On the Nonexistence of Finite-Variance Unbiased Estimators

David.Pacini@bristol.ac.uk

5 July 2023

Preliminary Draft

Comments are Welcome[1]

Abstract

This paper investigates the nonexistence of unbiased estimators with finite variance in parametric statistical models used in empirical economic studies. It presents two results. The first result is a sufficient condition for the nonexistence of finite-variance unbiased estimators. The second result is a method for verifying this sufficient condition. The application of this method gives support to the view that allowing for the possibility of overparametrizations, such as those that would occur when modelling instrumental variables, self-selection, endogeneity, or skewness, before knowing that they are not features of the data, can lead to the impossibility of constructing finite-variance unbiased estimators. The nonexistence of finite-variance unbiased estimators motivates considering alternative criteria for evaluating estimators. As a by-product, the second result gives a test of local identifiability with exact size control.

Keywords: Estimation, Unbiasedness, Minimum Variance, Metric Derivative, Over-parametrization Test.
JEL Codes: C13, C24, C26, C34.

§1. Introduction

Parametric statistical models serve to produce evidence relevant to economic analysis, among other uses. They offer the possibility of couching magnitudes representing economic behaviour, such as elasticities, as parameters of density functions. The practical use of these models requires choosing an estimator with desirable statistical properties for the parameters. A standard choice is the maximum likelihood estimator. This estimator can differ in a systematic manner from parameters; i.e., the maximum likelihood estimator can be

[1] I would like to thank Santiago Oliveros, Sami Stouli, and Pietro Spini and participants at the 2023 Universita di Siena Workshop on Econometric Theory and Applications, the IAAE 2023 Annual Conference for useful comments and discussions.

biased. An alternative is to convert a biased maximum likelihood estimator into an unbiased estimator using a debiasing method; see, e.g., Firth (1993), Müller and Wang (2019). What are parametric statistical models in which trying the conversion from biased to unbiased estimators is not desirable? Addressing this question is relevant, first, for justifying bias as a feature of certain modeling requirements, rather than as a consequence of a poor choice of an estimator, and, second, for motivating alternative criteria, other than unbiasedness, for choosing estimators. The first order of business is to deal with the problem of existence. If no unbiased estimator does exist, Doss and Sethuraman (1989) show that any nearly unbiased estimator can yield erratic estimates, and consequently produce evidence of unreliable quality, due to an inevitable nearly infinite variance. Little is known, however, about the nonexistence of unbiased estimators in parametric statistical models employed in economic studies, except for linear instrumental variable models; see, e.g., Hirano and Porter (2015).

This paper investigates the nonexistence of finite-variance unbiased estimators in a class of parametric density models in use by empirical economic studies. The class includes models that have not yet been examined in the literature dealing with the construction of finite-variance unbiased estimators. Examples are the normal self-selection model described in Gronau (1974); the normal bivariate binary response model with endogeneity in Heckman (1978) and employed by, e.g., Evans and Schwab (1995); a variant of this model with endogenous switching in Lokshin and Glinskaya (2009); the normal trivariate regression model with endogenous switching in Helpman et al. (2017); and the skew-normal binary response model in Stingo, Stanghellini and Capobianco (2012). It also includes the normal linear instrumental variable model, which has been largely studied in the literature; see, e.g., Hirano and Porter (2015) and Phillips (1980). This paper presents two results.

The first result shows that a sufficient condition for the nonexistence of finite-variance unbiased estimators is the presence of *Hellinger speedless points* in the parameter space. Models featuring this type of point have, for at least one of the model's parameters, no finite-variance unbiased estimator. This result provides motivation for considering alternative criteria, other than minimum-variance unbiasedness, when evaluating estimators in models with *Hellinger speedless points*. This result is related to, but not a particular case of, existing impossibility results for models with *nonidentifiable* points, see, e.g., Gourieroux and Monfort (1995, Chapter 6), or models with *singularities* in the mapping from reduced form to structural parameters, see, e.g., Hirano and Porter (2015). Hellinger speedless points are vanishing points of the Hellinger pseudo-distance derivative. Nonidentifiable points are observational equivalent points in the parameter space. Singularities are vertical asymptotes of the mapping from reduced form to structural parameters. I use a normal squared location model to illustrate the difference between the three. The reduced form parameter is the location, and the structural parameter is the squared location. The mapping from reduced form to structural parameters is the square function. In this illustration, zero is a Hellinger speedless point, but it is neither nonidentifiable nor a singularity of the mapping from reduced form to structural parameters. I use this example because it serves well as an illustration. I understand that it is of little practical relevance. The results of this paper concern more interesting models for empirical economic analysis, such as the parametric density models used by Lokshin and Glinskaya (2009), Evans and Schwab (1995), and Gronau (1974).

The second result is a method for detecting Hellinger speedless points. For quadratic mean differentiable models, I show that a point in the parameter space is Hellinger speedless if and only if the least eigenvalue of the Fisher matrix evaluated at that point is zero. In models with a closed-form Fisher matrix, e.g., the normal self-selection model, the detection of Hellinger speedless points only requires checking the existence of a solution to a system of linear equations. In models without a closed-form Fisher matrix, e.g., the normal bivariate binary response model with endogenous switching, the detection of Hellinger speedless points requires further elaboration. Naive methods, such as comparing the least eigenvalue of the sample Fisher matrix with zero or with a parametric bootstrap critical value, fail to control the probability of incorrectly rejecting the hypothesis that a point in the parameter space is Hellinger speedless. For these models, I propose a randomization test. The test uses the least eigenvalue of a simulated sample Fisher matrix as the test statistic. It exploits the invariance of the set of Hellinger speedless points to scalar reparametrizations for controlling the probability of incorrectly rejecting the hypothesis that a point in the parameter space is Hellinger speedless. As a by-product, when the point being tested is a regular point of the Fisher matrix in the sense defined by Rothenberg (1971), the test becomes a local identifiability test with exact size control. Alternative methods for constructing tests on eigenvalues exist in the literature (see e.g., Anderson, 2003; Cragg and Donald, 1993; Chen and Fang, 2019). These methods, if tailored to the problem of detecting Hellinger speedless points, could guarantee large-sample size control. The randomization test, by contrast, guarantees finite-sample size control. One could nevertheless use insights from these alternative methods to motivate the choice of a randomization test. Anderson (2003) shows that the sampling distribution of sample eigenvalues changes with the algebraic multiplicity of the eigenvalue of interest, which in my case is an unknown nuisance parameter. Since asymptotic or parametric bootstrap tests are, in general, unable to control size when the algebraic multiplicity of the zero eigenvalue is unknown, I consider using a randomization test instead.

The application of these results reveals that the nonexistence of finite-variance unbiased estimators can arise because of using overparametrizing models. The Fisher matrix in the normal self-selection model has a zero eigenvalue at the point in the parameter space representing absence of selection; see, e.g., Lee and Chesher (1986). The Fisher matrix in the skew-normal binary response model has a zero eigenvalue at the point representing absence of skewness; see, e.g., Stingo et al. (2012). The Fisher matrix in the normal bivariate binary response model with endogenous switching has a zero eigenvalue at points representing an irrelevant instrument, absence of endogeneity, or both. Consequently, there is no finite-variance unbiased estimator in these models. One can still choose a biased estimator, such as the maximum likelihood estimator, and then apply a bias-reduction method. The nonexistence of finite-variance unbiased estimators does not make this choice inappropriate. It rather suggests that, in these models, bias should not be pushed to zero.

Additional existing impossibility results that are applicable to parametric models are in van der Vaart (1991), on the nonexistence of regular estimators in models with a singular Fisher matrix; Liu and Brown (1993), on the nonexistence of finite-variance unbiased estimators for functions of parameters that are not uniformly continuous; Dufour (1997), on the nonexistence of valid confidence intervals with finite length in models with nonidentifiable points in the parameter space; Hirano and Porter (2012), on the nonexistence of regular estimators for nondifferentiable functions of parameters; and in Kaji (2021), on the nonexistence of equivariant-in-law estimators for weakly regular parameters. My randomization test is not a particular case of the results in these papers. One can use the results in van der Vaart (1991) to additionaly view my randomization test as a numerical method, useful in models without closed-form Fisher matrix, for predicting the nonexistence of regular estimators.

I organize the rest of the paper as follows. Section 2 presents motivating examples for the theoretical developments in each of the subsequent sections. Section 3 sets up a framework encompassing the motivating examples. It defines the notion of a Hellinger speedless point within this framework, and it presents the nonexistence of finite-variance unbiased estimators in models with Hellinger speedless points. Section 4 discusses the meaning of Hellinger speedless points in quadratic mean differentiable models, and it compares Hellinger speedless points with nonidentifiable points and singularities of the mapping from reduced form to structural parameters. Section 5 describes the randomization test for detecting Hellinger speedless points and establishes the validity of the test for controlling size in finite samples. Section 6 concludes. There are four appendices. Appendix A contains the proofs of the propositions in the text. Appendix B derives the score in the binary bivariate switching model. Appendix C establishes the impossiblity to construct regular estimators, as defined by Van der Vaart (1991), for Hellinger speedless points. Appendix D constructs an alternative resampling test for detecting Hellinger speedless points.

§2. Motivating Examples

I first give examples of models illustrating the issues I want to investigate and the challenges that they create. I set out, in the next Section, a framework that encompasses these examples and other parametric density models in use by empirical economic studies.

Example 1 (Normal Linear Instrumental Variable Model). Let $Y_i = (Y_{1i}, Y_{2i})$ denote a bivariate continuous random variable. Consider the specification

$$\begin{aligned} Y_{1i} &= \beta_0 Y_{2i} + U_{1i} \\ Y_{2i} &= \gamma_0 x_i + U_{2i} \end{aligned}, \text{ where } U_i|x_i \sim \mathcal{N}\begin{pmatrix} 0 & 1 & .5 \\ & , & \\ 0 & .5 & 1 \end{pmatrix}$$

with $U_i = (U_{1i}, U_{2i})$. This is a simplified parametric version of the linear triangular instrumental variable model. Consider the problem of constructing a minimum-variance unbiased estimator for β from a random sample $\{Y_i, x_i\}_{i=1}^N$. There are at least two strategies to show the nonexistence of such an unbiased estimator. The first strategy uses the observation that, see e.g., Hirano and Porter (2015), vertical asymptote in the mapping from reduced to structural parameters do not have unbiased estimators. Set $\theta = (\beta, \gamma)$ and $\pi = (\pi_1, \pi_2)$ with $\pi_1 = \beta\gamma$ and $\pi_2 = \gamma$. The mapping from reduced form to the structural parameter of interest is $\pi \to \kappa(\pi) = \pi_1/\gamma = \beta\gamma/\gamma$, so $\beta = \kappa(\theta)$. It has a vertical asymptote at $\gamma_0 = 0$. The second strategy uses the observation that, see e.g. Gourieroux and Monfort (1995, Chapter 6, Proposition 6.2, p. 129), nonidentifiable points do not have unbiased estimators. The specification in this example has an identifiability failure at $(\beta_0, \gamma_0) \in ((-\infty, \infty), 0)$. By contrast, I seek to establish the nonexistence of finite-variance unbiased estimators without having to check whether there is an identifiability failure or a vertical asymptote. My motivation comes from the following examples.

Example 2 (Normal Squared Location Model). This example is of little empirical interest but it will serve to illustrate sources of nonexistence of finite-variance unbiased estimators other than identifiability failures or vertical asymptotes. Let Y_i denote a continuous random variable. The specification is

$$Y_i = \theta_0^2 + U_i, \text{ where } U_i \sim \mathcal{N}(0, 1).$$

Consider the problem of constructing a minimum variance unbiased estimator for θ_0 from a random sample $\{Y_i\}_{i=1}^N$. The mapping $\theta \to \kappa(\theta) = \theta^2$ does not have a vertical asymptote. Any $\theta \in \mathbb{R}$ is a locally identifiable point, see e.g., Gourieroux and Monfort (1995, Chapter 3, Example 3.16, p. 89). The smallest Hellinger pseudo-distance derivative passing through $\theta_0 = 0$, which is defined in the next section, is zero. Section 3 below shows that this last observation implies that there is no finite-variance unbiased estimator for $\theta_0 = 0$ even when $\theta_0 = 0$ is locally identifiable and the mapping $\theta \to \theta^2$ does not have vertical asymptotes.

Example 3 (Normal Self-Selection Model). Let $Y_i^\star = (Y_{1i}^\star, Y_{2i}^\star)$ denote a latent bivariate continuous random variable. Consider the specification

$$\begin{array}{rc} Y_{1i}^\star = & x_{1i}\beta_0 + \sigma_1 U_{1i} \\ Y_{2i}^\star = & x_{1i}\gamma_1 + x_{2i}\gamma_2 - U_{2i} \end{array}, \text{ where } U_i | x_i \sim \mathcal{N}\begin{pmatrix} 0 & 1 & \rho_0 \\ & , & \\ 0 & \rho_0 & 1 \end{pmatrix}$$

with $U_i = (U_{1i}, U_{2i})$ and $x_i = (x_{1i}, x_{2i})$. The variables $Y_{1i}^\star$ and $Y_{2i}^\star$ are not observed. Instead, we observe

$$Y_{1i} = Y_{1i}^{\star} Y_{2i}$$
$$Y_{2i} = 1(Y_{2i}^{\star} > 0).$$

The specification in this example was proposed by Gronau (1974) to analyze reservation wages. It has been used since then in several different applications, including studies in health economics, see e.g., Jones (2000) for a review, and in empirical corporate finance, see e.g. Li and Prabhala (2007) for a review.

Consider the problem of constructing a minimum-variance estimator for β_0 from a random sample $\{Y_{1i}, Y_{2i}, x_i\}_{i=1}^{N}$. The parameter β_0 is identifiable. Section 4 below shows that there is no finite-variance unbiased estimator for β_0 without checking whether $\theta \to \kappa(\theta) = \beta_0$ has a vertical asymptote. I obtain this result from examining below the smallest Hellinger pseudo-distance derivative for curves passing through $(\beta_0, \gamma_1, \gamma_2 = 0, \sigma_1, \rho_0 = 0)$.

Example 4 (Normal Bivariate Binary Response with Endogenous Switching). This example will serve to motivate the introduction of a statistical test in Section 5. In this example, the Hellinger pseudo-distance derivative is not available in closed-form. Let $Y_i^{\star} = (Y_{1i}^{\star}, Y_{2i}^{\star})$ denote a latent bivariate continuous random variable. Consider the latent specification

$$\begin{array}{rl} Y_{1i}^{\star} = & x_{1i}\beta_1 + Y_{2i}\beta_2 + \tilde{x}_{1i} Y_{2i}\beta_3 - U_{1i} \\ Y_{2i}^{\star} = & x_{1i}\gamma_1 + x_{2i}\gamma_2 - U_{2i} \end{array}, \text{ where } U_i | x_i \sim \mathcal{N}\begin{pmatrix} 0 & 1 & \rho_0 \\ & , & \\ 0 & \rho_0 & 1 \end{pmatrix}$$

with $U_i = (U_{1i}, U_{2i})$, $x_i = (x_{1i}, x_{2i})$, and $x_{1i} = (1, \tilde{x}_{1i})'$. The variables $Y_{1i}^{\star}$ and $Y_{2i}^{\star}$ are not observed. Instead, we observe

$$Y_{1i} = 1(Y_{1i}^\star > 0)$$
$$Y_{2i} = 1(Y_{2i}^\star > 0).$$

Lokshin and Glinskaya (2009) use this specification to examine the effect of seasonal migration decisions (Y_1) by a husband on labor market participation decisions (Y_2) by his wife using data for Nepali households. The interactive covariate $Y_{2i}\tilde{x}_{1i}$ captures differences in home productivity $\tilde{x}_{1i}$ in households with and without a migrant. The disturbances U_{1i} and U_{2i} capture unobserved variables affecting each decision. Nonnegative correlation between the disturbances makes Y_{2i} to be correlated with U_{1i}, i.e., the migration decision and the interactive covariate are endogenous covariates. For the simplest case when $\tilde{x}_{1i}$ and x_{2i} have only one component, this model has one instrumental variable (x_{2i} measuring migration costs) and two endogenous covariates (Y_{2i} and $Y_{2i}\tilde{x}_{1i}$). The case with $\beta_3 = 0$ corresponds to the normal bivariate binary response model with endogeneity employed by, e.g., Evans and Schwab (1995) to analyze the effect of attending a catholic school on college completion.

Consider the problem of constructing a minimum-variance unbiased estimator for β_1 and $\beta_1 + \beta_3$ from a random sample $\{Y_{1i}, Y_{2i}, x_i\}_{i=1}^N$. Identifiability of β_1 or vertical asymptotes of $\theta \to \kappa(\theta) = \beta_1$ have either only investigated for the special case $\beta_3 = 0$ or have not yet been investigated in the literature. I show that there is no finite-variance unbiased estimator for β_1, without checking whether β_1 is identifiable or $\theta \to \kappa(\theta) = \beta_1$ has a vertical asymptote. I obtain this result using below a statistical test for the null hypothesis that the smallest Hellinger pseudo-metric derivative for curves passing through $(\beta, \gamma_1, \gamma_2 = 0, \rho_0 = 0)$ is zero.

§3. Nonexistence of Finite-Variance Unbiased Estimators

This Section sets up a statistical model that encompasses the examples in the previous Section. It defines the notion of a Hellinger speedless point within this model and it works out the consequences on the nonexistence of finite-variance unbiased estimators from the presence of these points.

Parametric Model. The observations consist of independent and identically distributed random vectors $Y_1, .., Y_N$ with values in a subset $\mathcal{Y}$ of the Euclidean space. The random vector Y_1 has a probability function P_{θ_0}. The parameter θ_0 belongs to the parameter space Θ. It has two components $\theta_0 = (\beta_0, \gamma_0)$. The parameter space $\Theta = B \times \Gamma$ is an open subset of $\mathbb{R}^2$. Assume that P_θ has a positive density function with respect to a σ-finite measure μ. To interchange limits and integrals when required, I also assume that there exists an integrable function $\bar{g}$ such that $f_\theta < \bar{g}^2$. The parametric model is the family of densities

$$\mathcal{F}_\Theta := \left\{ f_\theta : \mathcal{Y} \to \mathbb{R},\ f_\theta > 0,\ \int f_\theta d\mu = 1,\ f_\theta < \bar{g}^2,\ \int \bar{g}^2 d\mu < \infty,\ \theta \in \Theta \right\}.$$

One could work with the case when the parameter space is a more general space. However, the results are more easily described in the specific case $\Theta \subseteq \mathbb{R}^2$.

Estimation Problem. The parameter θ_0 is unknown. An unbiased estimator for β_0 is a measurable function $\hat{\beta}_N : \mathcal{Y}^N \to B$ from the sample space into the parameter space such that

$$\int \hat{\beta}_N \underbrace{f_\theta \times ... \times f_\theta}_{\text{N times}} d\mu = \beta \text{ for every } \theta \in \Theta.$$

I consider the problem of finding an unbiased estimator with the smallest variance among all unbiased estimators. The minimum variance for unbiased estimators may not be finite. Extreme estimates are likely to be encountered if such is the case. These extreme estimates could lead to distorted discoveries. This Section aims to clarify this phenomenon by using a measure of the ability of the model to distinguish θ_0 from relevant nearby points in the parameter space.

Hellinger Speedless Points. The square of the Hellinger distance between the densities f_θ and f_{θ_0} is

$$\rho(f_\theta, f_{\theta_0})^2 := \int \frac{\left|f_\theta^{1/2} - f_{\theta_0}^{1/2}\right|^2}{2} d\mu.$$

It arises from adding up (over the different realizations y of the random variable Y_i) the areas of one of the two equivalent equilateral triangles making the square with side lenght $|f_\theta^{1/2}(y) - f_{\theta_0}^{1/2}(y)|$. The Hellinger pseudo-distance between the points θ and θ_0 is

$$h(\theta, \theta_0) := \sqrt{\rho(f_\theta, f_{\theta_0})^2} = \sqrt{\int \frac{\left|f_\theta^{1/2} - f_{\theta_0}^{1/2}\right|^2}{2} d\mu}.$$

The Hellinger pseudo-distance takes values between zero and one. I call these values *Hellingers*. They are invariant to the choice of μ. There are functions other than the Hellinger distance to measure the distance between densities. I chose the Hellinger distance because it leads to the Theorem below about the non-existence of unbiased estimators with finite-variance. The metric properties of (Θ, h) are not the same as those of $(\mathcal{F}_\Theta, \rho)$. I am going to exploit this difference in the rest of the paper.

Let $c : [0, 1] \to \Theta$ denote a curve from the closed interval $[0, 1]$ to the parameter space that is differentiable on the open interval $(0, 1)$ and takes value θ_0 at zero. The curve $\epsilon \to c(\epsilon)$ is Hellinger pseudo-distance differentiable at zero if the limit

$$|\dot{h}_{\theta_0}|(c) := \lim_{\epsilon \to 0} \frac{h\big(c(\epsilon), \theta_0\big)}{|\epsilon|}$$

exists. I call $|\dot{h}_{\theta_0}|(c)$ the Hellinger pseudo-distance derivative of the function $\epsilon \to c(\epsilon)$. It is a particular case of the general concept of metric derivative for a curve studied by, e.g., Ambrosio, Gigli and Savare (2008, Chapter 1). Let $\mathcal{C}_{\theta_0}$ denote the set of Hellinger pseudo-distance differentiable curves. For $\epsilon = 1/N$, I interpret $|\dot{h}_{\theta_0}|(c)$ as the speed (in Hellingers per observation) when approaching θ_0 along the Hellinger pseudo-distance differentiable curve c. The Hellinger sensitivity at θ_0 is

$$v_{\theta_0} := \inf_{c \in \mathcal{C}_{\theta_0}} |\dot{h}_{\theta_0}|(c) = \inf_{c \in \mathcal{C}_{\theta_0}} \lim_{\epsilon \to 0} \frac{h\big(c(\epsilon), \theta_0\big)}{|\epsilon|}.$$

For $\epsilon = 1/N$, v_{θ_0} is the speed when approaching θ_0 along the slowest Hellinger pseudo-distance differentiable curve. The Hellinger sensitivity measures the ability of the model to distinguish θ_0 from nearby (in the Hellinger pseudo-distance) points in the parameter space.

Definition (Hellinger Speedless Point). A point θ_0 is Hellinger speedless if the Hellinger sensitivity at θ_0 is zero, i.e., $v_{\theta_0} = 0$.

At a Hellinger speedless point, there is a direction in the parameter space at which the Hellinger pseudo-distance function is flat because the pseudo-metric derivative is zero. A flat Hellinger pseudo-distance function cannot discriminate between nearby, in the Hellinger pseudo-distance sense, points in the parameter space. The definition of Hellinger speedless point does not presuppose that the mapping $\theta \mapsto f_\theta^{1/2}$ is differentiable, so it applies to models that may not be differentiable in quadratic mean.

Impossibility Result. The following result works out a negative consequence of the presence of Hellinger speedless points on the existence of minimum-variance unbiased estimators.

Theorem 1. An estimator $\hat{\beta}_N$ of β_0 cannot both be unbiased and have finite variance at any Hellinger speedless point.

This result follows from applying the Cauchy-Schwarz inequality to relate the Hellinger pseudo-distance derivative with the bias and variance of any estimator. The choice of the Hellinger distance induces an inner space structure, which, in turn, justifies the application of the Cauchy-Schwarz inequality.

§4. Characterizing Hellinger Speedless Points

I get further insight in this Section about Hellinger speedless points, and how to detect them, by considering models with *Fisher matrix*.

Definition (QMD Model). The density f_{θ_0} belongs to the family of QMD models if there is a measurable function $s_{\theta_0} : \mathcal{Y} \to \mathbb{R}^2$ such that

$$\left[\int (f_{\theta_0+\Delta}^{1/2} - f_{\theta_0}^{1/2} - \Delta^\top s_{\theta_0} f_{\theta_0}^{1/2})^2 d\mu \right]^{1/2} = o(\|\Delta\|_2) \text{ for any } \{\Delta \in \mathbb{R}^2\} \to 0 \text{ and any } \theta_0 \in \Theta.$$

The expectation of the outer product of g_θ is the Fisher matrix evaluated at θ:

$$\mathcal{I}_{\theta_0}(\theta) := \int s_\theta s_\theta^\top f_{\theta_0} d\mu.$$

Eigenvectors of the Fisher matrix represent the directions of the largest variance of the random vector s_{θ_0} while the eigenvalues represent the magnitude of this variance in those directions. The eigenvalues of the Fisher matrix depend on the particular parametrization chosen. If we reparametrize a model in terms of $\vartheta = \kappa(\theta)$ for a differentiable mapping κ, then the Fisher matrix scales by the square of of the derivative of κ. I use this observation in Section 5 below to construct a statistical test to detect Hellinger speedless points. The following result characterizes Hellinger speedless points in terms of the smallest eigenvalue of the Fisher matrix.

Proposition 1. Assume that f_{θ_0} belongs to the family of QMD models. Then, $\theta_0 \in \Theta$ is Hellinger speedless point if and only if the smallest eigenvalue of the Fisher matrix evaluated at θ_0 is zero.

According to Proposition 1, a Hellinger speedless point indicates that there is a direction in the parameter space at which the random vector s_{θ_0} has no variability. Three remarks are in order. First, Proposition 1 suggests that, for checking whether a given point in the parameter space is Hellinger speedless, one can consider testing the null hypothesis that the smallest eigenvalue of the Fisher matrix is zero versus the alternative that it is positive. I describe, in Section 5 below, a statistical test for these hypotheses.

Second, the following corollary to Proposition 1 follows as a direct application of Sard's Theorem.[2]

Corollary. The set of square-root densities with Hellinger speedless points has Lebesgue measure zero.

This Corollary indicates that, while the map $\theta \mapsto f_\theta^{1/2}$ may have many Hellinger speedless *points* in its domain, which is the parameter space, it must have critical *values* in its codomain, which is the manifold of square-root densities, with zero Lebesgue measure. This Corollary does not suggest that Hellinger speedless points should be neglected in applications; it just indicates that the class of Hellinger speedless densities is not dense in the space of squared-root densities.

[2]I would like to thank Pietro Spini for suggesting this Corollary.

Third, Proposition 1 suggests comparing Hellinger speedless points with *nonidentifiable* points characterized in terms of the Fisher matrix, see, e.g., Rothenberg (1971). I perform such a comparison in the next subsection.

4.1 Nonidentifiable Points

I proceed by recalling the definition of an identifiable point.

Definition (Identifiable Point). Two points θ and θ_0 in the parameter space are observational equivalent if and only if $f_\theta = f_{\theta_0}$ μ-ae. The point θ_0 is identifiable if there is no other θ in the parameter space that is observational equivalent to θ_0.

Proposition 2. Assume that f_{θ_0} belongs to the family of quadratic mean differentiable models. Then, every non-identifiable point is a Hellinger speedless point. The converse is not true, i.e., there are Hellinger speedless points that are identifiable.

The next two examples illustrate the differences and similarities between Hellinger speedless points and nonidentifiable points.

Example 1 (Normal Linear Instrumental Variable Model, cont'ed). This is an example of a Hellinger speedless point that is nonidentifiable. Consider the specification in Example 1. The model for the density of Y_i conditional on z_i is

$$f_\theta(y) = \frac{\exp\big(-\frac{1}{2}(y-\lambda_\theta)^\top \Omega_\beta^{-1}(y-\lambda_\theta)\big)}{\sqrt{(2\pi)^2 \det \Omega_\beta}}, \text{ where } \lambda_\theta = \begin{pmatrix} \beta\gamma z_i \\ \gamma z_i \end{pmatrix} \text{ and } \Omega_\beta = \begin{pmatrix} \beta^2+\beta+1 & \beta+\frac{1}{2} \\ \beta+\frac{1}{2} & 1 \end{pmatrix}.$$

The Hellinger pseudometric between θ and θ_0 induced by the $L_2(\mu)$ norm is (see e.g., Pardo, 2004, p. 51)

$$h(\theta,\theta_0)=\left[1-\frac{\det(\Omega_\beta)^{1/4}\det(\Omega_{\beta_0})^{1/4}}{\det\big((\Omega_\beta+\Omega_{\beta_0})/2\big)^{1/2}}\exp\Bigg(-\frac{1}{8}(\lambda_\theta-\lambda_{\theta_0})^\top\Big(\frac{\Omega_\beta+\Omega_{\beta_0}}{2}\Big)^{-1}(\lambda_\theta-\lambda_{\theta_0})\Bigg)\right]^{1/2}.$$

For $\theta_0=(0,0)=:0_2$, set the curve $\epsilon\mapsto e(\epsilon)=(\epsilon,0)$. Along this curve, we have $\lambda_{c(\epsilon)}-\lambda_{0_2}=(\epsilon,0)^\top$, $\det(\Omega_\epsilon)=3/4=\det(\Omega_0)$, and $\det\big((\Omega_\epsilon+\Omega_0)/2\big)=3/4$. Hence,

$$h(e(\epsilon),0_2)=\left[1-\frac{[(3/4)^{1/2}(3/4)^{1/2}]^{1/2}}{(3/4)^{1/2}}\exp(0)\right]^{1/2}=0.$$

We deduce that any $\theta_0\in\mathbb{R}\times 0$ is a nonidentifiable point. We also have

$$|\dot{h}_{\theta_0}(e)|:=\lim_{\epsilon\to 0}\frac{h(e(\epsilon),\theta_0)}{|\epsilon|}=\lim_{\epsilon\to 0}\frac{0}{|\epsilon|}=0\text{ and }0=|\dot{h}_{\theta_0}(e)|^2=\inf_{c\in\mathcal{C}_{\theta_0}}|\dot{h}_{\theta_0}(c)|^2.$$

The point $\theta_0=(0,0)$ is a Hellinger speedless point. From Proposition 1, we know that the smallest eigenvalue of the Fisher matrix evaluated at this point is zero.

Example 2 (Normal Squared Location Model, cont'ed). This example shows a Hellinger speedless point that is identifiable. Consider the specification in Example 2. The model for the density of Y_i is

$$f_\theta(y)=(\sqrt{2\pi})^{-1}\exp\Big(-(y-\theta^2)^2/2\Big)$$

This model has

$$h(\theta,\theta_0)=1-\exp\Big(-(\theta^2-\theta_0^2)^2/8\Big)\text{ and }\int g_\theta g_\theta^\top f_\theta d\mu=4\theta^2.$$

Consider the point $\theta_0=0$. We have $h(\theta,0)=1-\exp(-\theta^4/8)$ and the unique minimizer of $\theta\to h(\theta,0)$ is zero. Hence, $\theta_0=0$ is identifiable. We also have $\int g_0g_0^\top f_0d\mu=0$. Hence,

$\theta_0 = 0$ is also Hellinger speedless.

4.2 Singularities

I now compare Hellinger speedless points with singularities in the mapping from reduced form to structural parameters introduced by Hirano and Porter (2015). I proceed by recalling the definition of these singularities.

Definition (Singularity). Let $\kappa : \Theta \to \mathbb{R}$ denote a function from the parameter space into the real line. A point θ_0 is a singularity in κ if there is a convergent sequence $\{\theta_j \in \Theta\}_j$ such that $\lim_{j\to\infty} \theta_j = \theta_0$ and $\kappa(\theta_j) \to \pm\infty$ as $j \to \infty$.

A singularity in κ is a vertical asymptote. We have the following result.

Proposition 3. Every singularity in $\kappa : \Theta \to \mathbb{R}$ is a Hellinger speedless point. The converse is not true, i.e., there are Hellinger speedless points that are not singularities.

The next illustration provides an example of a Hellinger speedless point that is not a singularity of the mapping from reduced form to structural parameters.

Example 2 (Normal Squared Location Model, cont'ed). From our previous analysis, we known that $\theta_0 = 0$ is a Hellnger critical point. Set $\kappa(\theta) = \theta^2$ and consider any sequence $\{\theta_j \in \mathbb{R}\}_j$ converging to zero. For any of such sequences, we have $\kappa(\theta_j) \to 0$ because $\theta \to \theta^2$ is differentiable at $\theta_0 = 0$. Hence, $\theta_0 = 0$ is not a singularity of κ.

I conclude this section by observing that the absence of Hellinger speedless points is a necessary condition for the existence of finite-variance unbiased estimators (Theorem 1). This necessary condition is stronger than either the absence of nonidentifiable points (Proposition 2) or the absence of singularities in the mapping from reduced form to structural parameters (Proposition 3).

§5. Detecting Hellinger Speedless Points

The characterization of Hellinger speedless points in the previous Section suggests detecting these points by calculating the smallest eigenvalue of the Fisher matrix and comparing it with zero. I now develop this suggestion into a method.

5.1 Models with Fisher Matrix in Closed-Form

On the one hand, there are models where the Fisher matrix is available in analytical form. In these models, the detection of Hellinger speedless points only involves checking for the existence of a solution to a system of linear equations. Examples of models with closed-form Fisher matrix includes the normal self-selection model, see e.g., Lee and Chesher (1986), and the skew-normal binary response model, see e.g., Stingo et al. (1986). The analysis of these models suggests that Hellinger speedless points arise as a consequence of overparametrization. I use this insight later in this section to hypothesize the existence of Hellinger speedless points in models in which the smallest eigenvalue of the Fisher matrix is not already available in closed-form. As an illustration, consider the following example.

Example 3 (Normal Self-Selection Model, cont'ed). This example illustrates the detection of a Hellinger speedless point in a model a with closed-form Fisher matrix. The ultimate objective is to verify that the point $\theta_0 = (\beta, \gamma_1, \gamma_2 = 0, \rho = 0, \sigma_1)$ is Hellinger speedless. But

I simplify the calculations drastically for the sake of tractability. I confine calculations to the special case where x_{1i} only contains a constant. In this case, the Fisher matrix is, see e.g., Lee and Chesher (1986),

$$\mathcal{I}_{\theta_0}(\theta_0) = \begin{pmatrix} \frac{\Phi(\gamma_1)}{\sigma_1^2} & 0 & 0 & \frac{\phi(\gamma_1)}{\sigma_1} \\ 0 & \frac{\phi(\gamma_1}{\Phi(\gamma_1)[1-\Phi(\gamma_1)]} & 0 & 0 \\ 0 & 0 & \frac{1}{2\sigma_1^4}\Phi(\gamma_1) & 0 \\ \frac{\phi(\gamma_1)}{\sigma_1} & 0 & 0 & \frac{\phi(\gamma_1)^2}{\Phi(\gamma_1)} \end{pmatrix},$$

where $\phi(\gamma_1) = (2\pi)^{-1/2}\exp(-\gamma_1^2/2)$ and $\Phi(\gamma_1) = \int_{-\infty}^{\gamma_1}\phi(u)du$. Consider the first and last columns. The system of linear equations relating these columns

$$\begin{aligned} \frac{\Phi(\gamma_1)}{\sigma_1^2}v - \frac{\phi(\gamma_1)}{\sigma_1} &= 0 \\ \frac{\phi(\gamma_1)}{\sigma_1}v - \frac{\phi(\gamma_1)^2}{\Phi(\gamma_1)} &= 0 \end{aligned}$$

has solution $v = -\sigma_1\phi(\gamma_1)/\Phi(\gamma_1)$. This means that the Fisher matrix is singular, as the first and last columns are linearly dependent. Hence, the smallest eigenvalue of the Fisher matrix is zero. I now deduce, from Proposition 2, that θ_0 is a Hellinger speedless point. At $\rho = 0$, there is no self-selection, and at $\gamma_2 = 0$, there are no excluded covariates.

5.2 Models with Intractable Fisher Matrix

On the other hand, there are models in which calculating the Fisher matrix in closed-form is impractical or intractable. For illustrative purposes, consider the normal bivariate binary response model with endogenous switching in Example 4. The calculations that we have performed for the linear version of this model in Example 1 induce us to hypothesize the presence of Hellinger speedless points. I do not find it instructive to perform similar calculations for Example 4. For models like this one, this section proposes a simulation-based statistical test.

Consider the problem of testing

$$H_0 : \underline{e}(\theta_0) = 0 \text{ and } H_1 : \underline{e}(\theta_0) > 0.$$

The objective is to construct a test for this problem that controls the probability of incorrectly rejecting the hypothesis H_0 that θ_0 is a Hellinger speedless point. Define the sample eigenvalue statistic as

$$\widehat{\underline{e}}_0 = \hat{q}_N' \mathcal{I}_N(\theta_0) \hat{q}_N, \text{ where } \hat{q}_N \in \arg\min_{q \in \mathbb{S}} q' \mathcal{I}_N(\theta_0) q \text{ and } \mathcal{I}_N(\theta_0) = \frac{1}{N} \sum_{i=1}^{N} g_{\theta_0}(Y_i) g_{\theta_0}(Y_i)'.$$

The next proposition gives the asymptotic null distribution of this test statistic. This result is key to explaining the failure of the parametric bootstrap to deliver a test controlling size and provides motivation for the randomization test I describe below.

Proposition 4. Assume that f_θ belongs to a QMD model. Assume also that the fourth moment of s_{θ_0} is finite. Under $\underline{e}(\theta_0) = 0$,

$$\sqrt{N} \hat{\underline{e}}_0 \rightsquigarrow J_{\theta_0} = \arg\min_{q \in Q_0} q^\top \mathbb{G}_{\theta_0} q^\top$$

where $Q_0 = \arg\min_{q \in \mathbb{S}} q^\top \mathcal{I}_{\theta_0}(\theta_0) q$, and $\mathbb{G}_{\theta_0} \sim \mathcal{N}(0_{K \times K}, V_{\theta_0})$ is a $K^2 \times K^2$ Gaussian random matrix.

The asymptotic null distribution J_{θ_0} depends on Q_0, which gives the algebraic multiplicity of the zero eigenvalue. The parametric bootstrap fails to approximate the asymptotic distribution J_{θ_0} because the mapping $\theta \mapsto \underline{e}(\theta)$ fails to be differentiable unless Q_0 is a singleton.

To proceed, I show that the problem of testing H_0 remains invariant under scalar reparametrizations. The interested reader can find a review of the general theory of invariance in relation to statistical testing problems in for example, Lehmann and Romano (2005, Chapter 6). Write H_0 as $H_0 : \theta_0 \in \Theta_0$, where $\Theta_0 = \{\theta \in \Theta : \underline{e}(\theta) = 0\}$ is the set of Hellinger speedless points. Let $\overline{G}$ denote the group of scalar reparametrizations

$$\overline{g}\theta = \kappa\theta \text{ for any } \kappa \in \mathcal{U}_{(0,1)} := \{1 - j/(J+1)\}_{j=1}^{J},$$

where $\overline{g}\theta$ denote the action of $\overline{g} \in \overline{G}$ on $\theta \in \Theta$. Since $\underline{e}(g\theta_0) = \kappa^2 \underline{e}(\theta_0)$, the set of Hellinger speedless points is invariant under scalar reparametrizations: $\overline{g}\Theta_0 = \Theta_0$.

I now exploit the invariance of the set of Hellinger speedless points to construct a randomization test for H_0 against H_1. A review of the general theory of randomization tests can be found, for example, in Lehmann and Romano (2005, Chapter 15). Fix a nominal level α between zero and one. Let $\{Y_{i,j}\}_{i=1}^{N}$ denote the Monte Carlo sample simulated under $\kappa_j\theta_0$ for $\kappa_j \in \mathcal{U}_{[1,2)}$. Let $\widehat{\underline{e}_j}$ denote the value of the simulated sample eigenvalue statistic calculated using the $j-th$ simulated sample. Denote the ordered values of the simulated statistic as $\overline{g}$ varies in $\overline{G}$ by

$$\widehat{\underline{e}_{(1)}} \leq \ldots \leq \widehat{\underline{e}_{(j)}} \leq \ldots \leq \widehat{\underline{e}_{(J)}}.$$

Let $j_\star := \lceil J(1-\alpha) \rceil$ denote the smallest integer larger than $J(1-\alpha)$ and let

$$J_+ := \sum_{j=1}^{J} 1(\widehat{\underline{e}_{(j)}} > \widehat{\underline{e}_{(j_\star)}}) \text{ and } J_- := \sum_{j=1}^{J} 1(\widehat{\underline{e}_{(j)}} = \widehat{\underline{e}_{(j_\star)}})$$

denote the number of simulated statistics that are, respectively, greater than and equal to

$\widehat{\underline{e}_{(j_\star)}}$. The randomization test function is:

$$T_{\theta_0} = \begin{cases} 1 & \text{if } \widehat{\underline{e}_0} > \widehat{\underline{e}_{(j_\star)}} \\ (\alpha J - J_+)/J_0 & \text{if } \widehat{\underline{e}_0} = \widehat{\underline{e}_{(j_\star)}} \\ 0 & \text{if } \widehat{\underline{e}_0} < \widehat{\underline{e}_{(j_\star)}}. \end{cases}$$

The following result shows that the construction in the previous paragraph keeps under control the probability of incorrectly rejecting the hypothesis that θ_0 is a Hellinger speedless point.

Proposition 5. For any $\alpha \in (0,1)$, the test function T_{θ_0} satisfies

$$E_{\theta_0}(T_{\theta_0}) = \alpha \text{ whenever } \theta_0 \in \Theta_0.$$

Several remarks are in order. First, the randomization test T_{θ_0} controls size in finite samples: the nominal level α is, for any sample size, equal to the probability of incorrectly rejecting the hypothesis that θ_0 is Hellinger speedless. Appendix B presents an alternative resampling test and it shows that it controls size in large samples. Second, the randomization test is similar: T_{θ_0} has rejection probability α for any Hellinger speedless point. Third, for a point θ_0 that is regular to the Fisher matrix, the randomization test becomes a local identifiability test in view of the characterization of identifiability for regular points in Rothenberg (1971, Theorem 1). Fourth, for the point $\theta_0 = 0$, the randomization test is equivalent to the parametric bootstrap test. This is not in contradiction with the assertion made in the Introduction that, in the general case when $\theta_0 \neq 0$, the parametric bootstrap test does guarantee size control, even in large samples. The randomization test controls size in finite samples for any Hellinger speedless point.

5.3 Numerical Exercises

Implementing the randomization test requires computing the smallest eigenvalue of the sample Fisher matrix. There are alternative methods for carrying out such a computation. Each method is subject to computational numerical errors, and the computed eigenvalue may have low or even no accuracy, depending on the conditioning number of the matrix of eigenvalues. All that one can hope for is to find a method that gives a solution suitably close to the statistic $\underline{\widehat{e}}$. This Section explores three alternative methods to compute $\underline{\widehat{e}}$ and it warns about the need to take due care of computational numerical errors that could arise when implementing the randomization test function described above.

The first method uses the QR algorithm for computing the eigenvalues of a symmetric matrix, as implemented by the function `eigen` in the statistical software R. I denote the solution given by this method by $\underline{\tilde{e}}$ and I call it the QR computation of the smallest eigenvalue. The simulations show that the QR computation can return a negative number while $\underline{\widehat{e}}$ is non-negative because the Fisher matrix is positive semi-definite. The difference between the QR computation and the smallest eigenvalue is due to computational numerical errors. I explore two alternative methods. The first alternative uses the Rayleigh-Ritz characterization of the smallest eigenvalue of a square matrix M as the solution to the quadratic problem $\min_{q \in \mathbb{S}} q'Mq$, where $\mathbb{S}$ denotes the unit sphere. Let $\underline{\widetilde{q}}$ denote the eigenvalue corresponding to the Jacobi computation $\underline{\tilde{e}}$. The Rayleigh-Ritz-Jacobi computation of the smallest eigenvalue is $\underline{\breve{e}} = \underline{\widetilde{q}}'\hat{I}_\theta(\theta)\underline{\widetilde{q}}$. The second alternative uses the Jacobi algorithm for computing the eigenvalues of a symmetric matrix, as implemented by the function `Jacobi` in the statistical software R.

I use the normal self-selection model in Example 3 to evaluate the alternative methods to compute the smallest eigenvalue of the sample Fisher matrix. I have already verified in

Section 5.1 that the smallest eigenvalue of the Fisher matrix at the point $\beta_1 = (1, 0), \sigma_1 = 1, \rho = 0, \gamma_1 = (1, 0), \gamma_2 = 0$ is zero. At this point, Proposition 5 indicates the randomization test has a rejection rate equal to α. Table I reports the empirical rejection rates when the smallest eigenvalue is computed using the QR, QR-Rayleigh-Ritz, and Jacobi methods. I find that the numerical computation errors are non-negligible. The QR method gives the most suitable approximation.

Table I: Empirical Rejection Probabilities for the Randomization Test

Example 3: Normal Self-Selection Model with $\beta_1 = (1, 0), \sigma_1 = 1, \rho = 0, \gamma_1 = (1, 0), \gamma_2 = 0$

Sample Size	QR	QR-Rayleigh-Ritz	Jacobi
100	.052	.168	.185
400	.003	.204	.241
1,600	.000	.208	.172

Note: Nominal size is $\alpha = .05$. The number of simulations is $J = 99$.
The number of Monte Carlo replications is 999.

Table II: Empirical Rejection Probabilities for the Randomization Test

Example 3: Normal Self-Selection Model with Local Alternatives $\theta_0 + h/\sqrt{N}$

Sample Size	$h = 1$	$h = 2$	$h = 3$
100	.565	.595	.613
400	.797	.779	.815
1,600	.925	.945	.925

Note: Nominal size is $\alpha = .05$. $\theta_0 = (\beta_1 = (1, 0), \sigma_1 = 1, \rho = 0, \gamma_1 = (1, 0), \gamma_2 = 0)$.
The number of simulations is $J = 99$. Eigenvalues computed using the QR algorithm implementation in R. The number of Monte Carlo replications is 999.

The last numerical exercise explores the power to local alternatives of the randomization test. Table II reports the results.

5.4 Application: Endogenous Recursive Switching Model

The objective of this exercise is to illustrate the use of the least eigenvalue randomization test in a model without a closed-form expression for the Fisher matrix. The model is the one described in Example 4:

$$\begin{aligned} Y_{1i}^\star = &\quad x_{1i}\beta_1 + Y_{2i}\beta_2 + \tilde{x}_{1i}Y_{2i}\beta_3 - U_{1i} \\ Y_{2i}^\star = &\quad x_{1i}\gamma_1 + x_{2i}\gamma_2 - U_{2i} \end{aligned}, \text{ where } U_i|x_i \sim \mathcal{N}\begin{pmatrix} 0 & 1 & \rho \\ & , & \\ 0 & \rho & 1 \end{pmatrix},$$

$x_1 = (1, \tilde{x}_{1i})$ and $\tilde{x}_{1i}, x_{2i}$ are independent normal random variables with zero mean and variance .2 and .3, respectively. The parameters are $\theta := (\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \rho)$. Appendix B derives the individual score for this model. A non-normal version of this model, modeling the joint distribution of the disturbances with a copula, has been studied by Han and Lee (2019). The least eigenvalue randomization test also applies to parametric copula models of the disturbances.

I consider four types of points in the parameter space: points representing the absence of switching ($\beta_3 = 0$), points representing the absence of recursion ($\beta_2 = \beta_3 = 0$), points representing the absence of endogeneity ($\rho = 0$), and points representing irrelevant instrumental variables ($\gamma_2 = 0$). The least eigenvalue of the Fisher matrix evaluated at each of these points is the smallest root of an eight-degree polynomial. The roots of such a polynomial do not exist in closed form. I then use the least eigenvalue randomization test, which does not require to have a Fisher matrix in closed-form, to verify which of the four points are Hellinger speedless.

The randomization test indicates that points representing the absence of endogeneity and irrelevant instrumental variables are Hellinger speedless while points representing the absence of recursion or switching are not.

§6. Conclusion

There are econometric models where all finite-variance estimators are biased. Leading examples include linear instrumental variable models, normal self-selection models, normal bivariate binary response models with endogeneity and skew-normal binary response models. This impossibility result poses elementary questions, such as what the common cause of this result is, how it can be detected, and by what means it should be handled. This paper introduces Hellinger speedless points as a notion for explaining the nonexistence of finite-variance unbiased estimators. For quadratic mean differentiable models, a point is Hellinger speedless if and only if the smallest eigenvalue of the Fisher matrix evaluated at the point is zero. This characterization provides a method for diagnosing the nonexistence of finite-variance unbiased estimators. As a by-product, the method gives a test of local identifiability with exact size control. The application of this method gives support to the view that unwittingly modelling instrumental variables, self-selection, endogeneity, or skewness when they are not features of the data leads to the nonexistence of finite-variance unbiased estimators. More generally, these results provide motivation for considering alternative criteria other than minimum-variance unbiasedness when evaluating estimators in the presence of Hellinger speedless points.

This paper focuses on the small-sample properties of estimators in models with Hellinger speedless points in a finite-dimensional parameter space. The large-sample properties of the maximum likelihood and generalised method-of-moments estimators have been investigated by Rotnitzky, Cox, Bottai, and Robbins (2000) and Dovonon and Hall (2018), respectively. Semiparametric density models, e.g., Bonhomme (2012), Khan and Nekipelov (2018), have infinite-dimensional parameter spaces. To the best of my knowledge, the consequences of the presence of Hellinger speedless points on the construction of finite-variance unbiased estimators have not yet been systematically investigated in these models. The analysis of semiparametric density models using the tools employed to study parametric models looks promising. The notion of Hellinger-metric derivative does extend to infinite-dimensional spaces; see, e.g., Ambrosio, Gigli and Savare (2004, Chapter 1) for the general definition of metric derivative. The notion of the Fisher matrix does also extend to infinite-dimensional spaces; see, e.g., Escanciano (2022). Establishing the connection between these two notions in semiparametric models, and their consequences for the nonexistence of a finite-variance unbiased estimator is left for future work.

References

Ambrosio, L., N. Gigli and G. Savare (2004): *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhauser, Basel.

Anderson, T. (2003): *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Bonhomme, S. (2012): "Functional Differencing", *Econometrica.*

Brown, L. and R. Liu (1993): "Nonexistence of Informative Unbiased Estimators in Singular Problems", *Annals of Statistics.*

Chen, Q. and Z. Fang (2019): "Improved Inference on the Rank of a Matrix", *Quantitative Economics.*

Cragg, J. and S. Donald (1993): "Testing Identifiability and Specification in Instrumental Variables Models", *Econometric Theory.*

Doss, H. and J. Sethuraman (1989): "The Price of Bias Reduction when there is no Unbiased Estimate", *Annals of Statistics.*

Dovonon, P. and A. Hall (2018): "The Asymptotic Properties of GMM and Indirect Inference under Second-Order Identification", *Journal of Econometrics.*

Dufour, J-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models", *Econometrica.*

Escanciano, J.C. (2022): "Semiparametric Identification and Fisher Information", *Econometric Theory.*

Evans, W. and R. Schwab (1995): "Finishing High School and Starting College: Do Catholic School Make a Difference?", *Quarterly Journal of Economics.*

Firth, D. (1993): "Bias Reduction of Maximum Likelihood Estimates", *Biometrika.*

Gentzkow, M (2009): "Valuing New Goods in a Model with Complementarity: Online Newspapers ", *American Economic Review.*

Gronau, R. (1973): "Wage Comparisons - A Selectivity Bias", *Journal of Political Economy.*

Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models*, Cambridge University Press, Cambridge.

Han, S. and S. Lee (2019): "Estimation in a generalization of bivariate probit models with dummy endogenous regressors", *Journal of Applied Econometrics*.

Heckman, J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*.

Helpman, E., O. Itskhoki, M.-A. Muendler and S. Redding (2017): "Trade and Inequality: From Theory to Estimation", *Review of Economic Studies*.

Hirano, K. and J. Porter (2012): "Impossibility Results for Nondifferentiable Functions", *Econometrica*.

Hirano, K. and J. Porter (2015), "Location Properties of Point Estimators in Linear Instrumental Variable and Related Models", *Econometric Reviews*.

Jones, A. (2000): "Health Econometrics" in A. Culyer and J. Newhouse (Eds.): *Handbook of Health Economics*, Elsevier, Amsterdam.

Kaji, T. (2021): "Theory of Weak Identification in Semiparametric Models", *Econometrica*.

Khan, S. and D. Nekipelov (2018): "Information Structure and Statistical Information in Discrete Response Models", *Quantitative Economics*.

Lee, J-F. and A. Chesher (1986): "Specification Testing when Score Test Statistics are Identically Zero", *Journal of Econometrics*.

Lehmann, E. and J. Romano (2005): *Testing Statistical Hypothesis*, Springer.

Li, K. and N. Prabhala (2007): "Self-Selection Models in Corporate Finance", in B. Espen Eckbo (Ed.): *Handbook of Corporate Finance*, North-Holland, Amsterdam.

Lokshin, M. and E. Glinskaya (2009): "The Effect of Male Migration on Employment Patterns of Women in Nepal", *The World Bank Economic Review*.

Müller, U. and Y. Wang (2019): "Nearly Weighted Risk Minimal Unbiased Estimation", *Journal of Econometrics*.

Phillips, P. (1980): "The Exact Distribution of Instrumental Variable Estimators in an Equation containing n+1 Endogenous Variables", *Econometrica*.

Rothenberg, T. (1971): "Identification in Parametric Models", *Econometrica*.

Rotnitzky, A. D. Cox, M. Bottai and J. Robbins (2000): "Likelihood-based Inference with Singular Information Matrix", *Bernoulli*.

Stingo, F., E. Stanghelini, and R. Capobianco (2012): "On the Estimation of a Binary Response Model in a Selected Population", *Journal of Statistical Planning and Inference.*

van der Vaart, A. (1991): "On Differentiable Functionals", *Annals of Statistics*

Appendix A: Proofs

Proof of Theorem 1. I first derive an inequality relating the first two moments of the estimator to the Hellinger pseudo-metric derivative. Without loss of generality, set $N = 1$. Let $\hat{\beta}_1 : \mathcal{Y} \to B$ be an estimator with moments

$$m_1(\theta_0) := \int \hat{\beta}_1 f_{\theta_0} d\mu \text{ and } m_2(\theta_0) := \int \hat{\beta}_1^2 f_{\theta_0} d\mu.$$

The bias of $\hat{\beta}_1$ is $b_1(\theta_0) := m_1(\theta_0) - \beta_0$. Fix a curve $c : [0,1] \to \Theta$ in $\mathcal{C}_{\theta_0}$. Write

$$m_1\big(c(\epsilon)\big) - m_1(\theta_0) = \int \hat{\beta}_1 \big[f_{c(\epsilon)} - f_{\theta_0}\big] d\mu = \int \hat{\beta}_1 \big[f_{c(\epsilon)}^{1/2} + f_{\theta_0}^{1/2}\big]\big[f_{c(\epsilon)}^{1/2} - f_{\theta_0}^{1/2}\big] d\mu.$$

By the Cauchy-Schwarz Inequality

$$|m_1\big(c(\epsilon)\big) - m_1(\theta_0)|^2 \leq \int \hat{\beta}_1^2 \big[f_{c(\epsilon)}^{1/2} + f_{\theta_0}^{1/2}\big]^2 d\mu \int \big[f_{c(\epsilon)}^{1/2} - f_{\theta_0}^{1/2}\big]^2 d\mu.$$

Dividing both sides by ϵ^2, one has

$$\frac{|m_1\big(c(\epsilon)\big) - m_1(\theta_0)|}{|\epsilon|} \leq \sqrt{\int \hat{\beta}_1^2 \big[f_{c(\epsilon)}^{1/2} + f_{\theta_0}^{1/2}\big]^2 d\mu} \sqrt{\int \frac{\big[f_{c(\epsilon)}^{1/2} - f_{\theta_0}^{1/2}\big]^2 d\mu}{\epsilon^2}}.$$

Taking lim inf to the left-hand-side and lim sup to the right-hand-side of the inequality yields

$$\underbrace{\liminf_{\epsilon \downarrow 0} \frac{|m_1\big(c(\epsilon)\big) - m_1(\theta_0)|}{\epsilon}}_{(1)} \leq \sqrt{\underbrace{\limsup_{\epsilon \downarrow 0} \int \hat{\beta}_1^2 \big[f_{c(\epsilon)}^{1/2} + f_{\theta_0}^{1/2}\big]^2 d\mu}_{(2)}}\; \underbrace{\limsup_{\epsilon \downarrow 0} \frac{\sqrt{\int \big[f_{c(\epsilon)}^{1/2} - f_{\theta_0}^{1/2}\big]^2 d\mu}}{\epsilon}}_{(3)}$$

Define

$$\dot{m}_{1\theta_0}(c) := \liminf_{\epsilon \downarrow 0} \frac{|m_1\big(c(\epsilon)\big) - m_1(\theta_0)|}{\epsilon}.$$

Consider (2). Since we have assumed that there is an integrable function $\bar{g}$ such that $f_{c(\epsilon)} \leq \bar{g}$

for any ϵ, the reversed Fatou Lemma implies

$$\limsup_{\epsilon\downarrow 0}\int \hat{\beta}_1^2\big[f_{c(\epsilon)}^{1/2}+f_{\theta_0}^{1/2}\big]^2 d\mu \le \int \limsup_{\epsilon\downarrow 0}\hat{\beta}_1^2\big[f_{c(\epsilon)}^{1/2}+f_{\theta_0}^{1/2}\big]^2 d\mu.$$

Since $\lim_{\epsilon\downarrow 0} c(\epsilon)=\theta_0$, there is a subsequence $\{\epsilon_j\in[0,1]\}_{j\in\mathbb{N}}$ such that $\lim_{j\to\infty} c(\epsilon_j)=\theta_0$ and

$$\int \limsup_{\epsilon\downarrow 0}\hat{\beta}_1^2\big[f_{c(\epsilon)}^{1/2}+f_{\theta_0}^{1/2}\big]^2 d\mu = \int \lim_{j\to\infty}\hat{\beta}_1^2\big[f_{c(\epsilon_j)}^{1/2}+f_{\theta_0}^{1/2}\big]^2 d\mu = \int \hat{\beta}_1^2[2f_{\theta_0}]^2 d\mu = 4m_2(\theta_0).$$

We then have

$$\sqrt{\limsup_{\epsilon\downarrow 0}\int \hat{\beta}_1^2\big[f_{c(\epsilon)}^{1/2}+f_{\theta_0}^{1/2}\big]^2 d\mu} \le 2m_2(\theta_0)^{1/2}.$$

Consider now (3).

$$\begin{aligned}\limsup_{\epsilon\downarrow 0}\frac{\sqrt{\int\big[f_{c(\epsilon)}^{1/2}-f_{\theta_0}^{1/2}\big]^2 d\mu}}{\epsilon} &= \limsup_{\epsilon\to 0}\frac{\sqrt{\int\big[f_{c(\epsilon)}^{1/2}-f_{\theta_0}^{1/2}\big]^2 d\mu}}{|\epsilon|} = \limsup_{\epsilon\to 0}\sqrt{2}\frac{h\big(c(\epsilon),\theta_0\big)}{|\epsilon|}\\ &= \sqrt{2}\lim_{\epsilon\to 0}\frac{h\big(c(\epsilon),\theta_0\big)}{|\epsilon|},\end{aligned}$$

where the last equality follows because $c\in\mathcal{C}_{\theta_0}$. We then have

$$\limsup_{\epsilon\downarrow 0}\frac{\sqrt{\int\big[f_{c(\epsilon)}^{1/2}-f_{\theta_0}^{1/2}\big]^2 d\mu}}{\epsilon} = \sqrt{2}|\dot{h}_{\theta_0}(c)|.$$

Combining the results, we have

$$\dot{m}_{1\theta_0}(c) \le 2m_2(\theta_0)^{1/2}\sqrt{2}|\dot{h}_{\theta_0}(c)|.$$

Hence,

Lemma 1. $\inf_{c\in\mathcal{C}_{\theta_0}}\dot{m}_{1\theta_0}(c) \le \sqrt{8}m_2(\theta_0)^{1/2}v_{\theta_0}$, on the proviso that $m_2(\theta_0)$ is finite, where

$v_{\theta_0} = \inf_{c \in \mathcal{C}_{\theta_0}} |\dot{h}_{\theta_0}(c)|$.

We are now ready to use Lemma 1 to prove Theorem 1. The proof is by contradiction. An unbiased estimator has

$$b_1(\theta_0) = m_1(\theta_0) - \beta_0 = 0.$$

We have

$$\begin{aligned} 0 &= \liminf_{\epsilon \downarrow 0} \frac{|b_1\big(c(\epsilon)\big) - b_1(\theta_0)|}{\epsilon} = \liminf_{\epsilon \downarrow 0} \frac{|m_1\big(c(\epsilon)\big) - \beta_\epsilon - m_1(\theta_0) + \beta_0|}{\epsilon} \\ &= \liminf_{\epsilon \downarrow 0} \frac{|m_1\big(c(\epsilon)\big) - m_1(\theta_0)|}{\epsilon} - \liminf_{\epsilon \downarrow 0} \frac{|\beta_\epsilon - \beta_0|}{\epsilon}, \end{aligned}$$

where the last equality follows from the reversed triangle inequality. Since the curve c is differentiable, there is a constant $0 < \kappa_c < \infty$ such that $|\beta_\epsilon - \beta_0| = \kappa_c \epsilon$. Hence,

$$\dot{m}_{1\theta_0}(c) = \kappa_c > 0 \text{ for every } c \in \mathcal{C}_{\theta_0}.$$

Assume now that $\hat{\beta}_1$ is unbiased, which implies $\underline{\kappa} := \inf_{c \in \mathcal{C}_{\theta_0}} \dot{m}_{1\theta_0}(c)$ is such that $0 < \underline{\kappa} < \infty$, and it has finite variance, which implies $0 \le m_2(\theta_0) < \infty$, at a critical point θ_0, i.e., $v_{\theta_0} = 0$. This contradicts the inequality in Lemma 1. $\square$

Proof of Proposition 1. Let $\underline{\mathrm{e}}(\mathcal{I}_{\theta_0}(\theta_0))$ denote the smallest eigenvalue of the Fisher matrix $\mathcal{I}_{\theta_0}(\theta_0)$ evaluated at θ_0. Fix $c \in \mathcal{C}_{\theta_0}$. Let $\mathbb{S} := \{q \in \mathbb{R}^2 : \|q\|_2 = 1\}$ denote the unit sphere in $\mathbb{R}^2$. Write $c(\epsilon) = \theta_0 + \epsilon q_\epsilon$ for some $q_\epsilon \in \mathbb{R}^2$ such that $q_\epsilon \to q$, $q \in \mathbb{S}$.

We first verify that

$$|\dot{h}_{\theta_0}(c)| = \frac{1}{\sqrt{2}} \left\| \lim_{\epsilon \downarrow 0} \epsilon^{-1} (f^{1/2}_{\theta_0 + \epsilon q_\epsilon} - f^{1/2}_{\theta_0}) \right\|_{L_2(\mu)}.$$

Start from

$$|\dot{h}_{\theta_0}(c)| := \lim_{\epsilon\to 0}\frac{h(c(\epsilon),\theta_0)}{|\epsilon|} = \lim_{\epsilon\to 0}\frac{\left[\frac{1}{2}\int(f_{c(\epsilon)}^{1/2}-f_{\theta_0}^{1/2})^2 d\mu\right]^{1/2}}{|\epsilon|}$$

$$= \frac{1}{\sqrt{2}}\lim_{\epsilon\to 0}\left[\int \frac{(f_{\theta_0+\epsilon q_\epsilon}-f_{\theta_0}^{1/2})^2 d\mu}{|\epsilon|^2}\right]^{1/2} = \frac{1}{\sqrt{2}}\left[\int\left(\lim_{\epsilon\downarrow 0}\epsilon^{-1}(f_{\theta_0+\epsilon q_\epsilon}-f_{\theta_0}^{1/2})\right)^2 d\mu\right]^{1/2}$$

$$= \frac{1}{\sqrt{2}}\left\|\lim_{\epsilon\downarrow 0}\epsilon^{-1}(f_{\theta_0+\epsilon q_\epsilon}^{1/2}-f_{\theta_0}^{1/2})\right\|_{L_2(\mu)}.$$

We now verify

$$\left\|\lim_{\epsilon\downarrow 0}\epsilon^{-1}(f_{\theta_0+\epsilon q_\epsilon}^{1/2}-f_{\theta_0}^{1/2})\right\|_{L_2(\mu)} = \int q^\top s_{\theta_0}s_{\theta_0}^\top f_{\theta_0} q d\mu.$$

Since $\theta \to f_\theta^{1/2}$ is differentiable in quadratic mean,

$$\|f_{\theta_0+\Delta}^{1/2}-f_{\theta_0}^{1/2}-\Delta^\top s_{\theta_0}f_{\theta_0}^{1/2}\|_{L_2(\mu)} = o(\|\Delta\|_2), \text{ for every } \|\Delta\|_2 \to 0.$$

Set $\Delta = \epsilon q_\epsilon$. We have

$$\|f_{\theta_0+\epsilon q_\epsilon}^{1/2}-f_{\theta_0}^{1/2}-\epsilon q_\epsilon^\top s_{\theta_0}f_{\theta_0}^{1/2}\|_{L_2(\mu)} = \epsilon\|q_\epsilon\|_2 o(1) \text{ as } \epsilon\downarrow 0 \text{ and } q_\epsilon \to q.$$

and

$$\epsilon^{-1}(f_{\theta_0+\epsilon q_\epsilon}^{1/2}-f_{\theta_0}^{1/2}) = q^\top s_{\theta_0}f_{\theta_0}^{1/2}+o(1) \text{ as } \epsilon\downarrow 0 \text{ and } q_\epsilon\to q.$$

Taking $L_2(\mu)$ norm to both sides

$$\left\|\lim_{\epsilon\downarrow 0}\epsilon^{-1}(f_{\theta_0+\epsilon q_\epsilon}^{1/2}-f_{\theta_0}^{1/2})\right\|_{L_2(\mu)} = \left\|q^\top s_{\theta_0}f_{\theta_0}^{1/2}\right\|_{L_2(\mu)}$$

$$= \int q^\top s_{\theta_0}s_{\theta_0}^\top f_{\theta_0} q d\mu \text{ as } q_\epsilon \to q.$$

We deduce

$$|\dot{h}_{\theta_0}(c)| = \frac{1}{\sqrt{2}} \int q^\top s_{\theta_0} s_{\theta_0}^\top f_{\theta_0} q d\mu$$

Taking inf to both sides,

$$v_{\theta_0} := \inf_{c \in \mathcal{C}_{\theta_0}} |\dot{h}_{\theta_0}(c)| = \frac{1}{\sqrt{2}} \inf_{q \in \mathbb{S}} q^\top \mathcal{I}_{\theta_0}(\theta_0) q = \underline{e}(\mathcal{I}_{\theta_0}(\theta_0)),$$

where the last equality follows from the Rayleigh-Ritz Theorem (see e.g. Horn and Johnson, 1990, Theorem 4.2.2, p. 176) characterizing the smallest eigenvalue of a symmetric matrix. We deduce, from the last display, that θ_0 is a Hellinger speedless point if and only if $\underline{e}\big(\mathcal{I}_{\theta_0}(\theta_0)\big) = 0$. □

Proof of Proposition 2. Fix θ_0. By the Mean Value Theorem (see e.g., Coleman, 2012, Theorem 3.2), there is a point $\theta_\star$ between θ and θ_0 such that

$$f_\theta^{1/2} - f_{\theta_0}^{1/2} = (\theta - \theta_0)^\top s_{\theta_\star} f_{\theta_\star}^{1/2} \text{ for any } \theta.$$

Assume that θ_0 is a nonidentifiable point. Then, there is a sequence $\{\theta_j \in \Theta\}_{j \in \mathcal{N}}$ converging to θ_0 such that $f_{\theta_j}^{1/2} - f_{\theta_0}^{1/2} = 0$ and $\|\theta_j - \theta_0\|_2 \neq 0$. Evaluating the latter display at $\theta = \theta_j$ and dividing both sides by $\|\theta_j - \theta_0\|_2$ yields

$$0 = f_{\theta_j}^{1/2} - f_{\theta_0}^{1/2} = q_j s_{\theta_\star} f_{\theta_\star}^{1/2}, \text{ where } q_j = \frac{(\theta_j - \theta_0)}{\|\theta_j - \theta_0\|_2}.$$

The sequence $\{q_j \in \mathbb{S}\}_{j \in \mathcal{N}}$ converges (passing to a subsequence is necessary) to a limit $q_0 \in \mathbb{S}$ because $\mathbb{S}$ is compact. Hence, $q_j g_{\theta_\star} f_{\theta_\star}^{1/2} \to q_0 s_{\theta_0} f_{\theta_0}^{1/2} = 0$ and

$$\int q_0^\top s_{\theta_0} s_{\theta_0}^\top f_{\theta_0} q_0 d\mu = 0.$$

From the Rayleigh-Ritz Theorem, we have $\underline{e}(\mathcal{I}_{\theta_0}(\theta_0)) = 0$. We deduce from the characterization of Hellinger speedless points in Proposition 1 that every non-identifiable point is a Hellinger speedless point. The assertion that not every Hellinger speedless point is nonidentifiable follows from the example in Illustration I. $\square$

Proof of Proposition 3. Let θ_0 be a singularity of the mapping from reduced from to structural parameters. Let $\{\theta_j\}_{j\in\mathcal{N}}$ be a sequence converging to θ_0. Since $\kappa(\theta_j)$ diverges, we have that for every positive number ϵ, there is a sufficiently large j such that

$$\epsilon < |\kappa(\theta_j) - \kappa(\theta_0)|.$$

Define $b_0(\epsilon) = \sup_{\theta:h(\theta,\theta_0)\leq\epsilon} |\kappa(\theta) - \kappa(\theta_0)|$. Divide both sides of the inequality by $b_0(\epsilon)$ and take lower and upper limits

$$\liminf_{\epsilon\downarrow 0} \frac{\epsilon}{b_0(\epsilon)} \leq \limsup_{\epsilon\downarrow 0} \frac{|\kappa(\theta_j) - \kappa(\theta_0)|}{b_0(\epsilon)}$$

Consider the upper limit. Since $\epsilon < |\kappa(\theta_j) - \kappa(\theta_0)|$ and $\epsilon \to b_0(\epsilon)$ is decreasing, we have

$$b_0(\epsilon) = \sup_{\theta:h(\theta,\theta_0)\leq\epsilon} |\kappa(\theta) - \kappa(\theta_0)| \geq \sup_{\theta:h(\theta,\theta_0)\leq|\kappa(\theta_j)-\kappa(\theta_0)|} |\kappa(\theta) - \kappa(\theta_0)| \geq |\kappa(\theta_j) - \kappa(\theta_0)|.$$

Hence,

$$\limsup_{\epsilon\downarrow 0} \frac{|\kappa(\theta_j) - \kappa(\theta_0)|}{b_0(\epsilon)} \leq 0.$$

Consider now the lower limit. For a sufficiently small $\delta > 0$,

$$\inf_{\theta:|\theta-\theta_0|=\delta} h(\theta, \theta_0) = \inf_{\epsilon\in(0,\infty):b_0(\epsilon)>\delta} \epsilon.$$

Hence,

$$\liminf_{\epsilon\downarrow 0} \frac{\epsilon}{b_0(\epsilon)} = \lim\inf_{\theta\to\theta_0} \frac{h(\theta, \theta_0)}{|\theta-\theta_0|}.$$

There is a curve $c \in \mathcal{C}_{\theta_0}$ such that

$$\lim\inf_{\theta\to\theta_0} \frac{h(\theta, \theta_0)}{|\theta-\theta_0|} \geq \lim_{\epsilon\downarrow 0} \frac{h(c(\epsilon), \theta_0)}{|\epsilon|} \geq v_{\theta_0} \geq 0.$$

We deduce,

$$0 \leq v_{\theta_0} \leq \liminf_{\epsilon\downarrow 0} \frac{\epsilon}{b_0(\epsilon)}.$$

Combining the bounds on the lower and upper limits, one has

$$0 \leq v_{\theta_0} \leq 0.$$

Therefore, θ_0 is a Hellinger speedless point. The assertion that not every Hellinger speedless point is a singularity of $\theta \to \kappa(\theta)$ follows from the example in Illustration I. □

Proof of Proposition 4. This result is a direct application of the Delta Theorem for directionally differentiable statistics in Shapiro, Dentcheva and Ruszczynski (2009, Theorem 7.67, p.443). □

Proof of Proposition 5. Denote by gY_i the random variable with density $f_{\bar{g}\theta_0}$. By the invariance $\overline{g}\Theta_0 = \Theta_0$, one has

$$P_{\theta_0}(gY_i \in A) = P_{\overline{g}\theta_0}(Y_i \in A) \text{ for any } \theta_0 \in \Theta_0, \overline{g} \in \overline{G},$$

and any subset A in the sigma-algebra of $\mathcal{Y}$. Fix $\theta_0 \in \Theta_0$. By construction,

$$J\alpha = \sum_{\overline{g} \in \overline{G}} T_{\overline{g}\theta_0}.$$

Taking expectations to both sides

$$J\alpha = \sum_{\overline{g} \in \overline{G}} E_{\theta_0}(T_{\overline{g}\theta_0}).$$

By the invariance $\overline{g}\Theta_0 = \Theta_0$, we have $E_{\theta_0}(T_{\overline{g}\theta_0}) = E_{\theta_0}(T_{\theta_0})$. Hence,

$$J\alpha = JE_{\theta_0}(T_{\theta_0}) \text{ for any } \theta_0 \in \Theta_0,$$

which completes the proof. □

Appendix B: Endogenous Recursive Switching Model

This appendix presents the derivation of the individual score in the endogenous switching regression model described in Section 5. Recall that the model is

$$Y_{1i}^{\star} = X_{1i}'\beta_1 + Y_{2i}\beta_2 + Y_{2i}\tilde{X}_{1i}\beta_3 - U_{1i}$$
$$Y_{2i}^{\star} = X_i'\gamma_1 - U_{2i},$$

where $X_{1i} = (1, \tilde{X}_{1i})$, $X_i = (X_{1i}, X_{2i})'$, and $U_i = (U_{1i}, U_{2i})'$ is a standard normal bivariate random vector with a correlation coefficient ρ. The observed variables are (Y_{1i}, Y_{2i}, X_i), where $Y_{ji} = 1(Y_{ji}^{\star} \geq 0)$ for $j \in \{1, 2\}$. The joint density of the endogenous observed variables given the exogenous observed covariates for a given observation is:

$$\begin{aligned} f_\theta(y_i|x_i) = {} & P_0\big(Y_{1i}^{\star} \geq 0, Y_{2i}^{\star} \geq 0\big)^{y_{1i}y_{2i}} \times P_0\big(Y_{1i}^{\star} < 0, Y_{2i}^{\star} \geq 0\big)^{(1-y_{1i})y_{2i}} \\ & \times P_0\big(Y_{1i}^{\star} \geq 0, Y_{2i}^{\star} < 0\big)^{y_{1i}(1-y_{2i})} \times P_0\big(Y_{1i}^{\star} < 0, Y_{2i}^{\star} < 0\big)^{(1-y_{1i})(1-y_{2i})}, \end{aligned}$$

where $\theta = (\beta, \gamma, \rho)'$. I next derive the individual score

$$s_i(\theta) = \frac{\ln f_\theta(y_i|x_i)}{\partial\theta}$$

for this model. I use this individual score for calculating the randomization test statistic based on the Fisher matrix.

I start by rewriting the joint density of a given observation in a way amenable to the numerical computation of the individual scores. Define

$$t_{1i}(\beta) = x_{1i}'\beta_1 + y_{2i}\beta_2 + y_{2i}\tilde{x}_{1i}'\beta_3$$
$$t_{2i}(\gamma) = x_i'\gamma$$

Write $f_\theta(y_i|x_i)$ as

$$f_\theta(y_i|x_i) = P_0\big(U_{1i} \leq t_{1i}(\beta), U_{2i} \leq t_{2i}(\gamma)|x_i\big)^{y_{1i}y_{2i}} \times P_0\big(-U_{1i} < -t_{1i}(\beta), U_{2i} \leq t_{2i}(\gamma)|x_i\big)^{(1-y_{1i})y_{2i}}$$
$$\times P_0\big(U_{1i} \leq t_{1i}(\beta), -U_{2i} < -t_{2i}(\gamma)|x_i\big)^{y_{1i}(1-y_{2i})} P_0\big(-U_{1i} < -t_{1i}(\beta), -U_{2i} < -t_{2i}(\gamma)|x_i\big)^{(1-y_{1i})(1-y_{2i})}.$$

Let

$$\varphi_\rho(u_1, u_2) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\Big(-\frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{2 \times (1-\rho^2)}\Big)$$

denote the standard bivariate normal density with a correlation coefficient ρ. Let

$$\Phi_\rho(t_1, t_2) = \int_{-\infty}^{t_1}\int_{-\infty}^{t_2} \varphi_\rho(u_1, u_2) du_1 du_2$$

denote the standard bivariate normal cumulative distribution function with a correlation coefficient ρ. Write now $f_\theta(y_i|x_i)$ as

$$f_\theta(y_i|x_i) = \Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)^{y_{1i}y_{2i}} \times \Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)^{(1-y_{1i})y_{2i}}$$
$$\times \Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)^{y_{1i}(1-y_{2i})} \times \Phi_\rho\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)^{(1-y_{1i})(1-y_{2i})}.$$

I now calculate the individual scores. Using the chain rule, I have for the score associated to β

$$\begin{aligned}\frac{\partial \ln f_\theta(y_i|x_i)}{\partial \beta} &= \frac{y_{1i}y_{2i}}{\Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)}\frac{\partial \Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)}{\partial t_1}\frac{\partial t_{1i}(\beta)}{\partial \beta}\\ &+ \frac{(1-y_{1i})y_{2i}}{\Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)}\frac{\partial \Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)}{\partial t_1}\frac{\partial -t_{1i}(\beta)}{\partial \beta}\\ &+ \frac{y_{1i}(1-y_{2i})}{\Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)}\frac{\partial \Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)}{\partial t_1}\frac{\partial t_{1i}(\beta)}{\partial \beta}\\ &+ \frac{(1-y_{1i})(1-y_{2i})}{\Phi_\rho\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)}\frac{\partial \Phi_{-\rho}\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)}{\partial t_1}\frac{\partial -t_{1i}(\beta)}{\partial \beta}.\end{aligned}$$

Let

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-v^2/2)dv$$

denote the standard normal cumulative distribution function. Mukherjea et al. (1986) has noted

$$\frac{\partial f_\theta\big(t_1, t_2\big)}{\partial t_1} = \frac{1}{2\pi}\exp\Big(-\frac{t_1^2}{2}\Big)\Phi\Big(\frac{t_2}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}}t_1\Big).$$

Moreover, since

$$\frac{\partial t_{1i}(\beta)}{\partial \beta} = w_i, \text{ where } w_i = (x_i, y_{2i}, y_{2i}\tilde{x}_{1i})',$$

I have

$$\begin{aligned}
\frac{\partial \ln f_\theta(y_i|x_i)}{\partial \beta} &= \frac{y_{1i}y_{2i}}{\Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{t_{1i}(\beta)^2}{2}\bigg) \Phi\bigg(\frac{t_{2i}(\gamma)}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} t_{1i}(\beta)\bigg) w_i \\
&+ \frac{(1-y_{1i})y_{2i}}{\Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{-t_{1i}(\beta)^2}{2}\bigg) \Phi\bigg(\frac{t_{2i}(\gamma)}{\sqrt{1-\rho^2}} - \frac{(-\rho)}{\sqrt{1-\rho^2}} (-t_{1i}(\beta))\bigg) (-w_i) \\
&+ \frac{y_{1i}(1-y_{2i})}{\Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{t_{1i}(\beta)^2}{2}\bigg) \Phi\bigg(\frac{-t_{2i}(\gamma)}{\sqrt{1-\rho^2}} - \frac{(-\rho)}{\sqrt{1-\rho^2}} t_{1i}(\beta)\bigg) w_i \\
&+ \frac{(1-y_{1i})(1-y_{2i})}{\Phi_\rho\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{-t_{1i}(\beta)^2}{2}\bigg) \Phi\bigg(\frac{-t_{2i}(\gamma)}{\sqrt{1-\rho^2}} - \frac{(-\rho)}{\sqrt{1-\rho^2}} (-t_{1i}(\beta))\bigg) (-w_i).
\end{aligned}$$

Using a similar argument, I have for the score associated to γ

$$\begin{aligned}
\frac{\partial \ln f_\theta(y_i|x_i)}{\partial \gamma} &= = \frac{y_{1i}y_{2i}}{\Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{t_{2i}(\gamma)^2}{2}\bigg) \Phi\bigg(\frac{t_{1i}(\beta)}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} t_{2i}(\gamma)\bigg) x_i \\
&+ \frac{(1-y_{1i})y_{2i}}{\Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{-t_{2i}(\gamma)^2}{2}\bigg) \Phi\bigg(\frac{t_{1i}(\beta)}{\sqrt{1-\rho^2}} - \frac{(-\rho)}{\sqrt{1-\rho^2}} (-t_{2i}(\gamma))\bigg) (-x_i) \\
&+ \frac{y_{1i}(1-y_{2i})}{\Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{t_{2i}(\gamma)^2}{2}\bigg) \Phi\bigg(\frac{-t_{1i}(\beta)}{\sqrt{1-\rho^2}} - \frac{(-\rho)}{\sqrt{1-\rho^2}} t_{2i}(\gamma)\bigg) x_i \\
&+ \frac{(1-y_{1i})(1-y_{2i})}{\Phi_\rho\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)} \frac{1}{2\pi} \exp\bigg(-\frac{-t_{2i}(\gamma)^2}{2}\bigg) \Phi\bigg(\frac{-t_{1i}(\beta)}{\sqrt{1-\rho^2}} - \frac{(-\rho)}{\sqrt{1-\rho^2}} (-t_{2i}(\gamma))\bigg) (-x_i).
\end{aligned}$$

Finally, I have for the score associated to ρ

$$\begin{aligned}
\frac{\partial \ln f_\theta(y_i|x_i)}{\partial \rho} &= \frac{y_{1i}y_{2i}}{\Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)} \frac{\partial \Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)}{\partial \rho} \\
&+ \frac{(1-y_{1i})y_{2i}}{\Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)} \frac{\partial \Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)}{\partial \rho} \\
&+ \frac{y_{1i}(1-y_{2i})}{\Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)} \frac{\partial \Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)}{\partial \rho} \\
&+ \frac{(1-y_{1i})(1-y_{2i})}{\Phi_\rho\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)} \frac{\partial \Phi_{-\rho}\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)}{\partial \rho}.
\end{aligned}$$

Sungur (1990) has noted

$$\frac{\partial \Phi_\rho(t_1, t_2)}{\partial \rho} = \varphi_\rho(t_1, t_2).$$

Replacing this observation in the penultimate display,

$$\begin{aligned}\frac{\partial \ln f_\theta(y_i|x_i)}{\partial \rho} &= \frac{y_{1i}y_{2i}}{\Phi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big)}\varphi_\rho\big(t_{1i}(\beta), t_{2i}(\gamma)\big) \\ &+ \frac{(1-y_{1i})y_{2i}}{\Phi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big)}\varphi_{-\rho}\big(-t_{1i}(\beta), t_{2i}(\gamma)\big) \\ &+ \frac{y_{1i}(1-y_{2i})}{\Phi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big)}\varphi_{-\rho}\big(t_{1i}(\beta), -t_{2i}(\gamma)\big) \\ &+ \frac{(1-y_{1i})(1-y_{2i})}{\Phi_\rho\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big)}\varphi_\rho\big(-t_{1i}(\beta), -t_{2i}(\gamma)\big).\end{aligned}$$

Appendix C: Nonexistence of Regular Estimators

This appendix establishes the impossibility to construct regular estimators, in the sense defined by van der Vaart (1991), for Hellinger speedless points.

Let $\mathcal{P}$ be a class of probability functions on $(\mathcal{Y}, \mathcal{A})$, where $\mathcal{A}$ is a sigma-algebra on $\mathcal{Y}$. Let $\mathbb{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$. Set $\mathcal{P}_{\mathbb{H}} = \{P_\eta \in \mathcal{P} : \eta \in \mathbb{H}\}$. Fix η_0 in $\mathbb{H}$ and set $P_0 = P_{\eta_0}$. Let $L_{2,0}$ denote the set of P_0-square integrable functions with inner product $\langle s_1, s_2 \rangle_{P_0} = \int s_1 s_2 dP_0$. Let $\mathcal{H}_0$ be the class of paths $\epsilon \to \eta_\epsilon$ from $(0, \bar{\epsilon}) \subset \mathbb{R}_+$ to $\mathbb{H}$ such that:

$$\lim_{\epsilon \downarrow 0} \frac{\eta_\epsilon - \eta_0}{\epsilon} = \eta \tag{1}$$

for elements $\eta \in \mathbb{H}$. Let T_0 be the set of all η obtained. Assume that $\eta \to \eta_{h\epsilon}$ is in $\mathcal{H}_0$ for every $h \in \mathbb{R}_+$ if $\epsilon \to \eta_\epsilon$ is.

Furthermore, assume the existence of a continuous linear operator $A_0 : T_0 \to L_{2,0}$ satisfying

$$\lim_{\epsilon \downarrow 0} \int \left[\frac{dP_{\eta_\epsilon}^{1/2} - dP_0^{1/2}}{\epsilon} - \frac{1}{2} A_0 \eta dP_0^{1/2} \right]^2 = 0 \text{ for every } \eta_\epsilon \in \mathcal{H}_0 \text{ and some } \eta \in T_0. \tag{2}$$

To the set of paths $\epsilon \to P_{\eta_\epsilon}$ satisfying the condition in the last display corresponds a tanget set $\mathcal{T}_0$ consisting of all operators A_0 satisfying the condition in the last display. The tangent set is equal to the range of the operator A_0: $\mathcal{T}_0 = A_0 T_0 = R(A_0)$.

Let $\kappa : \mathcal{P}_{\mathbb{H}} \to \mathbb{B}$ be a function taking its values in an normed linear space $\mathbb{B}$. Assume that

$$\lim_{\epsilon \downarrow 0} \frac{\kappa(P_{\eta_\epsilon}) - \kappa(P_0)}{\epsilon} \text{ and } \kappa(P_\eta) = \varphi(\eta), \tag{3}$$

where $\varphi : \mathbb{H} \to \mathbb{B}$ is differentiable at η_0 in the sense that there exists a continuous linear map $\dot{\varphi}_0 : T_0 \to \mathbb{B}$ satisfying

$$\dot{\varphi}_0(\eta) = \lim_{\epsilon \downarrow 0} \frac{\varphi(\eta_\epsilon) - \varphi(\eta_0)}{\epsilon} \text{ for every } \eta_\epsilon \in \mathcal{H}_0. \tag{4}$$

The adjoint operator of A_0 is $A_0^\star : L_{2,0} \to T_0$ satisfying $\langle A_0\eta, s\rangle_{P_0} = \langle \eta, A_0^\star s\rangle_{\mathcal{H}}$ for every $\eta \in T_0$ and $s \in L_{2,0}$.

We have the following results:

Lemma 1. Let $N(\cdot)$ denote the null space of the operator $\cdot$ and let $\dot{\varphi}_0^\star$ denote the adjoint operator of $\dot{\varphi}_0$. Then,

$$\underset{\text{(range condition)}}{R(\dot{\varphi}_0^\star) \subset R(A_0^\star)} \implies \underset{\text{(nullspace condition)}}{N(A_0) \subset N(\dot{\varphi}_0)}$$

Proof. The claim follows from taking orthocomplements.

Lemma 2. $\underset{\text{(range condition)}}{R(\dot{\varphi}_0^\star) \subset R(A_0^\star)}$ holds if and only if there exists a continuous linear operator $\dot{\kappa}_0 : \mathcal{T}_0 \to \mathbb{B}$ such that

$$\dot{\kappa}_0(\eta) = \lim_{\epsilon \downarrow 0} \frac{\kappa(P_{\eta_\epsilon}) - \kappa(P_0)}{\epsilon} \text{ for every } \eta_\epsilon \in \mathcal{H}_0.$$
(differentiability condition)

Proof. van der Vaart (1991, Theorem 3.1).

Lemma 3. Consider estimators $T_N = t_N(Y_1, .., Y_N)$ of $\kappa(P_0)$ generated by maps $t_N : \mathcal{Y}^N \to \mathbb{B}$. Suppose that for every path $\epsilon P_{\eta_\epsilon}$ and every sequence $\{h_N \in \mathbb{R}\}_N$ such that $h_n \to h \in \mathbb{R}_+$ we have

$$\int^\star m\Big(\sqrt{N}\big(T_N - \kappa(P_{\eta_N})\big)\Big) dP_{\eta_N} \to \int m dL, \text{ where } \eta_N = \eta_{\epsilon_N} \text{ and } \epsilon_N = h_N/\sqrt{N},$$
(regularity condition)

for every continuous bounded function $m : \mathbb{B} \to \mathbb{R}$, where L is a fixed tight Borel law on $\mathbb{B}$ not depending on $h \in \mathbb{R}_+$.

Proof. van der Vaart (1991, Theorem 2.1).

An estimator satisfying the (regularity condition) in the last Lemma is called a regular estimator. Combining the Lemmata, we have the following chain of implications:

(regularity condition) $\Rightarrow$ (differentiability condition) $\Leftrightarrow$ (range condition) $\Rightarrow$ (nullspace condition)

For latter use, I deduce from this chain of implications the following Lemma:

Lemma 4. $\neg$(nullspace condition) $\Rightarrow$ $\neg$(differentiability condition) $\Rightarrow$ $\neg$(regularity condition).

Set $\mathbb{H} = \Theta$, $\eta_0 = \theta_0$, and $\varphi(\eta) = \theta$. We have $\dot{\varphi}_0 = I_K$, where I_K is the $K \times K$ identity matrix, and $A_0 = \mathcal{I}_{\theta_0}(\theta_0)$.

Proposition 7. If θ_0 is a Hellinger speedless point, then

(i) There is no function $\kappa : \mathcal{F} \to \Theta$ such that $\kappa(f_{\theta_0}) = \theta_0$ and

$$\dot{\kappa}(s) = \lim_{\epsilon \downarrow 0} \frac{\kappa(f_{\theta_\epsilon}) - \kappa(f_{\theta_0})}{\epsilon}, \text{ every } \theta_\epsilon \in \mathcal{H}_0,$$

for a continuous linear map $\dot{\kappa}_0 : \mathcal{T}_0 \to \Theta$.

(ii) There is no tight Borel law L on Θ, not depending on $h \in \mathbb{R}_+$, such that

$$\int m\Big(\sqrt{N}\big(T_N - \kappa(f_{\theta_N})\big)\Big) f_{\theta_N} d\mu \to \int m dL, \text{ where } \theta_N = \theta_{\epsilon_N} \text{ and } \epsilon_N = h_N/\sqrt{N},$$

for every continuous bounded function $m : \Theta \to \mathbb{R}$.

Proof of Proposition 7. Since θ_0 is a Hellinger speedless point, it follows from Proposition 1 that $N\big(\mathcal{I}_{\theta_0}(\theta_0)\big) \not\subset N(I_K)$. The claims then follow from Lemma 4.

Proposition 7 (ii) implies the nonexistence of regular estimators for Hellinger speedless points. The existence of regular estimators is a necessary condition for the validity of standard asymptotically normal inference based on Wald-type statistics. Furthermore, the existence of regular estimators is also a necessary condition for the consistency of standard parametric bootstrap inference; see e.g., Beran (1997). Since no regular estimator does exist for Hellinger speedless points, one cannot avoid using nonstandard asymptotic inference procedures in their presence.

Distinct but related results to Proposition 7 are the impossibility results in Hirano and Porter (2012) and in Kaji (2021). Hirano and Porter (2012, Theorem 2) show the nonexistence of regular estimators for the value of a function $\tau : \Theta \to \mathbb{R}$ that it is nondifferentiable at θ_0 and θ_0 is not a Hellinger speedless point, c.f. Hirano and Porter (2012, Assumption 1(b)) and Proposition 1. Kaji (2021, Theorem 2) shows the nonexistence of equivariant-in-law sequence of estimators for weakly regular parameters. Hellinger speedless points are not equivalent to the notion of weakly regular parameter, which is reproduced below for the sake of completeness. Fix $P_0 \in \mathcal{P}$. Let $\mathcal{P}_0$ denote the collection of paths $\epsilon \to P_\epsilon$ from $\epsilon \in (0,1]$ to $\mathcal{P}$ for which there exists a measurable function $s_0 \in L_{2,0}$ such that

$$\int \left[\frac{dP_\epsilon^{1/2} - dP_0^{1/2}}{\epsilon} - \frac{1}{2} s_0 dP_0^{1/2} \right]^2 \to 0 \text{ as } \epsilon \to 0.$$

Let $\mathcal{P}_0$ denote the collection of functions $s_0 \in L_{2,0}$ satisfying the last display and induced by any path in $\mathcal{P}_0$. Let $\mathcal{P}_b$ be a subset of $\mathcal{P}$ on which a parameter value $b \in \mathbb{B}$ is uniquely defined. Let $b : \mathcal{P}_b \to \mathbb{B}$ denote a function from $\mathcal{P}_b$ to the parameter space $\mathbb{B}$. The tangent set $\mathcal{S}_{0,b}$ pertinent to the submodel $\mathcal{P}_b$ is the set of scores s_0 such that there is exists a path $\epsilon \to P_\epsilon$ inducing s and every such paths shares the same limit of $b(P_\epsilon)$. Define $\mathcal{P}_{0,b}$ the set of paths inducing scores in $\mathcal{S}_{0,b}$. Kaji (2021) defines $b : \mathcal{P}_b \to \mathbb{B}$ as a weakly regular parameter at P_0 if there exists a map $b_0 : \mathcal{S}_{0,b} \to \mathbb{B}$ that is continuous and homogenous of degree zero such that $b(P_\epsilon) \to b_0(s_0)$ for every $P_\epsilon \in \mathcal{P}_{0,b}$. The definition above is for semi-parametric models. To compare with parametric models, set $\mathbb{B} = \Theta$ and $P_0 = P_{\theta_0}$. We have that $\{b(P_\epsilon) \in \Theta\}_\epsilon$ is a sequence in the parameter space. The notion of weakly regular parameter presuposses the existence of s_0 while the notion of Hellinger speedless point does not.

Appendix D: Subsampling Test

This Appendix constructs a resampling test for H_0 against H_1. It verifies that the test asymptotically controls size. For a subsample of size B_N, let $\hat{\underline{e}}_{B_N}(\theta_0)$ denote the smallest eigenvalue of the subsample analog of the Fisher matrix.

Step 1. Simulate under θ_0 a sample $\{Y_i\}_{i=1}^N$ of size N and compute the sample analog of the smallest Fisher matrix eigenvalue $\hat{\underline{e}}_N$.

Step 2. Sample without replacement S_N subsamples $\{Y_{i_j,s}\}_{j=1}^{B_N}$ of size $B_N < N$ from $\{Y_i\}_{i=1}^N$ and compute for each subsample s the subsample statistic $\hat{\underline{e}}_{B_N,s}$.

Step 3. Set the subsample critical value $cv_{B_N,\alpha}$ to the $1-\alpha$ quantile of $\{B_N^{1/2}(\hat{\underline{e}}_{B_N,s}-\hat{\underline{e}}_N)\}_{s=1}^{S_N}$.

Step 4. Reject H_0 if the test statistic $\sqrt{N}\hat{\underline{e}}_N$ is bigger than the subsample critical value $cv_{B_N,\alpha}$.

The following conditions guarantee the asymptotic validity of this procedure to keep size control. Let ${}^N C_{B_N} := \frac{N}{B_N!(N-B_N)!}$ denote the number of subsamples of size B_N.

Proposition 6. Let the assumptions of Proposition 4 hold. Also assume that: (i) $S_N \rightarrow^N C_{B_N}$, $B_N \rightarrow \infty$ and $B_N/N \rightarrow 0$ as $N \rightarrow \infty$; and (ii) J_{θ_0} is continuous at its $1-\alpha$ quantile. Then,

$$\lim_{N\rightarrow\infty} P_{\theta_0}(\sqrt{N}\hat{\underline{e}}_N \leq cv_{B_N,\alpha}) = \alpha \text{ whenever } \theta_0 \in \Theta_0.$$

Proof of Proposition 6. This result follows from Proposition 4 by a direct application of a result in Politis et al. (1999, Theorem 2.2.1, p. 43). $\square$

The study of the numerical properties of this subsampling test is out of the scope of this paper and it is left for future research.

References in the Appendices

Beran, R. (1997): "Diagnosing Bootstrap Success", *Annals of the Institute of Mathematical Statistics*.

Coleman, R. (2012): *Calculus on Normed Vector Spaces*, Springer, New York.

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

Mukherjea, A. A. Nakasis, J. Miyashita (1986): "Identification of Parameters by the Distribution of the Maximum Random Variable: The Anderson-Ghurye Theorems", *Journal of Multivariate Analysis*.

Politis, D., J. Romano and M. Wolf (1999): *Subsampling*, Springer.

Shapiro, A., D. Dentcheva and A. Ruszczynski (2014): *Lectures on Stochastic Programming: Modeling and Theory*, SIAM.

Sungur, E. (1990):"Dependence Information in Parametrized Copulas ", *Communications in Statistics: Simulation and Computations*.