Journal of Statistical Planning and Inference 174 (2016) 68–84

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Extension of the Schwarz Information Criterion for models sharing parameter boundaries

Le-Yu Chen [a,*], Jerzy Szroeter [b]

[a] *Institute of Economics, Academia Sinica, Taiwan*

[b] *Department of Economics, University College London, UK*

ARTICLE INFO

Article history:
Received 20 November 2014
Received in revised form 20 January 2016
Accepted 28 January 2016
Available online 4 February 2016

Keywords:
Model selection
Schwarz criterion
Consistency
Boundary
Dimension
Inequalities

ABSTRACT

The classic Schwarz Information Criterion, originally derived as an approximation to Bayes posterior probability, is widely used as a standalone likelihood-based measure of model fit. However, selection consistency is compromised when model sets partially include their parameter boundaries and when these in turn are partially shared by different models. This happens, for example, where sets represent mixed weak and strict inequality restrictions on parameters. To enable consistent selection of such models, a generic extension of the Schwarz criterion is required but does not appear to be available in the literature to date. In this paper, we define the boundary extended Schwarz criterion for a model to be the maximum of Schwarz-type criteria applied to the model parameter space and a systematically-generated list of boundary subsets. This entails new concepts of boundary interaction level and model dimension. A self-contained theory is presented along with examples and simulation.

1. Introduction

1.1. Objectives and literature

The classic Schwarz Information Criterion, originally derived as an approximation to Bayes posterior probability, is widely used as a standalone likelihood-based measure of model fit within the frequentist paradigm of statistical inference. In consequence, there is considerable interest in its sampling properties. Two particular concerns are consistency (choosing a smallest true model with probability tending to one as sample size increases) and efficiency (minimizing the expected loss in some sense of using the selected model with estimated rather than known parameters for some post-sample purpose such as prediction). These two properties are more fully described in McQuarrie and Tsai (1998, Chapter 1) and Claeskens and Hjort (2008, Chapter 4) who point out that consistency of Schwarz-type (1978) criteria comes at a price in terms of efficiency, this being reversed for Akaike-type (1974) criteria. In the present paper, we focus on the consistency aspect. Our concern is that the basic form of the Schwarz criterion actually fails to be consistent when model sets partially include their parameter boundaries and when these in turn are partially shared by different models. This happens, for example, where the sets represent mixed weak and strict inequality restrictions of various types.

The work of Sin and White (1996) on penalized likelihood criteria under conditions allowing misspecification is often cited to support the assertion that the Schwarz criterion is generally consistent. It does not, however, cover the situations

* Corresponding author.
E-mail addresses: lychen@econ.sinica.edu.tw (L.-Y. Chen), j.szroeter@uclmail.net (J. Szroeter).

http://dx.doi.org/10.1016/j.jspi.2016.01.010
0378-3758/

addressed in the present paper. For it proceeds (Proposition 4.2) under the assumption that the (Kullback–Leibler) pseudo-true parameter value of each model under choice lies in the interior of the particular parameter space of that model. This is somewhat restrictive for models having a common form of data density function but defined on different subsets of a larger space. For example, the model defined as an $N(\mu, 1)$ distribution on $\mu \in [a, b]$, some compact interval not containing the true μ_0, has pseudo-true μ value equal to either a or b.

The importance of the boundary (in part or in whole) is a matter of context. Some part of the boundary may well be deemed to be of no practical relevance. Other parts may have serious empirical meaning and represent a real-world feature perceived and acted upon as true. In Section 1.2 and Section 7, we give an example where correct policy advice to governments hinges on the boundary. Thus, the boundary cannot in general be re-allocated or ignored for the convenience of a statistical method. In Section 2.2 we give simple examples to illustrate how inconsistency on the boundary arises.

To deal with the inconsistency by a generally applicable principle, we define the boundary extended criterion for each model as the maximum of basic Schwarz-type criteria applied to the model parameter space and a systematically-generated list of boundary subsets. The length of the list depends on the boundary interaction level between the relevant model sets. A precise definition of the concept of boundary interaction level will be given in Section 3. Operationally a higher interaction level implies more terms are needed in the extension to ensure selection consistency. This in turn rests on two verifiable high-level regularity conditions on asymptotic behavior of likelihood ratios and two assumptions on dimension measure for parameter sets. The latter are required because the technical concept of model size in standard Schwarz theory is manifold dimension (loosely, the number of "free" parameters). In the present paper, however, the model and boundary parameter sets (original and derived) are not generally manifolds hence require a broader notion of set dimension. The framework, problems, examples and assumptions are presented in Sections 2.1–2.4.

The keystone of our approach is a thorough treatment of the boundary issue for binary selection between disjoint sets (models) whose closures may nonetheless have non-empty intersection This is presented in Sections 3.1–3.3. Section 3.4 illustrates the solution for two basic model types. We confirm that the general boundary-extended criterion reduces to the standard Schwarz version under parametric equality constraints and we show its specific form in models defined by mixed weak and strict inequalities.

In Section 4, the binary selection theory is applied to solve the multiple model selection problem: This is to consistently infer which of a given collection of models defined by subsets of common parameter space are true and which false allowing subset closures to overlap. When the models are disjoint or nested within each other, such an inference implies inferring the smallest true model. In Section 5 we indicate how the approach accommodates nuisance parameters. Section 6 illustrates theory by simulation. Section 7 illustrates the theory with an application to industrial economics. Section 8 concludes. Appendices A and B contain technical proofs.

We end this section with a review of related literature. Within the classical estimation and test branches of inference, substantial work exists on the effects of and methods for dealing with boundary-located true parameter values and the inequalities by which they arise. For this literature, see Andrews (1999, 2001, 2002), Silvapulle and Sen (2005), Chernozhukov et al. (2007), Andrews and Soares (2010), Chen and Liang (2010), Andrews and Barwick (2012) and references therein. Within the selection branch of inference by information or measure of fit criteria, Anraku (1999), Hughes and King (2003) and Kuiper et al. (2011, 2012) extend the Akaike (1974) criterion to models subject to parameter inequalities defining a closed convex cone but do not address the boundary issues raised in the present paper.

It is well-known that the Laplace analysis deriving the Schwarz (1978) statistic as an approximation of posterior probability fails at parameter points on the boundary. The modified criteria of Erkanli (1994, 1997), Hsiao (1997) and Pauler et al. (1999) improve that approximation but are not automatically consistent. Indeed, Hsiao (1997, p.662), has observed that the Bayes–Schwarz criterion in Haughton (1988) "may not be consistent in the boundary case". Dudley and Haughton (1997) have proposed two types of user-supplied bonus terms to enhance the Schwarz statistic. The first improves the Schwarz approximation to posterior probability. The second is a fixed addition to the Schwarz criterion to favor smaller models contained in larger ones when there is no dimensional difference. Nevertheless, there remains an inconsistency issue of the type we note after Example 2 in Section 2.2. The present paper also focuses on Schwarzian dimension-penalized likelihood. We do not, however, pursue Bayesian approximation issues but seek instead to extend the concept so as to systematically guarantee consistency at all boundary points.

1.2. Contextual example

As an example of the importance of boundaries in a real-world setting, we consider the production function, a structural relation which indicates the maximum output of goods or services by a firm in some specific industry for any stated levels of inputs such as capital, labor, fuel and raw materials. For the basic economic theory, the reader is referred to Besanko and Braeutigam (2011, Ch.6). For the specification of structural economic models to match theory to data, the reader is referred to Reiss and Wolak (2007) and Wolff (2014). Here it suffices to consider the Cobb–Douglas form

$$Y = \alpha K^{\theta_1} L^{\theta_2} F^{\theta_3} U \tag{1.1}$$

where Y and K, L, F denote observable measures of output and labor, capital, fuel inputs respectively and $\alpha, \theta_1, \theta_2, \theta_3$ are parameters which are fixed across all firms. The term U is an unobservable variable unique to each firm and represents