0 \in \Gamma$ and thus γ is non-testable. QED

Proof of Proposition 2.10

(1) Choose $\gamma_0 \in \Gamma$, $\psi_0 \in \Psi$, and define $\mathscr{P}_0 \equiv \{\mathsf{P} \in \mathscr{P} : \boldsymbol{\gamma}(\mathsf{P}) = \gamma_0\}$ and $\mathscr{P}^{\psi_0} \equiv \{\mathsf{P} \in \mathscr{P} : \boldsymbol{\psi}(\mathsf{P}) = \psi_0\}$. Then apply Lemma B.5 (see Appendix B) with $I = \Psi$. As this is true whatever the choice of γ_0, use Definition 2.3 to conclude.

(2) Consider $\theta_0 = (\gamma_0, \psi_0) \in \Theta$, and $H_0(\theta_0) : \overline{\theta} = \theta_0$ and $H_1(\theta_0) : \overline{\theta} \neq \theta_0$. If φ is an α-level test of $H_0(\theta_0)$, then it has level α for $H_0^{\psi_0}(\gamma_0)$. But as it is assumed that $H_0^{\psi_0}(\gamma_0)$ is non-testable against $H_1^{\psi_0}(\gamma_0)$, we must have $\sup_{\mathsf{P}\in\mathscr{P}_1\cap\mathscr{P}^{\psi_0}} \mathsf{E}_\mathsf{P}(\varphi_n) \leq \alpha$. But for any $\mathsf{P} \in \mathscr{P}_1 \cap \mathscr{P}^{\psi_0}$, the alternative $H_1(\theta_0)$ is true. Therefore, $H_0(\theta_0)$ is partially non-testable against $H_1(\theta_0)$. As this holds for any θ_0, $\boldsymbol{\theta}$ is partially non-testable. QED

Proof of Theorem 3.1 Choose and fix $N \in \{1, 2, \ldots\}$. For any P and Q in $\mathscr{P}$ define the pseudo distance $D(\mathsf{P}, \mathsf{Q}) \equiv \sup_{g\in\mathcal{B}_{Nd}} \left|\mathsf{E}_\mathsf{P}(g_N) - \mathsf{E}_\mathsf{Q}(g_N)\right|$. Take $f \in \mathcal{B}_{Nd}$, $\epsilon \in]0, +\infty[$, $\theta_0 \in \Theta$ and $\theta_1 \in \Theta$. Under BSE1, we may always find a P_0 in $\mathscr{P}(\theta_0)$. Take any $\pi \in]0, 1[$ such that $1 - \pi^N < \epsilon$. Under BSE2, we can find $\tilde{\mathsf{P}} \in \mathscr{P}$ such that $\mathsf{P}_1 \equiv \pi\mathsf{P}_0 + (1 - \pi)\tilde{\mathsf{P}}$ is in $\mathscr{P}(\theta_1)$. From Lemma B.2 (see the Appendix) we have $D(\mathsf{P}_0, \mathsf{P}_1) \leq 1 - \pi^N$, which, together with $1 - \pi^N < \epsilon$, implies $D(\mathsf{P}_0, \mathsf{P}_1) < \epsilon$. As this inequality is true for any $\epsilon \in]0, +\infty[$ and any $\mathsf{P}_0 \in \mathscr{P}(\theta_0)$, it shows that for any element P_0 of $\mathscr{P}(\theta_0)$, we can find an element of $\mathscr{P}(\theta_1)$ which is arbitrarily close to P_0 according to D. As this property holds for any $\theta_0 \in \Theta$ and any $\theta_1 \in \Theta$, we can conclude, using BSE1, that for any $\mathsf{P} \in \mathscr{P}$ and any $\theta \in \Theta$, we can find a P_0 in $\mathscr{P}(\theta)$ which is arbitrarily close to P. Therefore, for any $\theta \in \Theta$, $\mathscr{P}(\theta)$ is dense in $\mathscr{P}$ w.r.t. D. The function $\Lambda : \mathscr{P} \longrightarrow [0, 1]$ defined by $\Lambda(\mathsf{P}) = \mathsf{E}_\mathsf{P}(f_N)$ is obviously continuous w.r.t. the pseudo-distance D (see Lemma B.4). Indeed, fix $\epsilon \in]0, +\infty[$ and $\mathsf{P} \in \mathscr{P}$. If Q is such that $D(\mathsf{P}, \mathsf{Q}) < \epsilon$, then clearly $|\Lambda(\mathsf{P}) - \Lambda(\mathsf{Q})| = |\mathsf{E}_\mathsf{P}(f_N) - \mathsf{E}_\mathsf{Q}(f_N)| < \epsilon$. A straightforward application of Lemma B.4 with $A = \mathscr{P}(\theta)$, $B = \mathscr{P}$ and $h = \Lambda$ concludes the proof. QED

Proof of Corollary 3.2 Define $\mathscr{P}(\Theta_j) \equiv \{\mathsf{P} \in \mathscr{P} : \boldsymbol{\theta}(\mathsf{P}) \in \Theta_j\}$, $j = 0, 1$. We have for any $\theta \in \Theta_j$, $\sup_{\mathsf{P}\in\mathscr{P}(\theta)} \mathsf{E}_\mathsf{P}(\varphi_n) \leq \sup_{\mathsf{P}\in\mathscr{P}(\Theta_j)} \mathsf{E}_\mathsf{P}(\varphi_n) \leq \sup_{\mathsf{P}\in\mathscr{P}} \mathsf{E}_\mathsf{P}(\varphi_n)$. As $\varphi_n \in \Xi_{nq}$, Theorem 3.1 (with $N = 1$ and $d = nq$) implies that the LHS and the RHS are equal, and thus $\sup_{\mathsf{P}\in\mathscr{P}(\Theta_j)} \mathsf{E}_\mathsf{P}(\varphi_n) = \sup_{\mathsf{P}\in\mathscr{P}} \mathsf{E}_\mathsf{P}(\varphi_n)$. But this is true for $j = 0, 1$, and then

$$\sup_{\mathsf{P}\in\mathscr{P}(\Theta_1)} \mathsf{E}_\mathsf{P}(\varphi_n) = \sup_{\mathsf{P}\in\mathscr{P}(\Theta_0)} \mathsf{E}_\mathsf{P}(\varphi_n). \tag{A.1}$$

By the same argument, $\inf_{\mathsf{P}\in\mathscr{P}(\theta)} \mathsf{E}_\mathsf{P}(\varphi_n) \geq \inf_{\mathsf{P}\in\mathscr{P}(\Theta_j)} \mathsf{E}_\mathsf{P}(\varphi_n) \geq \inf_{\mathsf{P}\in\mathscr{P}} \mathsf{E}_\mathsf{P}(\varphi_n)$ and Theorem 3.1 entails

$$\inf_{\mathsf{P}\in\mathscr{P}(\Theta_1)} \mathsf{E}_\mathsf{P}(\varphi_n) = \inf_{\mathsf{P}\in\mathscr{P}(\Theta_0)} \mathsf{E}_\mathsf{P}(\varphi_n). \tag{A.2}$$

If φ has level α, the RHS of (A.1) is less than α, which proves 1. If φ is similar at level α, we have $\inf_{\mathsf{P}\in\mathscr{P}(\Theta_0)} \mathsf{E}_\mathsf{P}(\varphi_n) = \sup_{\mathsf{P}\in\mathscr{P}(\Theta_0)} \mathsf{E}_\mathsf{P}(\varphi_n) = \alpha$, which yields $\mathsf{E}_\mathsf{P}(\varphi_n) = \alpha$, $\forall\mathsf{P} \in \mathscr{P}(\Theta_0)$. If $\mathscr{P}$ is complete, this entails $\varphi_n = \alpha$, a.s.-P, $\forall\mathsf{P} \in \mathscr{P}(\Theta_0)$. Equations

(A.1) and (A.2) imply $\inf_{\mathsf{P}\in\mathscr{P}(\Theta_1)} \mathsf{E}_{\mathsf{P}}(\varphi_n) = \sup_{\mathsf{P}\in\mathscr{P}(\Theta_1)} \mathsf{E}_{\mathsf{P}}(\varphi_n) = \alpha$. Thus $\mathsf{E}_{\mathsf{P}}(\varphi_n) = \alpha$, $\forall \mathsf{P} \in \mathscr{P}(\Theta_1)$. This proves 2. If φ is not similar, we can find $\mathsf{P}_0 \in \mathscr{P}(\Theta_0)$ such that $\mathsf{E}_{\mathsf{P}_0}(\varphi_n) < \alpha$. Hence by (A.2) we have $\mathsf{E}_{\mathsf{P}_1}(\varphi_n) < \alpha$ for some $\mathsf{P}_1 \in \mathscr{P}(\Theta_1)$. Thus φ is biased. QED

Proof of Proposition 3.3 Let φ be an α-level test of $H_0 : \overline{\theta} = \theta_0$ against $H_0 : \overline{\theta} \neq \theta_0$. Choose $\mathsf{P}_1 \notin \mathscr{P}(\theta_0)$. For any $\epsilon > 0$ and any $\mathsf{P} \in \mathscr{V}_\epsilon(\mathsf{P}_1)$ we have either $\theta(\mathsf{P}) = \theta_0$ or $\theta(\mathsf{P}) \neq \theta_0$. As θ is non-testable, in both cases we have $\mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \alpha$. QED

Proof of Proposition 3.4 Let φ be an α-level test of H_0 and assume $\Theta_1^* \neq \varnothing$. Assume $\mathsf{P}_{\theta_1}\big(\partial\{X : \varphi(X) = 1\}\big) = 0$ for some $\theta_1 \in \Theta_1^*$. By definition of Θ_1^* we may find a sequence $\{\theta_{0,n} : n \geq 1\}$ in Θ_0 with limit θ_1. Under the above assumption and under assumption BSE5, we have $\mathsf{P}_{\theta_{0,n}}(\varphi = 1) \longrightarrow \mathsf{P}_{\theta_1}(\varphi = 1)$, as $n \to \infty$. But as φ has level α, $\mathsf{P}_{\theta_{0,n}}(\varphi = 1) \leq \alpha$, $\forall n$, which implies $\mathsf{P}_{\theta_1}(\varphi = 1) \leq \alpha$. Now if this is true for any $\theta_1 \in \Theta_1^*$ and if $\Theta_1^* = \Theta_1$, the same argument yields $\mathsf{P}_{\theta_1}(\varphi = 1) \leq \alpha$, $\forall \theta_1 \in \Theta_1$. QED

Proof of Proposition of 4.2 For each n, let φ_n be a test of $H_n = (H_{0,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{0,n}\,,\, H_{1,n} : \overline{\mathsf{P}}_n \in \mathscr{P}_{1,n})$.

(1) If for any n, H_n is partially non-testable, then $\forall n, \exists \mathsf{P}_{1,n} \in \mathscr{P}_{1,n}$ such that

$$\mathsf{E}_{\mathsf{P}_{1,n}}(\varphi_n) \leq \sup_{\mathsf{P}_n\in\mathscr{P}_{0,n}} \mathsf{E}_{\mathsf{P}_n}(\varphi_n). \tag{A.3}$$

If φ_n is asymptotically α-level, $\alpha \in]0,1[$, then the $\limsup$ of the RHS of (A.3) is $\alpha < 1$ and φ_n is not everywhere consistent.

(2) If for any n, H_n is non-testable, then (A.3) holds for any n and any $\mathsf{P}_{1,n} \in \mathscr{P}_{1,n}$. If φ_n is somewhere consistent, then there exists a sequence $\{\mathsf{P}_{1,n} : n \geq 1\}$ with $\mathsf{P}_{1,n} \in \mathscr{P}_{1,n}\ \forall n$, such that the $\liminf$ of the LHS of (A.3) converges to 1. Thus φ_n has level 1 asymptotically. QED

Proof of Corollary 4.3

(1) Assume the sequence of tests $\{\varphi_n : n \geq 1\}$ is somewhere consistent. As H_n is non-testable for all n, we conclude from Proposition 4.2 that

$$\liminf_{n\to\infty} \sup_{\mathsf{P}_{0,n}\in\mathscr{P}_{0,n}} \mathsf{E}_{\mathsf{P}_{0,n}}(\varphi_n) = 1 = \lim_{n\to\infty} \sup_{\mathsf{P}_{0,n}\in\mathscr{P}_{0,n}} \mathsf{E}_{\mathsf{P}_{0,n}}(\varphi_n).$$

(2) Assume the sequence of type-1 risks associated with $\{\varphi_n : n \geq 1\}$ converges uniformly to $\alpha < 1$, while $\liminf_{n\to\infty} \mathsf{E}_{\mathsf{P}_{1,n}}(\varphi_n) = 1$, for any sequence $\{\mathsf{P}_{1,n} : n \geq 1\}$ such that $\mathsf{P}_{1,n} \in \mathscr{P}_{1,n}$ for all $n \geq 1$. According to Proposition 4.2, this is a contradiction when H_n is non-partially testable for all n. QED

Proof of Proposition 5.1 Denote $\Theta = \mathcal{M}_{p,q}$, the set of (p,q) matrices with real entries. We must show that for any $\theta_0 \in \mathcal{M}_{p,q}$, $H(\theta_0) = \big(H_0(\theta_0), H_1(\theta_0)\big)$ is non-testable, where $H_0(\theta_0) : \overline{\mathsf{P}} \in \mathscr{P}(\theta_0)$, $H_1(\theta_0) : \overline{\mathsf{P}} \in \mathscr{P} \setminus \mathscr{P}(\theta_0)$, and $\mathscr{P}(\theta_0) \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) = \theta_0\}$. Fix any $\mu_0 \in \mathbb{R}^{p+q}$ and define $\mathscr{P}^{\mu_0} \equiv \{\mathsf{P} \in \mathscr{P} : \mathsf{E}_{\mathsf{P}}(Z) = \mu_0\}$. We first show that BSE1 and BSE6 to BSE8 (see Lemma B.5) hold for $\mathscr{P}^{\mu_0}$ and $\theta(\mathsf{P}) = \mathsf{Cov}_{\mathsf{P}}(X, Y)$. To show BSE1, we show that for any any given $\theta \in \mathcal{M}_{p,q}$ the family $\mathscr{P}^{\mu_0}$ contains the $(p+q)$ dimensional normal distribution $\mathcal{N}(\mu_0, V)$, where

$$V \equiv \begin{pmatrix} \sigma^2 I_p & \theta \\ \theta' & \sigma^2 I_q \end{pmatrix},$$

for some $\sigma \in \,]0, \infty[$ such that V is positive semidefinite. The matrix V is positive semidefinite iff for all $a \in \mathbb{R}^{p+q}$ with $\|a\| = 1$ we have $\sigma^2 + a'\Lambda a \geq 0$, where $\|\cdot\|$ denotes the Euclidian norm and Λ is the $(p+q, p+q)$ matrix defined by

$$\Lambda \equiv \begin{pmatrix} 0 & \theta \\ \theta' & 0 \end{pmatrix}.$$

We may look for the element a^* of $\mathbb{R}^{p+q}$, with $\|a^*\| = 1$, for which $a'\Lambda a$ is minimized. This is the Rayleigh-Ritz problem whose solution a^* is the eigenvector of Λ associated with the smallest eigenvalue of Λ, which we denote λ^* [see Horn and Johnson (1985, Theorem 4.2.2)]. We then have $a^{*\prime}\Lambda a^* = \lambda^*$. Thus if we choose $\sigma^2 = |\lambda^*| + \epsilon$ for some $\epsilon > 0$, the matrix

$$\begin{pmatrix} \sigma^2 I_p & \theta \\ \theta' & \sigma^2 I_q \end{pmatrix}$$

is positive semidefinite. This shows BSE1 holds in $\mathscr{P}^{\mu_0}$. BSE7 is obviously true. We next show BSE6 and BSE8 are also true. Let P_1 and P_2 be two distributions in $\mathscr{P}^{\mu_0}$. Take $\pi \in \,]0, 1[$ and consider $\mathsf{P} = \pi\mathsf{P}_1 + (1-\pi)\mathsf{P}_2$. Consider the random vector $(Z_1', Z_2', U)'$ where $Z_k \equiv (X_k', Y_k')'$ is distributed as P_k, $k = 1, 2$, and $U \sim \mathscr{B}(\pi)$, with that U, Z_1 and Z_2 are independent. Clearly $Z \equiv UZ_1 + (1-U)Z_2$ is distributed as P and

$$\mathsf{E}(Z) = \mathsf{E}(UZ_1) + \mathsf{E}\big((1-U)Z_2\big) = \pi\mu_0 + (1-\pi)\mu_0 = \mu_0. \tag{A.4}$$

Also $\mathsf{V}(UZ_1)$ and $\mathsf{V}\big((1-U)Z_2\big)$ exist, which entails the existence of $\mathsf{V}(Z)$. Hence $\mathsf{P} \in \mathscr{P}^{\mu_0}$ which shows BSE6 holds. Moreover, as $U(1-U) = 0$, $U^2 = U$ and $(1-U)^2 = (1-U)$ always, we have

$$\mathsf{V}(Z) = \pi\mathsf{E}(Z_1 Z_1') + (1-\pi)\mathsf{E}(Z_2 Z_2') - \mu_0\mu_0' = \pi\mathsf{V}(Z_1) + (1-\pi)\mathsf{V}(Z_2). \tag{A.5}$$

Therefore, denoting $\mu_0 = (\mu'_{0X}, \mu'_{0Y})'$ we have

$$\begin{aligned} \mathsf{Cov}(X, Y) &= \pi \mathsf{E}(X_1 Y_1') + (1-\pi)\mathsf{E}(X_2 Y_2') - \mu_{0X}\mu'_{0Y} && \text{(A.6)} \\ &= \pi \mathsf{Cov}(X_1, Y_1) + (1-\pi)\mathsf{Cov}(X_2, Y_2), && \text{(A.7)} \end{aligned}$$

which is BSE8. From Lemma B.1, we conclude that BSE1 and BSE2 hold for the parameter θ in model $\mathscr{P}^{\mu_0}$. Define the hypotheses $H_0^{\mu_0}(\theta_0) : \overline{\mathsf{P}} \in \mathscr{P}(\theta_0) \cap \mathscr{P}^{\mu_0}$ and $H_1^{\mu_0}(\theta_0) : \overline{\mathsf{P}} \in (\mathscr{P} \setminus \mathscr{P}(\theta_0)) \cap \mathscr{P}^{\mu_0}$, and the testing problem $H^{\mu_0}(\theta_0) = \big(H_0^{\mu_0}(\theta_0), H_1^{\mu_0}(\theta_0)\big)$. From above, for any $\mu_0 \in \mathbb{R}^{p+q}$, $H^{\mu_0}(\theta_0)$ is non-testable. We have $\mathscr{P} = \bigcup_{\mu_0 \in \mathbb{R}^{p+q}} \mathscr{P}^{\mu_0}$ and Lemma B.5 applies. Therefore $H(\theta_0)$ is non-testable. As this is true for any $\theta_0 \in \Theta$, we conclude that the parameter θ(P) is non-testable.

To prove this holds in the restricted model $\mathscr{P}^*$, we consider $\mathscr{P}^{*\mu_0} \equiv \mathscr{P}^* \cap \mathscr{P}^{\mu_0}$ for which we show BSE1 and BSE6 to BSE8 hold. BSE1 is true because for any $\theta \in \mathcal{M}_{p,q}$, the matrix

$$\begin{pmatrix} \sigma^2 I_p & \theta \\ \theta' & \sigma^2 I_q \end{pmatrix},$$

with $\sigma^2 > |\lambda^*|$ is real, symmetric and positive definite, where λ^* is defined as above. Thus it is invertible. BSE7 is clearly satisfied. Equality (A.4) also holds in $\mathscr{P}^{*\mu_0}$ BSE6 holds. Next consider equation (A.5). If $\mathsf{V}(Z_1)$ and $\mathsf{V}(Z_2)$ are invertible, they are positive definite. Invertibility of $\pi\mathsf{V}(Z_1) + (1-\pi)\mathsf{V}(Z_2)$, for $\pi \in \,]0, 1[$, follows from theorem 22 of Magnus and Neudecker (1988). Thus BSE8 holds. We conclude using Lemma B.5 as in the first part of the proof. QED

Proof of Proposition 5.2 Consider the testing problem $H_0(0) : \theta(\overline{\mathsf{P}}) = 0$ against $H_1(0) : \theta(\overline{\mathsf{P}}) \neq 0$. Clearly, this is equivalent to testing $\mathsf{Cov}_{\mathsf{P}}(Y, X) = 0$ against $\mathsf{Cov}_{\mathsf{P}}(Y, X) \neq 0$. From Proposition 5.1, this problem is non-testable. Consider a given $\theta_0 \in \mathcal{M}_{p,q}$ and define $Z \equiv Y - X\theta_0$. The regression coefficients of Z on X are given by

$$\mathsf{Cov}_{\mathsf{P}}(Z, X)\mathsf{V}_{\mathsf{P}}(X)^{-1} = \big[\mathsf{Cov}_{\mathsf{P}}(Y, X) - \theta_0 \mathsf{V}_{\mathsf{P}}(X)\big]\mathsf{V}_{\mathsf{P}}(X)^{-1} = \mathsf{Cov}_{\mathsf{P}}(Y, X)\mathsf{V}_{\mathsf{P}}(X)^{-1} - \theta_0.$$

Thus testing $H_0(\theta_0) : \theta(\overline{\mathsf{P}}) = \theta_0$ against $H_1(\theta_0) : \theta(\overline{\mathsf{P}}) \neq \theta_0$ amounts to testing $\mathsf{Cov}_{\mathsf{P}}(Y - X\theta_0, X) = 0$ against $\mathsf{Cov}_{\mathsf{P}}(Y - X\theta_0, X) \neq 0$, which is non-testable. QED

Proof of Proposition 5.3 Fix a $\theta^* \in \Theta = \,]0, +\infty[$. Define $\mathscr{P}^{\mu_0} \equiv \{\mathsf{P} \in \mathscr{P} : \mathsf{E}_{\mathsf{P}}(X) = \mu_0\}$, $\mathscr{P}_0(\theta^*) \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) \geq \theta^*\}$ and $\mathscr{P}_1(\theta^*) \equiv \mathscr{P} \setminus \mathscr{P}_0(\theta^*)$. Also define $\mathscr{P}_k^{\mu_0}(\theta^*) \equiv \mathscr{P}_k(\theta^*) \cap \mathscr{P}^{\mu_0}$, $k = 0, 1$. Notice that for any $\lambda \in [0, 1]$ and any P and Q in $\mathscr{P}^{\mu_0}$ we have $\theta\big(\lambda\mathsf{P} + (1-\lambda)\mathsf{Q}\big) = \lambda\theta(\mathsf{P}) + (1-\lambda)\theta(\mathsf{Q})$ [see (A.4), (A.5) and (A.7)]. Now choose and fix $\epsilon \in \,]0, +\infty[$, $\theta_0 \in [\theta^*, +\infty[$ and $\mathsf{P}_1 \in \mathscr{P}_1^{\mu_0}(\theta^*)$. Denote $\theta_1 \equiv \theta(\mathsf{P}_1)$. Take a $\pi \in]0, 1[$ and $k \geq n$ such that $1 - \pi^k < \epsilon$. Consider the real number $\tilde{\theta} \equiv \theta_0 \frac{1}{1-\pi} - \frac{\pi}{1-\pi}\theta_1$. Notice we have $\theta_1 < \theta^* \leq \theta_0$, and as $\pi \in \,]0, 1[$, $\tilde{\theta} \in \Theta$. Therefore, we can find a $\tilde{\mathsf{P}} \in \mathscr{P}^{\mu_0}$ such that $\theta(\tilde{\mathsf{P}}) = \tilde{\theta}$ [take for instance $\tilde{\mathsf{P}} = \mathcal{N}(\mu_0, \tilde{\theta})$]. Consider the distribution P_0 defined by $\mathsf{P}_0 = \pi\mathsf{P}_1 + (1-\pi)\tilde{\mathsf{P}}$. We have

$\theta(\mathsf{P}_0) = \pi\theta(\mathsf{P}_1) + (1-\pi)\theta(\tilde{\mathsf{P}}) = \pi\theta_1 + (1-\pi)\tilde{\theta} = \theta_0$. Moreover $\mathsf{E}_{\mathsf{P}_0}(X) = \mu_0$. Hence $\mathsf{P}_0 \in \mathscr{P}_0^{\mu_0}(\theta^*)$. Now we may apply Lemma B.2 and we get $D(\mathsf{P}_1, \mathsf{P}_0) < \epsilon$. As this is true for any choice of ϵ and $\mathsf{P}_1 \in \mathscr{P}_1^{\mu_0}(\theta^*)$, it implies that $\mathscr{P}_0^{\mu_0}(\theta^*)$ is dense in $\mathscr{P}_1^{\mu_0}(\theta^*)$. Then $\forall \mathsf{P}_1 \in \mathscr{P}_1^{\mu_0}(\theta^*)$, we may find a sequence $\{\mathsf{P}_{0,m}, m \geq 1\}$ of $\mathscr{P}_0^{\mu_0}(\theta^*)$ such that $D(\mathsf{P}_{0,m}, \mathsf{P}_1) \to 0$, $m \to \infty$. As for any test φ, $\mathsf{E}_{\mathsf{P}}(\varphi_n)$ is a continuous function of P (see the proof of Theorem 3.1), we have $\mathsf{E}_{\mathsf{P}_{0,m}}(\varphi_n) \to \mathsf{E}_{\mathsf{P}_1}(\varphi_n)$, $m \to \infty$. But if φ has level α, we must have $\mathsf{E}_{\mathsf{P}_{0,m}}(\varphi_n) \leq \alpha$, $\forall m \geq 1$. Taking the limit when $m \to \infty$ yields $\mathsf{E}_{\mathsf{P}_1}(\varphi_n) \leq \alpha$. As this holds for any $\mathsf{P}_1 \in \mathscr{P}_1^{\mu_0}(\theta^*)$, we have $\sup_{\mathsf{P} \in \mathscr{P}_1(\theta^*)} \mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \alpha$. This also holds for any α-level test φ and thus $H_0^{\mu_0}(\theta^*) : \overline{\mathsf{P}} \in \mathscr{P}_0^{\mu_0}(\theta^*)$ is non-testable against $H_1^{\mu_0}(\theta^*) : \overline{\mathsf{P}} \in \mathscr{P}_1^{\mu_0}(\theta^*)$. This is true for any $\mu_0 \in \mathbb{R}$, and obviously $\mathscr{P} = \bigcup_{\mu_0 \in \mathbb{R}} \mathscr{P}^{\mu_0}$ and $\mathscr{P}_k(\theta^*) = \bigcup_{\mu_0 \in \mathbb{R}} \mathscr{P}_k^{\mu_0}(\theta^*)$. We may thus apply Lemma B.5 and $H_0(\theta^*)$ is non-testable against $H_1(\theta^*)$. This is true for any $\theta^* \in \Theta$.

To prove partial non-testability of the variance, define $\Theta =]0, +\infty[$ and fix $\theta_0 \in \Theta$. Consider the testing problem $H_0(\theta_0) : \overline{\theta} \in \{\theta_0\}$ against $H_1(\theta_0) : \overline{\theta} \in \Theta_1$, where $\Theta_1 \equiv \Theta \setminus \{\theta_0\}$. For $\alpha \in [0, 1]$, let φ be an α-level test of $H_0(\theta_0)$ against $H_1(\theta_0)$. Consider $\tilde{\Theta}_1 = \{\theta_1\}$, where θ_1 is some real number such that $0 < \theta_1 < \theta_0$. Now, fix $\mathsf{P}_1 \in \theta^{-1}(\tilde{\Theta}_1)$. Proceeding as above, $\forall \epsilon \in]0, +\infty[$, $\exists \mathsf{P}_0 \in \theta^{-1}(\Theta_0)$ such that $D(\mathsf{P}_0, \mathsf{P}_1) < \epsilon$. Therefore, we may find a sequence $\{\mathsf{P}_{0,m}, m \geq 1\}$ of $\theta^{-1}(\Theta_0)$ such that $D(\mathsf{P}_{0,m}, \mathsf{P}_1) \to 0$, $m \to \infty$. As before, we get $\mathsf{E}_{\mathsf{P}_1}(\varphi_n) \leq \alpha$. This is true for any $\mathsf{P}_1 \in \theta^{-1}(\{\theta_1\})$. Hence $\sup_{\mathsf{P} \in \theta^{-1}(\tilde{\Theta}_1)} \mathsf{E}_{\mathsf{P}}(\varphi_n) \leq \alpha$. This result holds for any $\theta_0 \in \Theta$. QED

Proof of Proposition 5.4 For any $\pi \in]0, 1[$ and any P_0, P_1 in $\mathscr{P}$, we clearly have $\theta(\pi\mathsf{P}_0 + (1-\pi)\mathsf{P}_1) = \max\{\theta(\mathsf{P}_0), \theta(\mathsf{P}_1)\}$. Now choose any $\mathsf{P}_1 \in \mathscr{P} \setminus \mathscr{P}(\theta^*)$ and let θ_1 denote the value of $\theta(\mathsf{P}_1)$. Let $\tilde{\mathsf{P}}$ be an element of $\mathscr{P}(\theta^*)$, $\epsilon \in]0, +\infty[$, $k \geq n$ and $\pi \in]0, 1[$ such that $1 - \pi^k < \epsilon$. Define $\mathsf{P}_0 \equiv \pi\mathsf{P}_1 + (1-\pi)\tilde{\mathsf{P}}$. We have $\theta(\mathsf{P}_0) = \theta(\tilde{\mathsf{P}})$, hence $\mathsf{P}_0 \in \mathscr{P}(\theta^*)$. Thus, using Lemma B.2, for any $\epsilon > 0$, for any $\mathsf{P}_1 \in \mathscr{P} \setminus \mathscr{P}(\theta^*)$, there exists $\mathsf{P}_0 \in \mathscr{P}(\theta^*)$ such that $D(\mathsf{P}_0, \mathsf{P}_1) < \epsilon$. Using the same argument as in the proof of Proposition 5.3, for any α-level test φ of H_0, we have $\mathsf{E}_{\mathsf{P}_1}(\varphi_n) \leq \alpha$, $\forall \mathsf{P}_1 \in \mathscr{P} \setminus \mathscr{P}(\theta^*)$. Partial non-testability of θ follows from an argument similar to that of Proposition 5.3. QED

Proof of Proposition 6.1

(1) Fix $\beta_0 \in \mathbb{R}^K$ and $\mu_0 \in \mathbb{R}$, and define $\mathscr{P}^{\beta_0} \equiv \{\mathsf{P} \in \mathscr{P} : \beta(\mathsf{P}) = \beta_0\}$, $\mathscr{P}_0(\mu_0) \equiv \{\mathsf{P} \in \mathscr{P} : \mu(\mathsf{P}) = \mu_0\}$, $\mathscr{P}_1(\mu_0) \equiv \mathscr{P} \setminus \mathscr{P}_0(\mu_0)$ and $\mathscr{P}_k^{\beta_0}(\mu_0) \equiv \mathscr{P}_k(\mu_0) \cap \mathscr{P}^{\beta_0}$, $k = 0, 1$. Consider the test of $H_0^{\beta_0}(\mu_0) : \overline{\mathsf{P}} \in \mathscr{P}_0^{\beta_0}(\mu_0)$ against $H_1^{\beta_0}(\mu_0) : \overline{\mathsf{P}} \in \mathscr{P}_1^{\beta_0}(\mu_0)$. For $\mu_1 \neq \mu_0$, let $(X', Y')'$ have distribution P_1 such that $\theta(\mathsf{P}) = (\mu_1, \beta_0')'$. Clearly, $\mathsf{P}_1 \in \mathscr{P}_1^{\beta_0}(\mu_0)$. For $\pi \in]0, 1[$, let also $U = (U_1, \ldots, U_n)'$ be a n-vector of i.i.d. Bernoulli $\mathscr{B}(\pi)$ random variables, independent of $(X', Y')'$. Define $X^* \equiv X$ and for $i = 1, \ldots, n$,

$$Y_i^* \equiv U_i Y_i + (1 - U_i)\left(\beta_0' X_i + \frac{\mu_0 - \pi\mu_1}{1 - \pi}\right). \tag{A.8}$$

Obviously,

$$Y_i^* - \beta_0' X_i^* = U_i(Y_i - \beta_0' X_i) + (1 - U_i)\frac{\mu_0 - \pi\mu_1}{1-\pi}, \qquad i = 1, \dots, n. \tag{A.9}$$

Define $(\tilde{X}', \tilde{Y}')' \equiv \big(X', (X\beta_0 + \frac{\mu_0+\pi\mu_1}{1-\pi}\iota_n)'\big)'$. Let $\tilde{\mathsf{P}}$ and P_0 denote the distribution of $(\tilde{X}', \tilde{Y}')'$ and $(X^{*\prime}, Y^{*\prime})'$, respectively. From (A.8), we have $\mathsf{P}_0 = \pi\mathsf{P}_1 + (1-\pi)\tilde{\mathsf{P}}$. It is easily checked from (A.8) and (A.9) that $\mathsf{E}_{\mathsf{P}_0}(Y_i^*|X^*) = \mu_0 + \beta_0' X_i^*$ and $Y_1^* - \beta_0' X_1^*, \dots, Y_n^* - \beta_0' X_n^*$ are identically distributed, conditionally on X^*. Thus $\mathsf{P}_0 \in \mathscr{P}_0^{\beta_0}(\mu_0)$. This holds for any $\pi \in \,]0,1[$, any P_1 such that $\theta(\mathsf{P}_1) = (\mu_1, \beta_0')'$ and $\mu_1 \neq \mu_0$. Thus BSE2 is satisfied and $H_0^{\beta_0}(\mu_0)$ is non-testable against $H_1^{\beta_0}(\mu_0)$. As this holds for any μ_0, the parameter μ is non-testable in $\mathscr{P}^{\beta_0}$. As this is true for any $\beta_0 \in \mathbb{R}^K$, Lemma B.5 implies that μ is non-testable.

(2) We have shown in the above point that for all $\mu_0 \in \mathbb{R}$ and all $\beta_0 \in \mathbb{R}^K$, $H_0 : \{\overline{\mu} = \mu_0, \overline{\beta} = \beta_0\}$ is non-testable against $H_1 : \{\overline{\mu} \neq \mu_0, \overline{\beta} = \beta_0\}$. The result is then an immediate consequence of Proposition 2.10.

(3) This follows from Proposition 4.2. QED

Proof of Proposition 6.2 We first consider the case $\beta_0 = 0$ and $n_{X,0} = n$. We show that for any $\beta \in \mathbb{R}^K$, $\beta \neq 0$, and any $\alpha \in [\frac{1}{n}, 1[$, there exist an α-level test φ and a distribution P_1 in $\mathscr{P}$ with $\beta(\mathsf{P}_1) \neq 0$ such that $\mathsf{E}_{\mathsf{P}_1}(\varphi_n) > \alpha$. Choose $\beta_1 \in \mathbb{R}^K \backslash \{0\}$ such that $\#\{\beta_1' X_1, \dots, \beta_1' X_n\} = n$. Let $\Pi_X : \{1, \dots, n\} \to \{1, \dots, n\}$ be the permutation defined by $\Pi_X(i) = j$ iff $\beta_1' X_j$ has rank i when $\beta_1' X_1, \dots, \beta_1' X_n$ are ranked in increasing order. Define $T_i = \beta_1' X_{\Pi_X(i)}$ and $\tilde{Y}_i = Y_{\Pi_X(i)}$, $i = 1, \dots, n$. For any distribution P, any real number a and any $i = 1, \dots, n$, we have

$$\mathsf{P}(\tilde{Y}_i \leq a|X) = \sum_{k=1}^n \mathsf{P}(\tilde{Y}_i \leq a, \Pi_X(i) = k|X) = \sum_{k=1}^n \mathrm{I}(\Pi_X(i) = k)\mathsf{P}(Y_k \leq a|X). \tag{A.10}$$

For a strictly positive real number δ_X such that $\delta_X < \min\{T_i - T_{i-1}, i = 1, \dots, n\}$, define

$$I_1 =]-\infty, T_1+\delta_X], \quad I_i =]T_{i-1}+\delta_X, T_i+\delta_X], \; i = 2, \dots, n-1, \quad I_n =]T_{n-1}+\delta_X, +\infty[\tag{A.11}$$

Notice $T_i \in I_i$, $i = 1, \dots, n$. Consider the test φ defined by $\varphi_n = \mathrm{I}(\tilde{Y}_i \in I_i, i = 1, \dots, n)$. If $\beta(\overline{\mathsf{P}}) = 0$, $Y_1, \dots, Y_n$ are identically distributed conditionally on X. From (A.10) we see that $\tilde{Y}_1, \dots, \tilde{Y}_n$ are also identically distributed, conditionally on X. Then we have

$$\mathsf{E}_{\overline{\mathsf{P}}}(\varphi_n|X) = \overline{\mathsf{P}}(\bigcap_{i=1}^n \{\tilde{Y}_i \in I_i\}|X) \leq \min\{\overline{\mathsf{P}}(\tilde{Y}_i \in I_i|X), i = 1, \dots, n\} \tag{A.12}$$

$$= \min\{\overline{\mathsf{P}}(\tilde{Y}_1 \in I_i|X), i = 1, \dots, n\}. \tag{A.13}$$

As $(I_1, \ldots, I_n)$ is a partition of $\mathbb{R}$, we must have $\min\{\overline{\mathsf{P}}(\tilde{Y}_1 \in I_i|X), i = 1, \ldots, n\} \leq \frac{1}{n}$. As this is true unconditionally on X, the test φ has level α for any choice of $\alpha \geq \frac{1}{n}$. Now choose $0 < \eta_X < \min\{\delta_X, T_i - T_{i-1} - \delta_X, i = 1 \ldots, n\}$, $\xi \in]0, 1[$, and let $\nu = (\nu_1, \ldots, \nu_n)'$ be a real random n-vector such that $\nu|X \sim N(0, \sigma_X^2 I_n)$, where σ_X^2 is such that

$$\mathsf{P}(|\nu_1| < \eta_X|X) > 1 - \xi. \tag{A.14}$$

Let P_1 be the distribution of $(X', Y')'$, with $Y_i = \beta_1' X_i + \nu_i$, $i = 1, \ldots, n$. We clearly have $\mathsf{P}_1 \in \{\mathsf{P} \in \mathscr{P} : \upbeta(\mathsf{P}) \neq 0\}$. Thus, using (A.10) we get

$$\mathsf{P}_1(\tilde{Y}_i \in [T_i - \eta_X, T_i + \eta_X]|X) = \sum_{j=1}^{n} \mathrm{I}(\Pi_X(i) = j) \times \\ \mathsf{P}_1(Y_j \in [\beta_1' X_j - \eta_X, \beta_1' X_j + \eta_X]|X) > 1 - \xi,$$

$i = 1, \ldots, n$ where the last inequality results from (A.14). This relation also implies $\mathsf{P}_1(\tilde{Y}_i \in I_i|X) > 1 - \xi$, $i = 1, \ldots, n$ and therefore

$$\mathsf{E}_{\mathsf{P}_1}(\varphi_n|X) = \prod_{i=1}^{n} \mathsf{P}_1(\tilde{Y}_i \in I_i|X) > (1 - \xi)^n. \tag{A.15}$$

For any given $\xi \in]0, 1[$, there always exists suitable choices of η_X and σ_X^2 such that (A.15) holds. Therefore $\mathsf{E}_{\mathsf{P}_1}(\varphi_n|X)$ can be made arbitrarily close to 1.

Next, for $\beta_0 \neq 0$, consider a test of $H_0(\beta_0) : \overline{\beta} = \beta_0$ against $H_1(\beta_0) : \overline{\beta} \neq \beta_0$. We define $\varepsilon_i(\beta_0) \equiv Y_i - \beta_0' X_i$, $i = 1, \ldots, n$. The same steps as above, with Y_i replaced by $\varepsilon_i(\beta_0)$ and, may be followed to derive a somewhere powerful α-level test of $H_0(\beta_0)$ against $H_1(\beta_0)$, for any $\alpha \geq \frac{1}{n}$.

Finally, if $n_{X,0} \equiv \max_{\beta \in \mathbb{R}^K \setminus \{\beta_0\}} \#\{(\beta - \beta_0)' X_1, \ldots, (\beta - \beta_0)' X_n\} < n$, we proceed as above, choosing $\beta_1 \neq \beta_0$ such that $\#\{(\beta_1 - \beta_0)' X_1, \ldots, (\beta_1 - \beta_0)' X_n\} = n_{X,0}$ and keeping only $n_{X,0}$ observations with pairwise distinct $(\beta_1 - \beta_0)' X_i$.

Testability of the parameter $\upbeta$ then follows from the above result and Definition 2.3. QED

Proof of Proposition 6.3 Choose $\beta_0 \in \mathbb{R}^K$ and consider the problem of testing $H_0(\beta_0) : \overline{\mathsf{P}} \in \mathscr{P}_0(\beta_0)$ against $H_1(\beta_0) : \overline{\mathsf{P}} \in \mathscr{P}_1(\beta_0)$, where $\mathscr{P}_0(\beta_0) \equiv \{\mathsf{P} \in \mathscr{P} : \upbeta(\mathsf{P}) = \beta_0\}$ and $\mathscr{P}_1(\beta_0) = \mathscr{P} \setminus \mathscr{P}_0(\beta_0)$. It follows from the definition of $\mathscr{P}$ that $\mathscr{P}_0(\beta_0)$ and $\mathscr{P}_1(\beta_0)$ are non empty. Take $\mathsf{P}_1 \in \mathscr{P}_1(\beta_0)$ and let β_1 be the parameter associated with P_1: $\beta_1 = \upbeta(\mathsf{P}_1)$. Choose $\epsilon \in]0, \infty[$ and $\pi_i \in]0, 1[$, $i = 1, \ldots, n$, such that $\sum_{i=1}^n (1 - \pi_i) < \epsilon$. We may always find a probability space $(\Omega, \mathscr{A}, \mathsf{m})$ and a real random $(nK + 2n)$-vector $\Upsilon = (X', Y', U')'$ defined on Ω with a distribution such that

1. $(X', Y')' \sim \mathsf{P}_1$, 2. $Y \perp U|X$, 3. $U|X \sim \bigotimes_{i=1}^n \mathscr{B}(\pi_i)$,

where $A \perp B|C$ stands for A is independent of B conditionally on C. Next define $\tilde{Y}_i \equiv [\mu_0 + X_i'\beta_0 - \pi_i(\mu_1 + X_i'\beta_1)]/(1 - \pi_i)$, $i = 1, \ldots, n$, where $\mu_1 \equiv \upmu(\mathsf{P}_1)$ and μ_0 is any real number. Clearly $\mathsf{E}(\tilde{Y}_i|X) = [\mu_0 + X_i'\beta_0 - \pi_i(\mu_1 + X_i'\beta_1)]/(1-\pi_i)$. Now define $X^* = X$ and construct the n real random variables $Y_i^* \equiv U_i Y_i + (1-U_i)\tilde{Y}_i$, $i = 1, \ldots, n$. From 1 to 3 above, we have $\mathsf{E}(Y_i^*|X^*) = \pi_i \mathsf{E}(Y_i|X) + (1-\pi_i)\mathsf{E}(\tilde{Y}_i|X) = \mu_0 + X_i'\beta_0$. In other words, the distribution P_0 of $(X^{*\prime}, Y^{*\prime})'$ lies in $\mathscr{P}_0(\beta_0)$. Now, a straightforward application of Lemma B.3 yields $\sup_{g \in \mathcal{B}_{nK+n}} \left|\mathsf{E}\big(g(X^*, Y^*)\big) - \mathsf{E}\big(g(X,Y)\big)\right| \leq \sum_{i=1}^n (1-\pi_i)$. The choice of $\pi_1, \ldots, \pi_n$ implies this supremum is less that ϵ. This holds for any $\epsilon \in]0, +\infty[$ and any $\mathsf{P}_1 \in \mathscr{P}_1(\beta_0)$. It follows that for any test φ of $H_0(\beta_0)$ we have

$$\forall \mathsf{P}_1 \in \mathscr{P}_1(\beta_0), \forall \epsilon \in]0, +\infty[, \exists \mathsf{P}_0 \in \mathscr{P}_0(\beta_0) \text{ such that } |\mathsf{E}_{\mathsf{P}_0}(\varphi_n) - \mathsf{E}_{\mathsf{P}_1}(\varphi_n)| < \epsilon.$$

Therefore, the power of any α-level test of $H_0(\beta_0)$ against $H_1(\beta_0)$ is uniformly bounded from above by α. As this is true for any β_0, the parameter β is non-testable. QED

Proof of Proposition 6.4 We fix $\beta_0 \in \mathbb{R}^K$ and consider the test of $H_0(\beta_0) : \upbeta(\overline{\mathsf{P}}) = \beta_0$ against $H_1(\beta_0) : \upbeta(\overline{\mathsf{P}}) \neq \beta_0$. This testing problem is successively investigated in various statistical models, all being subsets of $\mathscr{P}$. These are defined from the following families of probability distributions:

$\mathscr{P}^{\mathrm{H}} \equiv \{\mathsf{P} \in \mathscr{P} : \mathsf{V}(Y_i|X) = \mathsf{V}(Y_j|X), \forall i, j = 1, \ldots, n\}$,

$\mathscr{P}^{\mathrm{U}} \equiv \{\mathsf{P} \in \mathscr{P} : \mathsf{Cov}(Y_i, Y_j|X) = 0, \forall i, j = 1, \ldots, n, i \neq j\}$,

$\mathscr{P}^{\mathrm{I}} \equiv \{\mathsf{P} \in \mathscr{P} : Y_1 - \mathsf{E}(Y_1|X), \ldots, Y_n - \mathsf{E}(Y_n|X)$ are independent conditionally on $X\}$.

In each case considered below, the model is a set $\mathscr{P}^{\mathrm{L}}$ of family distributions, where L is a character string made of some combination of the letters H, U and I. We write the null and alternative hypotheses as $H_0(\beta_0) : \overline{\mathsf{P}} \in \mathscr{P}_0^{\mathrm{L}}(\beta_0)$ and $H_0(\beta_0) : \overline{\mathsf{P}} \in \mathscr{P}_1^{\mathrm{L}}(\beta_0)$, where $\mathscr{P}_0^{\mathrm{L}}(\beta_0) \equiv \{\mathsf{P} \in \mathscr{P}^{\mathrm{L}} : \upbeta(\mathsf{P}) = \beta_0\}$ and $\mathscr{P}_1^{\mathrm{L}}(\beta_0) \equiv \mathscr{P}^{\mathrm{L}} \setminus \mathscr{P}_0^{\mathrm{L}}(\beta_0)$.

1. Model $\mathscr{P}^{\mathrm{H}}$: a linear regression model with conditionally homoskedastic error terms. Choose any element $\mathsf{P}_1^{\mathrm{H}}$ of $\mathscr{P}_1^{\mathrm{H}}(\beta_0)$ and let $\beta_1 \equiv \upbeta(\mathsf{P}_1^{\mathrm{H}})$. Choose $\epsilon \in]0, \infty[$. For any $x \in \mathbb{R}^{nK}$, let x_i denote the vector of the $[(i-1)K+1]$-th to iK-th coordinates of x, $i = 1, \ldots, n$. It is easy to check that for all $x \in \mathbb{R}^{nK}$, we may always find a strictly positive real number κ such that $\sum_{i=1}^n \frac{\kappa(x_i'\beta_1 - x_i'\beta_0)^2}{1+\kappa(x_i'\beta_1 - x_i'\beta_0)^2} < \epsilon$. Let $c : \mathbb{R}^{nK} \to \mathbb{R}$ be the mapping which associates such a κ with $x \in \mathbb{R}^{nK}$. We may always find a probability space $(\Omega, \mathscr{A}, \mathsf{m})$ and a real random $(nK + 3n)$-vector $\Upsilon^{\mathrm{H}} = (X', Y', \tilde{\varepsilon}', U')'$ defined on Ω with a distribution such that:

   (a$^{\mathrm{H}}$) $(X', Y')' \sim \mathsf{P}_1^{\mathrm{H}}$,

   (b$^{\mathrm{H}}$) $\mathsf{E}(\tilde{\varepsilon}|X) = 0$ and $\mathsf{V}(\tilde{\varepsilon}|X) = \sigma^2 I_n$,

   (c$^{\mathrm{H}}$) $U|X \sim \bigotimes_{i=1}^n \mathscr{B}(\pi_i)$,

   (d$^{\mathrm{H}}$) $Y \perp U|X$ and $\tilde{\varepsilon} \perp U|X$.

   where σ^2 is the conditional variance of Y_1 given X, and $\pi_i \equiv \left[1 + c(X)(X_i'\beta_1 - X_i'\beta_0)^2\right]^{-1}$, $i = 1, \ldots, n$. One easily checks that this choice of the π_i and the definition of the mapping c entail

$$\sum_{i=1}^n (1-\pi_i) < \epsilon, \quad \text{always.} \tag{A.16}$$

Now define $\mu_1 \equiv \mu(\mathsf{P}_1^{\mathrm{H}})$, $X^* \equiv X$ and

$$\tilde{Y}_i \equiv \mu_1 + \frac{X_i'\beta_0 - \pi_i X_i'\beta_1}{1-\pi_i} + \tilde{\varepsilon}_i, \tag{A.17}$$

$$Y_i^* \equiv U_i Y_i + (1-U_i)\tilde{Y}_i, \tag{A.18}$$

for $i = 1, \ldots, n$. Let $\mathsf{P}_0^{\mathrm{H}}$ denote the distribution of $(X^{*\prime}, Y^{*\prime})' \equiv (X_1^{*\prime}, \ldots, X_n^{*\prime}, Y_1^*, \ldots, Y_n^*)'$. Under (a$^{\mathrm{H}}$) to (d$^{\mathrm{H}}$) above, it is easy to verify that $\mathsf{E}(Y_i^*|X^*) = \mu_1 + X_i^{*\prime}\beta_0$ and $\mathsf{V}(Y_i^*|X^*) = \pi_i \mathsf{E}(Y_i^2|X) + (1-\pi_i)\mathsf{E}(\tilde{Y}_i^2|X) - (\mu_1 + X_i'\beta_0)^2$, $i = 1, \ldots, n$. Also

$$\mathsf{E}(\tilde{Y}_i^2|X) = \sigma^2 + \mathsf{E}(\tilde{Y}_i|X)^2 = \sigma^2 + \mu_1^2 + \frac{2\mu_1(X_i'\beta_0 - \pi_i X_i'\beta_1)}{1-\pi_i} + \frac{(X_i'\beta_0)^2 + \pi_i^2(X_i'\beta_1)^2 - 2\pi_i X_i'\beta_0 X_i'\beta_1}{(1-\pi_i)^2},$$

and $\mathsf{E}(Y_i^2|X) = \sigma^2 + \mu_1^2 + (X_i'\beta_1)^2 + 2\mu_1 X_i'\beta_1$. Substituting and gathering terms, we get $\mathsf{V}(Y_i^*|X^*) = \sigma^2 + \frac{\pi_i}{1-\pi_i}(X_i^{*\prime}\beta_1 - X_i^{*\prime}\beta_0)^2 = \sigma^2 + \frac{1}{c(X^*)}$, $i = 1, \ldots, n$, where the last equality results from substituting the expression of π_i. The above computations of the conditional expectation and variance of Y^* given X^* show that $\mathsf{P}_0^{\mathrm{H}} \in \mathscr{P}_0^{\mathrm{H}}(\beta_0)$. Using equation (A.16) above and Lemma B.3, we have $\sup_{g \in \mathcal{B}_{nK+n}} \left|\mathsf{E}\big(g(X^*, Y^*)\big) - \mathsf{E}\big(g(X, Y)\big)\right| < \epsilon$. We conclude as in the proof of Proposition 6.3.

2. Model $\mathscr{P}^{\mathrm{U}}$: a linear regression model with conditionally uncorrelated, possibly dependent, possibly heteroskedastic error terms. Choose any element $\mathsf{P}_1^{\mathrm{U}}$ of $\mathscr{P}_1^{\mathrm{U}}(\beta_0)$ and let $\beta_1 \equiv \beta(\mathsf{P}_1^{\mathrm{U}})$. Also choose any $\epsilon \in]0, \infty[$ and any $(\pi_1, \ldots, \pi_n)$ in $]0, 1[^n$ with $\sum_{i=1}^n (1-\pi_i) < \epsilon$. As in the previous point, we introduce the random vector $\Upsilon^{\mathrm{U}} = (X', Y', U)'$ satisfying conditions (a$^{\mathrm{U}}$) to (c$^{\mathrm{U}}$) where:

   (a$^{\mathrm{U}}$) $(X', Y')' \sim \mathsf{P}_1^{\mathrm{U}}$, (b$^{\mathrm{U}}$) $U \perp X \sim \bigotimes_{i=1}^n \mathscr{B}(\pi_i)$, (c$^{\mathrm{U}}$) $Y \perp U | X$.

   Set $\mu_1 \equiv \mu(\mathsf{P}_1^{\mathrm{U}})$, and define $X^* \equiv X$, $\tilde{Y}_i \equiv \mu_1 + \frac{X_i'\beta_0 - \pi_i X_i'\beta_1}{1-\pi_i}$ and $Y_i^* \equiv U_i Y_i + (1-U_i)\tilde{Y}_i$, $i = 1, \ldots, n$. Notice $\mathsf{Cov}(Y_i, \tilde{Y}_j|X) = \mathsf{Cov}(\tilde{Y}_i, \tilde{Y}_j|X) = 0$, $\forall i, j$. Let $\mathsf{P}_0^{\mathrm{U}}$ denote the distribution of $(X^{*\prime}, Y^{*\prime})'$. Under (a$^{\mathrm{U}}$) to (c$^{\mathrm{U}}$) we have $\mathsf{E}(Y_i^*|X^*) = \mu_1 + X_i^{*\prime}\beta_0$ and $\mathsf{Cov}(Y_i^*, Y_j^*|X^*) = \pi_i\pi_j \mathsf{Cov}(Y_i, Y_j) = 0$, for $i \neq j$. Therefore $\mathsf{P}_0^{\mathrm{U}} \in \mathscr{P}_0^{\mathrm{U}}(\beta_0)$. Conclude as in 1 above.

3. Model $\mathscr{P}^{\mathrm{UH}} \equiv \mathscr{P}^{\mathrm{U}} \cap \mathscr{P}^{\mathrm{H}}$: a linear regression model with conditionally uncorrelated and homoskedastic but possibly dependent error terms. Choose any element $\mathsf{P}_1^{\mathrm{UH}}$ of $\mathscr{P}_1^{\mathrm{UH}}(\beta_0)$ and let $\beta_1 \equiv \beta(\mathsf{P}_1^{\mathrm{UH}})$. Also choose any $\epsilon \in]0, \infty[$. Consider the random vector $\Upsilon^{\mathrm{UH}} = (X', Y, \tilde{\varepsilon}', U')'$ such that (a$^{\mathrm{UH}}$) to (d$^{\mathrm{UH}}$) hold where

   (a$^{\mathrm{UH}}$) $(X', Y')' \sim \mathsf{P}_1^{\mathrm{UH}}$,

   and (b$^{\mathrm{UH}}$), (c$^{\mathrm{UH}}$), (d$^{\mathrm{UH}}$) identical to (b$^{\mathrm{H}}$),(c$^{\mathrm{H}}$), (d$^{\mathrm{H}}$), respectively, and with $\pi_1, \ldots, \pi_n$ as in 1 above. We define $X^* \equiv X$, and $\tilde{Y}_i$ and Y_i^*, $i = 1, \ldots, n$, as in (A.17) and

(A.18) and we let $\mathsf{P}_0^{\mathrm{UH}}$ denote the distribution of $(X^{*\prime}, Y^{*\prime})'$. Using (a$^{\mathrm{UH}}$) to (d$^{\mathrm{UH}}$), the same computations as in 1 and 2 above yield $\mathsf{P}_0^{\mathrm{UH}} \in \mathscr{P}_0^{\mathrm{UH}}$. Conclude as in 1.

4. Model $\mathscr{P}^{\mathrm{I}}$: a linear regression model with conditionally independent and possibly heteroskedastic error terms. Choose any element $\mathsf{P}_1^{\mathrm{I}}$ of $\mathscr{P}_1^{\mathrm{I}}(\beta_0)$ and let $\beta_1 \equiv \beta(\mathsf{P}_1^{\mathrm{I}})$. Also choose any $\epsilon \in]0, \infty[$ and $\pi \in]0, 1[$ such that $1 - \pi^n < \epsilon$. Consider the random vector $\Upsilon^{\mathrm{I}} = (X', Y, U')'$ such that (a$^{\mathrm{I}}$) to (c$^{\mathrm{I}}$) hold where

   (a$^{\mathrm{I}}$) $(X', Y')' \sim \mathsf{P}_1^{\mathrm{I}}$, (b$^{\mathrm{I}}$) $U|X \sim \mathscr{B}(\pi)^{\otimes n}$, (c$^{\mathrm{I}}$) $Y \perp U|X$.

   Set $\mu_1 \equiv \mu(\mathsf{P}_1^{\mathrm{U}})$ and define $X^* \equiv X$, $\tilde{Y}_i \equiv \mu_1 + \frac{X_i'\beta_0 - \pi X_i'\beta_1}{1-\pi}$ and Y_i^* as in (A.18) with $\pi_i = \pi$, $i = 1 \dots, n$. Let $\mathsf{P}_0^{\mathrm{I}}$ denote the distribution of $(X^{*\prime}, Y^{*\prime})'$. Conditions (a$^{\mathrm{I}}$) to (c$^{\mathrm{I}}$) entail $\mathsf{E}(Y_i^*|X^*) = \mu_1 + X_i^{*\prime}\beta_0$. Moreover, the same conditions imply that $(Y', \tilde{Y}', U')'$ is a vector of independent random variables, conditionally on X. Therefore $Y_1^*, \dots, Y_n^*$ are independent, conditionally on X^*. Hence $\mathsf{P}_0^{\mathrm{I}} \in \mathscr{P}_0^{\mathrm{I}}(\beta_0)$. Moreover, Lemma B.3 implies $\sup_{g \in \mathcal{B}_{nK+n}} \left|\mathsf{E}\big(g(X^*, Y^*)\big) - \mathsf{E}\big(g(X, Y)\big)\right| \leq 1 - \pi^n$. Then conclude as in 1 above.

5. Model $\mathscr{P}^{\mathrm{IH}} \equiv \mathscr{P}^{\mathrm{I}} \cap \mathscr{P}^{\mathrm{H}}$: a linear regression model with conditionally independent and homoskedastic error terms. Choose any element $\mathsf{P}_1^{\mathrm{IH}}$ of $\mathscr{P}_1^{\mathrm{IH}}(\beta_0)$ and let $\beta_1 \equiv \beta(\mathsf{P}_1^{\mathrm{IH}})$. Also choose any $\epsilon \in]0, \infty[$ Consider the random vector Υ^{IH} $(X', Y, \tilde{\varepsilon}', U')'$ such that (a$^{\mathrm{IH}}$) to (d$^{\mathrm{IH}}$) hold where

   (a$^{\mathrm{IH}}$) $(X', Y')' \sim \mathsf{P}_1^{\mathrm{IH}}$,

   (b$^{\mathrm{IH}}$) $U \perp X \sim \bigotimes_{i=1}^n \mathscr{B}(\pi_i)$,

   (c$^{\mathrm{IH}}$) $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$ are i.i.d. conditionally on X, $\mathsf{E}(\tilde{\varepsilon}_1|X) = 0$, $\mathsf{V}(\tilde{\varepsilon}_1|X) = \sigma^2$,

   (d$^{\mathrm{IH}}$) $(Y, U, \tilde{\varepsilon})$ are independent conditionally on X,

   and σ^2 denotes the conditional variance of Y_1 given X, and $\pi_1, \dots, \pi_n$ are set as in 1 above. Set $\mu_1 \equiv \mu(\mathsf{P}_1^{\mathrm{IH}})$ and define $X^* \equiv X$, $\tilde{Y}_i$ and Y_i^* as in (A.17) and (A.18), $i = 1, \dots, n$. Let $\mathsf{P}_0^{\mathrm{IH}}$ denote the distribution of $(X^{*\prime}, Y^{*\prime})'$. Conditions (a$^{\mathrm{IH}}$) to (d$^{\mathrm{IH}}$) entail $\mathsf{E}(Y_i^*|X^*) = \mu_1 + X_i^{*\prime}\beta_0$. Also, conditionally on X^*, $Y_1^*, \dots, Y_n^*$ are homoskedastic as in 1 above. Moreover, under the same conditions, $Y_1^*, \dots, Y_n^*$ are independent conditionally on X^*. In other words $\mathsf{P}_0^{\mathrm{IH}} \in \mathscr{P}_0^{\mathrm{IH}}(\beta_0)$. Conclude as in 1 above. QED

Proof of Proposition 6.5 Using the same notation as in the proof of Proposition 6.4, define $\mathscr{P}^{\mathrm{R}} \equiv \{\mathsf{P} \in \mathscr{P} : \text{C1 and C3 hold}\}$. Choose any element $\mathsf{P}_1^{\mathrm{R}}$ of $\mathscr{P}_1^{\mathrm{R}}(\beta_0)$ and let $\beta_1 \equiv \beta(\mathsf{P}_1^{\mathrm{R}})$. Choose $\epsilon \in]0, \infty[$ and $\pi \in]0, 1[$ such that $1 - \pi^n < \epsilon$. Consider the random vector $\Upsilon = (X', Y', U')'$ such that (a$^{\mathrm{R}}$) $(X', Y')' \sim \mathsf{P}_1^{\mathrm{R}}$ and (b$^{\mathrm{R}}$) $U|(X', Y')' \sim \mathscr{B}(\pi)^{\otimes n}$ hold. Now define $\mu_1 \equiv \mu(\mathsf{P}_1^{\mathrm{R}})$, $X^* \equiv X$, $\tilde{Y}_i \equiv \mu_1 + \frac{X_i'\beta_0 - \pi X_i'\beta_1}{1-\pi}$ and $Y_i^* \equiv U_i Y_i + (1 - U_i)\tilde{Y}_i$, $i = 1, \dots, n$. Let $\mathsf{P}_0^{\mathrm{R}}$ denote the distribution of $(X^{*\prime}, Y^{*\prime})'$. Under the above conditions (a$^{\mathrm{R}}$) and (b$^{\mathrm{R}}$), we check that $\mathsf{P}_0^{\mathrm{R}} \in \mathscr{P}_0^{\mathrm{R}}(\beta_0)$ as in the proof of Proposition 6.4. The same conclusion applies. QED

Proof of Proposition 7.1 We know that in a pure i.i.d. sampling scheme, the expectation is a non-testable parameter when no restrictions are put on the family of distributions (see Bahadur and Savage (1956)), while it becomes testable when those distributions are imposed to be bounded, with known bounds (see Anderson (1967)). In other words, while $\big(H_0 : \mathsf{E}_{\mathsf{P}}(Z_i) = 0, \forall i; H_1 : \mathsf{E}_{\mathsf{P}}(Z_i) \neq 0 \text{ for some } i\big)$ is non-testable, $\big(\tilde{H}_0 : \mathsf{E}_{\mathsf{P}}\big(h(Z_i,\theta)\big) = 0, \forall i; \tilde{H}_1 : \mathsf{E}_{\mathsf{P}}\big(h(Z_i,\theta)\big) \neq 0 \text{ for some } i\big)$ is a testable problem, for any bounded function h, with known bounds. However, the equivalence of Proposition 7.1 implies these two problems are equivalent. QED

Proof of Proposition 7.2 Let $Z = (X', Y')' \sim \mathsf{P}_1 \in \mathscr{P}_1(\theta_0)$, with $\theta(\mathsf{P}_1) = \theta_1 \neq \theta_0$. For any $(\pi_1, \ldots, \pi_n)' \in \,]0,1[^n$, let $U = (U_1, \ldots, U_n)$ be distributed as $\bigotimes_{i=1}^n \mathscr{B}(\pi_i)$, independent of Z. Define $X^* \equiv X$, $\tilde{Y}_i \equiv \frac{g(X_i,\theta_0) - \pi_i g(X_i,\theta_1)}{1-\pi_i}$, and $Y_i^* \equiv U_i Y_i + (1-U_i)\tilde{Y}_i$, $i = 1, \ldots, n$. Let P_0 denote the distribution of $(X^{*\prime}, Y^{*\prime})'$. We have

$$\mathsf{E}_{\mathsf{P}_0}(Y_i^*|X^*) = \pi_i \mathsf{E}_{\mathsf{P}_0}(Y_i|X) + (1-\pi_1)\frac{g(X_i,\theta_0) - \pi_i g(X_i,\theta_1)}{1-\pi_i} = 0, \quad i = 1, \ldots, n.$$

Thus $\mathsf{P}_0 \in \mathscr{P}_0(\theta_0)$. This holds for any $\theta_0 \in \Theta$. We conclude as above. QED

Proof of Proposition 7.3 Choose $\theta_0 \in \Theta$ such that is equivalence class $\mathscr{E}(\theta_0)$ is not a singleton and choose $\theta_1 \in \mathscr{E}(\theta_0)$, $\theta_1 \neq \theta_0$. Define $\mathscr{P}_0 \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) = \theta_0\}$ and $\mathscr{P}_1 \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) = \theta_1\}$, choose any $\mathsf{P}_1 \in \mathscr{P}_1$. For any $\pi \in]0,1[$, we construct the random vector $\Upsilon \equiv (X', Y', U)'$ such that

(a) $(X', Y')' \sim \mathsf{P}_1$, (b) $U : X \sim \mathscr{B}(\pi)^{\otimes n}$, (c) U and Y are independent, conditionally on X.

Define $X^* \equiv X$, $\tilde{Y}_i \equiv \frac{g(X_i,\theta_0) - \pi g(X_i,\theta_1)}{1-\pi}$ and $Y_i^* = U_i Y_i + (1-U_i)\tilde{Y}_i$, $i = 1, \ldots, n$. We let P_0 denote the distribution of $(X^{*\prime}, Y^{*\prime})'$. We show $\mathsf{P}_0 \in \mathscr{P}_0$. Under conditions (a) to (c) above, one easily checks that $\mathsf{E}(Y_i^*|X^*) = g(X_i^*, \theta_0)$, $i = 1, \ldots, n$ and that $(Y', \tilde{Y}', U')'$ is a vector of independent random variables, conditionally on X. It follows that $\varepsilon_1^*, \ldots, \varepsilon_n^*$ are independent conditionally on X^*, where $\varepsilon_i^* \equiv Y_i^* - g(X_i^*, \theta_0)$, $i = 1, \ldots, n$. Now we also have

$$\begin{aligned} \varepsilon_i^* &= U_i\big[Y_i - g(X_i,\theta_1)\big] + (1-U_i)\frac{g(X_i,\theta_0) - \pi g(X_i,\theta_1)}{1-\pi} + U_i g(X_i,\theta_1) - g(X_i,\theta_0) \\ &= U_i\big[Y_i - g(X_i,\theta_1)\big] + \frac{\pi - U_i}{1-\pi}\big[g(X_i,\theta_0) - g(X_i,\theta_1)\big]. \end{aligned}$$

As $\theta_1 \in \mathscr{E}(\theta_0)$, we can find $\mu \in \mathbb{R}$ such that $g(X_i,\theta_0) - g(X_i,\theta_1) = \mu$, $\forall i = 1, \ldots, n$. Thus under (a) to (c), $\varepsilon_1^*, \ldots, \varepsilon_n^*$ are i.i.d., conditionally on X^*. Therefore, $\mathsf{P}_0 \in \mathscr{P}_0$ as announced. From Lemma B.3, it follows that $\sup_{g \in \mathcal{B}_{np+n}} \big|\mathsf{E}_{\mathsf{P}_0}\big(g(X^*, Y^*)\big) - \mathsf{E}_{\mathsf{P}_1}\big(g(X, Y)\big)\big| \leq n(1-\pi)$. As this is true for any $\pi \in]0,1[$ and any $\mathsf{P}_1 \in \mathscr{P}_1$, we conclude that $H_0(\theta_0)$ is non-testable against $H_1(\theta_1)$. QED

Proof of Proposition 7.4 We propose a proof similar to that of Proposition 6.2. Let $\{\mathscr{E}_\ell : \ell \in L\}$ be the family of all equivalent classes generated by $\mathcal{R}$. In each $\mathscr{E}_\ell$ choose a representative element $\theta_\ell \in \mathscr{E}_\ell$. For $i = 1, \dots, n$ define $\varepsilon_i(\theta_\ell) \equiv Y_i - g(X_i, \theta_\ell) = \varepsilon_i + g\big(X_i, \theta(\overline{\mathsf{P}})\big) - g(X_i, \theta_\ell)$, where $\varepsilon_i \equiv Y_i - \mathsf{E}_{\overline{\mathsf{P}}}(Y_i|X) = Y_i - g\big(X_i, \theta(\overline{\mathsf{P}})\big)$.

We first assume $n_{X,\ell} = n$. In this case, we may find $k \neq \ell$ such that $\#\{g(X_1, \theta_k) - g(X_1, \theta_\ell), \dots, g(X_n, \theta_k) - g(X_n, \theta_\ell)\} = n$. Define $G_i \equiv g(X_i, \theta_k) - g(X_i, \theta_\ell)$, $i = 1, \dots, n$, and $\Pi_X : \{1, \dots, n\} \to \{1, \dots, n\}$ such that $T_1 < \cdots < T_n$, where $T_i \equiv G_{\Pi_X(i)}$, $i = 1, \dots, n$. Also choose $0 < \delta_X < \min\{T_i - T_{i-1}, i = 1, \dots, n\}$ and define the intervals I_i, $i = 1, \dots, n$ as in (A.11). Consider the test φ defined as $\varphi_n = \mathrm{I}(\tilde{\varepsilon}_i(\theta_\ell) \in I_i, i = 1, \dots, n)$, where $\tilde{\varepsilon}_i \equiv \varepsilon_{\Pi_X(i)}$, $i = 1, \dots, n$.

Assume $\overline{\mathsf{P}} \in \mathscr{P}_\ell$. Then $\varepsilon_i(\theta_\ell) \equiv Y_i - g(X_i, \theta_\ell) = \varepsilon_i + \mu_\ell$, $\forall i = 1, \dots, n$, for some $\mu_\ell \in \mathbb{R}$. Condition C8 implies that under $H_0(\ell)$, $\varepsilon_1(\theta_\ell), \dots, \varepsilon_n(\theta_\ell)$ are identically distributed, conditionally on X. Thus, using (A.10) and (A.13) for $\varepsilon_1(\theta_\ell), \dots, \varepsilon_n(\theta_\ell)$, we get $\mathsf{E}_{\overline{\mathsf{P}}}(\varphi_n|X) \leq \frac{1}{n}$. Thus φ has level α provided we choose $\alpha \geq \frac{1}{n}$. Next, choose $\xi \in]0, 1[$ and $0 < \eta_X < \min\{\delta_X, T_i - T_{i-1} - \delta_X, i = 1, \dots, n\}$ and let $\nu = (\nu_1, \dots, \nu_n)'$ be the real random vector of the proof of Proposition 6.2. Now let P_1 be the distribution of $(X', Y')'$, with $Y_i = g(X_i, \theta_1) + \nu_i$, $i = 1, \dots, n$. Clearly $\theta(\mathsf{P}_1) = \theta_1 \in \mathscr{E}_k$ and $\mathsf{P}_1 \notin \mathscr{P}_\ell$. Then

$$\mathsf{P}_1(G_i - \eta_X \leq \varepsilon_i(\theta_\ell) \leq G_i + \eta_X|X) = \mathsf{P}_1(-\eta_X \leq \nu_i \leq \eta_X|X) > 1 - \xi, \quad i = 1, \dots, n,$$

which implies $\mathsf{P}_1(T_i - \eta_X \leq \tilde{\varepsilon}_i(\theta_\ell) \leq T_i + \eta_X|X) > 1 - \xi$, $i = 1, \dots, n$. The choice of η_X and σ_X^2 entails $\mathsf{P}_1(\tilde{\varepsilon}_i(\theta_\ell) \in I_i|X) \geq \mathsf{P}_1(T_i - \eta_X \leq \tilde{\varepsilon}_i(\theta_\ell) \leq T_i + \eta_X|X)$, $i = 1, \dots, n$. Thus $\mathsf{E}_{\mathsf{P}_1}(\varphi_n) = \prod_{i=1}^n \mathsf{P}_1(\tilde{\varepsilon}_i(\theta_\ell) \in I_i|X) > (1 - \xi)^n$. As this can be obtained for any $\xi \in]0, 1[$, we can find a distribution not in $\mathscr{P}_\ell$ for which the power of φ_n is arbitrarily close to 1.

If $n_{X,\ell} < n$, let $k \neq \ell$ be such that $n_{X,\ell} = \#\{g(X_1, \theta_k) - g(X_1, \theta_\ell), \dots, g(X_n, \theta_k) - g(X_n, \theta_\ell)\}$. Then proceed as above keeping only $n_{X,\ell}$ observations i with pairwise distinct values of $g(X_i, \theta_k) - g(X_i, \theta_\ell)$. QED

Proof of Proposition 7.5 Define $\mathscr{P}(\theta_0) \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) = \theta_0\}$ and $\mathscr{P}(\theta_1) \equiv \{\mathsf{P} \in \mathscr{P} : \theta(\mathsf{P}) = \theta_1\}$ Choose $\mathsf{P}_1 \in \mathscr{P}(\theta_1)$, $\epsilon \in]0, 1[$ and $\pi \in]0, 1[$ such that $n(1 - \pi) < \epsilon$. Consider the random real $(nK + 2n)$-vector $\Upsilon = (X', Y', U')'$ such that

1. $(X', Y')' \sim \mathsf{P}_1$, 2. $U|X \sim \mathscr{B}(\pi)^{\otimes n}$, 3. $U \perp Y|X$.

Define $X^* \equiv X$, $\tilde{Y}_i \equiv \frac{\theta_0(X_i) - \pi\theta_1(X_i)}{1-\pi}$ and $Y_i^* \equiv U_i Y_i + (1 - U_i)\tilde{Y}_i$ $i = 1, \dots, n$. Let P_0^* be the distribution of $(X_i^{*\prime}, Y_i^*)'$. We have $\mathsf{E}(\tilde{Y}_i|X) = \frac{\theta_0(X_i) - \pi\theta_1(X_i)}{1-\pi}$ and then $\mathsf{E}(Y_i^*|X^*) = \theta_0(X_i^*)$, under 2 and 3 above. Moreover, under the same conditions, the conditional distribution of U given $(X', Y)'$ is $\mathscr{B}(\pi)^{\otimes n}$. Therefore, under 1 to 3 above, $\Upsilon \sim \mathscr{B}(\pi)^{\otimes n} \otimes \mathsf{P}_1 = \big(\mathscr{B}(\pi) \otimes \mathsf{P}_{11}\big)^{\otimes n}$, where P_{11} is the distribution of $(X_1', Y_1)'$. In other words, $(X_i', Y_i, U_i)'$, $i = 1, \dots, n$, are i.i.d. Now as $(X_i^{*\prime}, Y_i^*)' = g(X_i, Y_i, U_i)$, $i = 1, \dots, n$, for some mapping g from $\mathbb{R}^{K+2}$ into $\mathbb{R}^{K+1}$, it follows that

$(X_1^{*\prime}, Y_1^*)', \ldots, (X_n^{*\prime}, Y_n^*)'$ are also i.i.d. Thus $\mathsf{P}_0^* \in \mathscr{P}_0(\theta_0)$ and the result follows as in the proof of Proposition 6.3. QED

Proof of Proposition 7.7 Choose $\mathsf{P}_1 \notin \mathscr{P}_W$ and define $\theta_1 \equiv \theta(\mathsf{P}_1)$. Choose $\epsilon \in]0,1[$ and $\pi \in]0,1[$ such that $n(1-\pi) < \epsilon$. Consider the random real $(nK+2n)$-vector $\Upsilon = (X', Y', U')'$ such that

1. $(X', Y')' \sim \mathsf{P}_1$, 2. $U|X \sim \mathscr{B}(\pi)^{\otimes n}$, 3. $U \perp Y | X$.

Define $X^* \equiv X$, $\tilde{Y}_i \equiv -\frac{\pi}{1-\pi}\theta_1(X_i)$ and $Y_i^* \equiv U_i Y_i + (1-U_i)\tilde{Y}_i$, $i = 1, \ldots, n$. Let P_0^* be the distribution of $(X_i^{*\prime}, Y_i^*)'$. Under 2 and 3 we have $\mathsf{E}(Y_i^*|X^*) = \mathsf{E}(Y_i^*|W_i^*) = 0$. The fact that $(X_1^{*\prime}, Y_1^*)', \ldots, (X_n^{*\prime}, Y_n^*)'$ are i.i.d. follows from the same argument as in the proof of Proposition 7.5 above. Thus $\mathsf{P}_0^* \in \mathscr{P}_W$. QED

B Lemmas

Lemma B.1 (Sufficient conditions for BSE2) *In addition to BSE1, assume the following conditions hold:*

BSE6. $\forall \pi \in]0,1[, \forall \mathsf{P}_1 \in \mathscr{P}, \forall \mathsf{P}_2 \in \mathscr{P}, \pi\mathsf{P}_1 + (1-\pi)\mathsf{P}_2 \in \mathscr{P}$.

BSE7. $\forall \lambda \in [1, +\infty[$, $\forall \theta_1 \in \Theta$, $\forall \theta_2 \in \Theta$, $\lambda\theta_1 + (1-\lambda)\theta_2 \in \Theta$.

BSE8. $\forall \pi \in]0,1[, \forall \mathsf{P}_1 \in \mathscr{P}, \forall \mathsf{P}_2 \in \mathscr{P}$ *we have* $\theta\big(\pi\mathsf{P}_1 + (1-\pi)\mathsf{P}_2\big) = \pi\theta(\mathsf{P}_1) + (1-\pi)\theta(\mathsf{P}_2)$.

Then BSE2 holds.

Proof: Choose any two elements θ_1 and θ_2 of Θ and any $\pi \in]0,1[$. Under BSE7, we can find $\tilde{\theta} \in \Theta$ such that $\theta_2 = \pi\theta_1 + (1-\pi)\tilde{\theta}$. Under BSE1, $\mathscr{P}(\theta_1)$ and $\mathscr{P}(\tilde{\theta})$ are not empty. Then we can find $\mathsf{P}_1 \in \mathscr{P}(\theta_1)$ and $\tilde{\mathsf{P}} \in \mathscr{P}(\tilde{\theta})$. Define $\mathsf{P}_2 \equiv \pi\mathsf{P}_1 + (1-\pi)\tilde{\mathsf{P}}$. From BSE6, $\mathsf{P}_2 \in \mathscr{P}$. Now BSE8 implies $\theta(\mathsf{P}_2) = \pi\theta(\mathsf{P}_1) + (1-\pi)\theta(\tilde{\mathsf{P}}) = \pi\theta_1 + (1-\pi)\theta(\tilde{\mathsf{P}}) = \theta_2$. Hence $\mathsf{P} \in \mathscr{P}(\theta_2)$. This holds for any θ_1 and θ_2 in Θ, any $\mathsf{P}_1 \in \theta$ and any $\pi \in]0,1[$. Thus BSE2 is true. QED

Lemma B.2 (Generalization of BS's Lemma 1) *Let d be some strictly positive integer and P, Q and H be three probability distributions on $\mathbb{R}^d$ such that $\mathsf{Q} = \pi\mathsf{P} + (1-\pi)\mathsf{H}$ for some $\pi \in]0,1[$. For any integer $N \geq 1$, we have*

$$\sup_{g \in \mathcal{B}_{Nd}} \big|\mathsf{E}_{\mathsf{P}}(g_N) - \mathsf{E}_{\mathsf{Q}}(g_N)\big| \leq 1 - \pi^k, \quad \forall k \geq N,$$

where for any probability F on $\mathbb{R}^d$, $\mathsf{E}_{\mathsf{F}}(g_N) \equiv \int_{\mathbb{R}^{Nd}} g(t_1, \ldots, t_N)\mathrm{d}\mathsf{F}^{\otimes N}(t_1, \ldots, t_N)$.

Proof: Let $\mathscr{B}(\pi)$ denote the Bernoulli distribution with parameter π. Given P and H, define $\boldsymbol{\mu} \equiv \mathsf{P} \otimes \mathsf{H} \otimes \mathscr{B}(\pi)$. For any integer $N \geq 1$, we may always find a probability space $(\Omega, \mathscr{A}, \mathsf{m})$ and a real random vector $\boldsymbol{W} : \Omega \longrightarrow \mathbb{R}^{N(2d+1)}$ such that $\boldsymbol{W} \sim \boldsymbol{\mu}^{\otimes N}$. Thus we may write $\boldsymbol{W} = (W_1', \ldots, W_N')'$, where $W_1, \ldots, W_N$ are random real $(2d+1)$-vectors i.i.d. $\boldsymbol{\mu}$. In turn, we may write $W_i = (X_i', Y_i', U_i)'$, where $X_i \sim \mathsf{P}$, $Y_i \sim \mathsf{H}$, $U_i \sim \mathscr{B}(\pi)$ and X_i, Y_i, U_i independent, $i = 1, \ldots, N$. Now define $Z_i \equiv U_i X_i + (1 - U_i) Y_i$, $i = 1, \ldots, N$. It is easy to see that $Z_1, \ldots, Z_N$ are i.i.d. Q. Next define the event $A \equiv \bigcap_{i=1}^N A_i$, where $A_i \equiv \{\omega : U_i(\omega) = 1\} \in \mathscr{A}$, $i = 1, \ldots, N$. Notice $A_1, \ldots, A_N$ are independent, each with probability $\mathsf{m}(A_1) = \pi$. Thus $\mathsf{m}(A) = \pi^N$. Let $A^{\complement} \equiv \Omega \setminus A$. We have

$$\mathsf{E}_{\mathsf{Q}}(g_N) = \int_{\mathbb{R}^{Nd}} g(t_1, \ldots, t_N) \mathrm{d}\mathsf{Q}^{\otimes N}(t_1, \ldots, t_N) = \int_\Omega (g \circ Z)(\omega) \mathrm{d}\mathsf{m}(\omega)$$
$$= \int_A (g \circ Z)(\omega) \mathrm{d}\mathsf{m}(\omega) + \int_\Omega (g \circ Z)(\omega) \mathrm{I}_{A^{\complement}}(\omega) \mathrm{d}\mathsf{m}(\omega).$$

Because $0 \leq g \leq 1$ always, we have $(g \circ Z)\mathrm{I}_{A^{\complement}} \leq \mathrm{I}_{A^{\complement}}$. Thus

$$\int_\Omega (g \circ Z)(\omega) \mathrm{I}_{A^{\complement}}(\omega) \mathrm{d}\mathsf{m}(\omega) \leq \int_\Omega \mathrm{I}_{A^{\complement}}(\omega) \mathrm{d}\mathsf{m}(\omega) = \mathsf{m}(A^{\complement}) = 1 - \pi^N.$$

Notice that from the definition of $Z_1, \ldots, Z_N$, we have $Z(\omega) = X(\omega)$ for all $\omega \in A$. Therefore $\int_A (g \circ Z)(\omega) \mathrm{d}\mathsf{m}(\omega) \int_A (g \circ X)(\omega) \mathrm{d}\mathsf{m}(\omega)$. As $0 \leq g \leq 1$ always,

$$\int_A (g \circ X)(\omega) \mathrm{d}\mathsf{m}(\omega) \leq \int_\Omega (g \circ X)(\omega) \mathrm{d}\mathsf{m}(\omega) = \mathsf{E}_{\mathsf{m}}(g \circ X) = \mathsf{E}_{\mathsf{P}}(g_N).$$

Then $\mathsf{E}_{\mathsf{Q}}(g_N) \leq \mathsf{E}_{\mathsf{P}}(g_N) + 1 - \pi^N$. As this inequality is true for any $g \in \mathcal{B}_{Nd}$, it is also true for $1 - g$ and thus $\mathsf{E}_{\mathsf{P}}(g_N) - \mathsf{E}_{\mathsf{Q}}(g_N) \leq 1 - \pi^N$. Combining these results, we get

$$\big|\mathsf{E}_{\mathsf{Q}}(g_N) - \mathsf{E}_{\mathsf{P}}(g_N)\big| \leq 1 - \pi^N.$$

As this holds for all $g \in \mathcal{B}_{Nd}$, the desired result follows. QED

Lemma B.3 *Let $(\Omega, \mathscr{A}, \mathsf{m})$ be a probability space and N and d be strictly positive integers. Let $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$ be N real random d-vectors defined on Ω. Let U_i be a Bernoulli $\mathscr{B}(\pi_i)$ random variable defined on Ω, for some $\pi_i \in]0, 1[$, $i = 1, \ldots, N$. Define the N random d-vectors $Z_i \equiv U_i X_i + (1 - U_i) Y_i$, $i = 1, \ldots, N$. Then*

$$\sup_{g \in \mathcal{B}_{Nd}} \big|\mathsf{E}_{\mathsf{m}}\big(g(Z)\big) - \mathsf{E}_{\mathsf{m}}\big(g(X)\big)\big| \leq \sum_{i=1}^N (1 - \pi_i),$$

where $\mathsf{E}_{\mathsf{m}}\big(g(X)\big) \equiv \int_\Omega g\big(X_1(\omega), \ldots, X_N(\omega)\big) \mathrm{d}\mathsf{m}(\omega)$. If $U_1, \ldots, U_N$ are independent, the RHS of the inequality may be replaced with $1 - \prod_{i=1}^N \pi_i$.

Proof: Define $A \equiv \bigcap_{i=1}^{N}\{U_i(\omega)=1\}$ and $A^{\complement}$ its complement in Ω. Notice that

$$\mathsf{m}(A^{\complement}) \le \sum_{i=1}^{N} \mathsf{m}\big(\{U_i(\omega)=0\}\big) = \sum_{i=1}^{N}(1-\pi_i). \qquad \text{(A.1)}$$

Thus for any $g \in \mathcal{B}_{Nd}$, and using the same arguments as in Lemma B.2, we have

$$\mathsf{E}_{\mathsf{m}}\big(g(Z)\big) = \int_A g\big(Z(\omega)\big)\mathsf{dm}(\omega) + \int_{A^{\complement}} g\big(Z(\omega)\big)\mathsf{dm}(\omega) \le \mathsf{E}_{\mathsf{m}}\big(g(X)\big) + \mathsf{m}(A^{\complement}).$$

Inequality (A.1) implies $\mathsf{E}_{\mathsf{m}}\big(g(Z)\big)-\mathsf{E}_{\mathsf{m}}\big(g(X)\big) \le \sum_{i=1}^{N}(1-\pi_i)$. As this is true for any $g \in \mathcal{B}_{Nd}$, it is also true for $1-g$ and thus $\mathsf{E}_{\mathsf{m}}\big(g(X)\big)-\mathsf{E}_{\mathsf{m}}\big(g(Z)\big) \le \sum_{i=1}^{N}(1-\pi_i)$. Combining these results, we get $\big|\mathsf{E}_{\mathsf{m}}\big(g(Z)\big) - \mathsf{E}_{\mathsf{m}}\big(g(X)\big)\big| \le \sum_{i=1}^{N}(1-\pi_i)$. Now when $U_1,\ldots,U_n$ are independent, $\mathsf{m}(A^{\complement}) = 1-\prod_{i=1}^{n}\pi_i$. As this holds for all $g \in \mathcal{B}_{Nd}$, the desired result follows. QED

Lemma B.4 *Let A and B be two subsets of a set S, with $A \subset B$. Let d be a pseudo-distance on S. Let h be a real function defined on B. Continuity of h w.r.t. d is defined in the same way as when d is a distance. Also, "A* is dense in *B" is defined as when d is a distance.*

If (w.r.t. d) A is dense in B and h is continuous, then $\sup_A h(x) = \sup_B h(x)$ *and* $\inf_A h(x) = \inf_B h(x)$.

Proof: As h is continuous, $\forall b \in B$, $\forall \epsilon \in]0;+\infty[$, $\exists \delta \in]0;+\infty[$ such that $x \in B_\delta^d(b) \implies |h(x)-h(b)| < \epsilon$. As A is dense in B, the neighborhood $B_\delta^d(b)$ necessarily contains a $a \in A$. This implies that

$$\forall b \in B,\ \forall \epsilon \in]0;+\infty[,\ \exists a \in A \text{ such that } |h(a)-h(b)| < \epsilon. \qquad \text{(A.2)}$$

As $A \subset B$, suppose $\sup_A h(x) < \sup_B h(x)$ and define $\eta \in]0;+\infty[$ as $\eta \equiv \sup_B h(x) - \sup_A h(x)$. Then clearly $\sup_B h(x) - h(a) > \epsilon$ for all $a \in A$ and all $\epsilon \in]0;\eta[$. Hence $\exists b \in B$ such that $h(b)-h(a) \ge \epsilon$, $\forall \epsilon \in]0;\eta[$, $\forall a \in A$, which contradicts (A.2). Now suppose $\inf_A h(x) > \inf_B h(x)$ and define $\nu \in]0;+\infty[$ as $\nu \equiv \inf_A h(x) - \inf_B h(x)$. Then $h(a) - \inf_B h(x) > \epsilon$ for all $a \in A$ and all $\epsilon \in]0;\nu[$. Hence $\exists b \in B$ such that $h(a)-h(b) \ge \epsilon$, $\forall \epsilon \in]0;\nu[$, $\forall a \in A$, which also contradicts (A.2). QED

Lemma B.5 *Let $\mathscr{P}$ be a statistical model and define the testing problem $H = (H_0, H_1)$, where $H_k : \mathsf{P} \in \mathscr{P}_k$, with $\varnothing \ne \mathscr{P}_k \subset \mathscr{P}$, $k = 0,1$, and $\mathscr{P}_1 = \mathscr{P} \backslash \mathscr{P}_0$. Assume that for some index set I we may write $\mathscr{P} = \bigcup_{i \in I} \mathscr{P}^i$, where the $\mathscr{P}^i$s are subsets of $\mathscr{P}$. For any $i \in I$, define $\mathscr{P}_k^i \equiv \mathscr{P}_k \cap \mathscr{P}^i$, $k = 0,1$ and the testing problem $H^i = (H_0^i, H_1^i)$, where $H_k^i : \mathsf{P} \in \mathscr{P}_k^i$, $k = 0,1$. If for any $i \in I$, H^i is non-testable, then H is non-testable.*

Proof : Let φ be an α-level test of H. For any $i \in I$, as $\mathscr{P}_0^i \subset \mathscr{P}_0$, φ is also an α-level test of H^i. Since for any $i \in I$ it is assumed H^i is non-testable, we have $\sup_{\mathsf{P}\in\mathscr{P}_1^i} \mathsf{E}_\mathsf{P}(\varphi_n) \leq \alpha$, $\forall i \in I$, which entails $\sup_{i\in I} \sup_{\mathsf{P}\in\mathscr{P}_1^i} \mathsf{E}_\mathsf{P}(\varphi_n) \leq \alpha$. Now, $\mathscr{P}_1 = \bigcup_{i\in I} \mathscr{P}_1^i$ and the LHS of the last inequality is $\sup_{\mathsf{P}\in\mathscr{P}_1} \mathsf{E}_\mathsf{P}(\varphi_n)$. As $\sup_{\mathsf{P}\in\mathscr{P}_1} \mathsf{E}_\mathsf{P}(\varphi_n) \leq \alpha$ holds for any α-level test of H, this problem is non-testable. QED

References

Adichie, J. N. (1967). Asymptotic efficiency of a class of nonparametric tests for regression scores. *Annals of Mathematical Statistics*, 38:884–893.

Aït-Sahalia, Y. Stoker, T. M., and Bickel, P. J. (2001). Goodness-of-fit tests for regression using kernel methods. *Journal of Econometrics*, 105:363–412.

Anderson, T. W. (1967). Confidence limits for the expected value of an arbitrary bounded random variable with a continuous distribution function. *Bulletin of the International Statistical Institute*, 43:249–251.

Andrews, D. W. K. (1994). Empirical process methods in econometrics. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, chapter 37. North Holland.

Arellano, M. and Honoré, B. (2001). Panel data models : Some recent developments. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, chapter 53. North Holland.

Bahadur, R. R. and Savage, L. J. (1956). Nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, 25:1115–1122.

Bickel, P. J. (1971). Analogs of linear combinations of order statistics in the linear model. In *Statistical Decision Theory and Related Topics*. Academic Press.

Blough, S. R. (1992). The relationship between power and level for generic unit root tests in finite samples. *Journal of Applied Econometrics*, 7:295–308.

Davidson, R. and MacKinnon, J. G. (1999). Bootstrap testing in nonlinear models. *Econometric Theory*, 15:361–376.

Devroye, L. and Lugosi, G. (2002). Almost sure classification of densities. *Journal of Nonparametric Statistics*, 14:675–698.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics, with applications to structural and dynamic models. *Econometrica*, 65:1365–1389.

Dufour, J.-M. (2003). Identification, weak instruments and statistical inference in econometrics, presidential address to the canadian economics association. *Canadian Journal of Economics*, 36:767–808.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, 64:865–890.

Faust, J. (1999). Conventional confidence intervals for points on spectrum have confidence level zero. *Econometrica*, 67:629–637.

Forchini, G. (2005). Some properties of tests for possibly unidentified parameters. Working Paper 21/05, Departement of Econometrics and Business Statistics, Monash University.

Gleser, L. and Hwang, J. (1987). The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter. *Annals of Statistics*, 15:1351–1362.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Härdle, W. and Linton, O. (1994). Applied nonparametric methods. In Engle, R. F. and .McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, chapter 38. North-Holland.

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press.

Jureckova, J. (1971). Nonparametric estimates of regression coefficients. *Annals of Mathematical Statistics*, 42:1328–1338.

Lavergne, P. and Vuong, Q. (2000). Nonparametric significance testing. *Econometric Theory*, 16(4):576–601.

Lee, M. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. Springer.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. Statistics/Probability Series. Wadsworth & Brooks/Cole, 2nd edition.

Lehmann, E. L. and D'Abrera, H. J. M. (1998). *Nonparametrics: Statistical Methods Based on Ranks*. Prentice-Hall, rev. ed. edition.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, chapter 42. North Holland.

Pfanzagl, J. (1998). The nonexistence of confidence sets for discontinuous functionals. *Journal of Statistical Planning and Inference*, 75:9–20.

Pötscher, B. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots and estimation of long memory parameters. *Econometrica*, 70:1035–1066.

Powell, J. L. (1994). Estimation of semiparametric models. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, chapter 41. North Holland.

Puri, M. L. and Sen, P. K. (1985). *Nonparametric Methods in General Linear Models*. John Wiley & Sons.

Romano, J. P. (2004). On nonparametric testing, the uniform behavior of the t-test and related problems. *Scandinavian Journal of Statistics*, 31(4):567–584.

Romano, J. P. and Wolf, M. (2000). Finite sample nonparametric inference and large sample efficiency. *Annals of Statistics*, 28:756–778.

Serfling, R. (1980). *Approximations Theorems of Mathematical Statistics*. Wiley.

Sherman, R. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61:123–137.

Tibshirani, R. and Wasserman, L. A. (1988). Sensitive parameters. *Canadian Journal of Statistics*, 16(2):185–192.