Article:
BRAVO, FRANCESCO (2026) A uniform model selection test for semiparametric models. Statistics & Probability Letters. 110658. ISSN: 0167-7152

https://doi.org/10.1016/j.spl.2026.110658

A uniform model selection test for semiparametric models

Francesco Bravo

Department of Economics
University of York, United Kingdom

Abstract

This paper proposes a new simple test for model selection between two possibly misspecified competing semiparametric models. An important feature of the test is that it controls uniformly its size regardless as to whether the competing models are nested, overlapping or non nested and can be applied to overidentified models with weakly dependent observations.

Keywords: Exponential tilting, Strong mixing, Uniform size control
2020 MSC: 62G05, 62G10

1. Introduction

Let $\{Z_{nt}, n \geq 1, 1 \leq t \leq n\}$ be a triangular array of random vectors on the probability space $(\Omega, \mathcal{Z}, P)$, and let

$$M_{nf}(\beta, h) = E_{P_f}(f(Z_{nt}, \beta, h)), \quad M_{ng}(\theta, l) = E_{P_g}(g(Z_{nt}, \theta, l))$$

denote two competing semiparametric models, where $f : \mathbb{R}^{d_Z} \times B \times \mathcal{H} \to \mathbb{R}^{d_f}$, $g : \mathbb{R}^{d_Z} \times \Theta \times \mathcal{L} \to \mathbb{R}^{d_g}$, $B \subset \mathbb{R}^k$, $\Theta \subset \mathbb{R}^l$, $\mathcal{H} = \mathcal{H}_1 \times .... \times \mathcal{H}_f$, $\mathcal{L} = \mathcal{L}_1 \times ... \times ....\mathcal{L}_g$ are pseudo metric spaces of functions, that is $M_{nf}(\beta, h)$ and $M_{ng}(\theta, l)$ consist of the set of all distributions defined by the moment conditions $E_{P_f}(f(\cdot))$ and $E_{P_g}(g(\cdot))$. $M_{nf}(\beta, h)$ is said to be correctly specified if for all t, n $M_{nf}(\beta, h) = 0$ for some $\beta \in B, h \in \mathcal{H}$ and misspecified if for all t, n, $M_{nf}(\beta, h) \neq 0$ for all $\beta \in B, h \in \mathcal{H}$ (and similarly for $M_{ng}(\theta, l)$).

Email address: `francesco.bravo@york.ac.uk` (Francesco Bravo)
URL: `https://sites.google.com/a/york.ac.uk/francescobravo` (Francesco Bravo)

Valuable comments from an Associate Editor and two referees are gratefully acknowledged. The usual disclaimer applies.

This paper proposes a test for model selection between possibly misspecified $M_{nf}(\beta, h)$ and $M_{ng}(\theta, l)$, which is quite general because it applies to possibly overidentified models with weakly dependent observations, which is empirically relevant because, for example, it is well known that many commonly used asset pricing and economics models (see for example Hansen and Jagannathan (1997) and Hansen and Marinacci (2016), respectively) are misspecified and are characterized by observations that exhibit some form of serial dependence. The test has a number of important and desirable features: First, it is based on estimators that have an information theoretic interpretation (as the unique minimizers of the Kullback-Liebler distance between the unknown distribution of the observations and the empirical distribution). This compares favorably with the alternative test based on generalized method of moments estimators, which are not unique as they depend on the choice of the weighting matrix (see Hall and Pelletier (2011) for more details). Second, it uniformly controls its size, which is an important property because if the model selection procedure were based only on pointwise correct asymptotic levels, for any sample size N there would exist a sequence of distributions P_n consistent with the null hypothesis such that for any $n \geq N$, the rejection probability under such a sequence would be arbitrarily close to one. This possibility is ruled out when the test is uniformly of correct asymptotic level, that is, for any $\varepsilon > 0$, there is a sample size N such that, for all $n \geq N$ the rejection probability under any sequence P_n consistent with the null hypothesis is at most $\alpha + \varepsilon$. Third, the test statistic is asymptotically normal regardless as to whether the two competing models are non nested, overlapping or nested [2] under the null hypothesis that both models are observationally equivalent (that is they have the same Kullback-Liebler distance). Finally, the test is computationally easy to implement. The test of this paper is related to the one proposed by Liao and Shi (2020), which also has uniform size control and has power against $n^{1/2}$ local alternatives. The two tests complement each other because they are based on different semiparametric models, a different way of achieving uniform size control and a different sampling scheme. From a technical point of view, the main contribution of the paper is to derive a central limit theorem for semiparametric models with triangular arrays of strong mixing random vectors which extends results of Andrews and Pollard (1994), Ekström (2014) and

[2] The two semiparametric models $M_{nf}(\beta, h)$ and $M_{ng}(\theta, l)$ are (strictly) non nested if given the true distribution P_0 of the observations, $P_f \cap P_g = \varnothing$; they are overlapping if $P_0 \cap (P_f \cap P_g) \neq \varnothing$ with $P_f \subsetneq P_g$ and $P_g \supsetneq P_f$; they are nested if either $P_f \subset P_g$ (P_f is nested in P_g) or $P_g \subset P_f$ (P_g is nested in P_f).

Bravo et al. (2017) among others.
The main results of the paper are contained in the next section, whereas Section 3 briefly discusses how to choose the regularization parameter $\widehat{\epsilon}_n$. A Supplemental Appendix contains the regularity conditions, all the proofs, a detailed discussion on how to choose the regularization parameter $\widehat{\epsilon}_n$ and some simulation results.

2. Main results

The estimators for $M_{nf}(\beta, h)$ and $M_{ng}(\theta, l)$ are a semiparametric extension of the exponential tilting method proposed by Kitamura (2000) in the context of misspecified parametric estimating equations models. They are two-step in nature and rather general because they simply assume the existence of preliminary consistent (in a suitable norm) estimators, say $\widehat{h}_n$ and $\widehat{l}_n$, of the unknown infinite dimensional parameters h, l. The estimators are

$$\left[\widehat{\beta}_n', \widehat{\lambda}_n'\right]' = \arg\max_{\beta\in B} \arg\min_{\lambda\in\Lambda(B)} \mathcal{X}_{nf}\left(\beta, \lambda, \widehat{h}_n\right) \tag{1}$$

$$\left[\widehat{\theta}_n', \widehat{\gamma}_n'\right]' = \arg\max_{\theta\in\Theta} \arg\min_{\gamma\in\Gamma(\Theta)} \mathcal{X}_{ng}(\theta, \gamma, \widehat{l}_n) \tag{2}$$

where $\Lambda(B) \subset \mathbb{R}^{d_f}$, $\mathcal{X}_{nf}(\beta, \lambda, h) = \sum_{t=1}^{n} \exp\left(\lambda' f_{nt}(\beta, h)\right)/n$, $\lambda =: \lambda(\beta, h)$, $f_{nt}(\beta, h) =: f(Z_{nt}, \beta, h)$, $\Gamma(\Theta) \subset \mathbb{R}^{d_g}$, $\mathcal{X}_{ng} = \sum_{t=1}^{n} \exp(\gamma' g_{nt}(\beta, l))/n$, $\gamma =: \gamma(\theta, l)$ and $g_{nt}(\beta, l) =: g(Z_{nt}, \beta, l)$. It is important to note that the solution to (1) corresponds to

$$\mathcal{X}_{nf}\left(\widehat{\beta}_n, \widehat{\lambda}_n, \widehat{h}_n\right) = \exp\left(-\min_{P\in P_f} KL(P, P_n)\right),$$

and similarly for (2), where $KL(\cdot,\cdot)$ is the Kullback-Liebler distance, that is, the proposed two-step semiparametric estimator can be interpreted as the one that estimates among the set of distributions P_f defined by $M_{nf}(\beta, h)$ the closest one, as measured by the Kullback-Liebler distance, to the empirical distribution P_n. In this context, β_* is the pseudo-true value assumed to be the unique minimizer of $\min_{P\in P_f} KL(P, P_0)$, where P_0 is the true unknown distribution of the observations, and λ_* is the unique minimizer of the corresponding convex minimization population problem $\min_{\lambda\in\Lambda(B)} E(\exp(\lambda' f(Z_{nt}, \beta_*, h_*)))$ and similarly for θ_* and γ_*. The null hypothesis is

$$H_0 : E(\mathcal{X}_{nf}(\beta_*, \lambda_*, h_*)) = E(\mathcal{X}_{ng}(\theta_*, \gamma_*, l_*)), \tag{3}$$

that is the two competing semiparametric models are observationally equivalent in terms of their Kullback-Liebler distance, which implies that no model selection is possible. To test for (3) we use the same sample splitting technique as that of Schennach and Wilhelm (2017), which avoids the discontinuity problem[3] of Vuong (1989)'s model selection procedure, and propose the following test statistic

$$V_n\left(\widehat{\epsilon}_n\right)=\frac{1}{n^{1/2}}\sum_{t=1}^{n}\left(\omega_t\left(\widehat{\epsilon}_n\right)\exp\left(\widehat{\lambda}_n' f_{nt}\left(\widehat{\beta}_n,\widehat{h}_n\right)\right)-\omega_{t+1}\left(\widehat{\epsilon}_n\right)\exp\left(\widehat{\gamma}_n' g_{nt}\left(\widehat{\theta}_n,\widehat{l}_n\right)\right)\right), \tag{4}$$

where $\left(\omega_t\left(\widehat{\epsilon}_n\right)\right)_{t=1}^{n+1}$ is a sequence of weights depending on a possibly data-dependent real valued regularization parameter $\widehat{\epsilon}_n$ such that

$$\omega_t\left(\widehat{\epsilon}_n\right)=\begin{cases}1 \text{ for } t \text{ odd}\\ 1+\widehat{\epsilon}_n \text{ for } t \text{ even}\end{cases}$$

A straightforward calculation shows that the asymptotic variance of (4) $\sigma_*^2\left(\beta_*,\lambda_*,h_*,\theta_*,\gamma_*,l_*,\epsilon_n\right):=\sigma_*^2\left(\epsilon_n\right)$ is

$$\begin{aligned}\sigma_*^2\left(\epsilon_n\right)=\left(1+\epsilon_n\right)\sigma_*^2+\frac{\epsilon_n^2}{2}\lim_{n\to\infty}&\left[Var\left(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\left(\exp\left(\lambda_*' f_{nt}\left(\beta_*,h_*\right)\right)\right)\right)+\right.\\ &\left.Var\left(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\exp\left(\gamma_*' g_{nt}\left(\theta_*,l_*\right)\right)\right)\right],\end{aligned} \tag{5}$$

where

$$\sigma_*^2=\lim_{n\to\infty}Var\left(\frac{1}{n^{1/2}}\sum_{t=1}^{n}(\exp(\lambda_*' f_{nt}(\beta_*,h_*)-\exp(\gamma_*' g_{nt}(\theta_*,l_*))\right),$$

and ϵ_n is the (possibly probability) limit of $\widehat{\epsilon}_n$. Thus $\sigma_*^2(\epsilon_n)$ is always positive regardless as to whether the two competing models are non nested (which implies $\sigma_*^2>0$) or overlapping (which implies $\sigma_*^2=0$). Let $\widehat{\sigma}_n^2(\widehat{\beta}_n,\widehat{\lambda}_n,\widehat{h}_n,\widehat{\theta}_n,\widehat{\gamma}_n,\widehat{\gamma}_n,\widehat{l}_n,\widehat{\epsilon}_n)$ denote an estimator[4] for $\sigma_*^2(\epsilon_n)$,

$$t_n(\epsilon_n)=\frac{V_n(\epsilon_n)}{\widehat{\sigma}_n(\widehat{\beta}_n,\widehat{\lambda}_n,\widehat{h}_n,\widehat{\theta}_n,\widehat{\gamma}_n,\widehat{l}_n,\widehat{\epsilon}_n)},$$

[3]See Section 1 in the supplemental Appendix for a discussion of this important point.
[4]See Section 4 in the supplemental Appendix for an example of such an estimator.

and let $\mathcal{P}_*$ denote the set of distributions that satisfy Assumptions A1-A6 (given in the supplemental Appendix) and the null hypothesis (3).

Theorem 1. *Under Assumptions A1-A6 and the null hypothesis* (3)

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_*} P\left(|t_n\left(\widehat{\epsilon}_n\right)| \geq z_{1-\alpha/2}\right) = \alpha.$$

Next, we consider the local power of $t_n(\widehat{\epsilon}_n)$, and consider the case where

$$n^{1/2}(E(\mathcal{X}_{fn}(\beta_*, \lambda_*, h_*)) - E(\mathcal{X}_{ng}(\theta_*, \gamma_*, l_*))) \to \pi, \tag{6}$$

as $n \to \infty$ for some $\pi \in \mathbb{R}$. Let $\mathcal{P}_{n\pi}$ denote the sequence of distributions satisfying Assumptions A1-A7 (in the supplemental Appendix) along which (6) is satisfied. Let $P_n \in \mathcal{P}_{n\pi}$.

Theorem 2. *Under A1-A7 and P_n,*

$$t_n(\widehat{\epsilon}_n) \xrightarrow{d} N(\Pi, 1),$$

where

$$\Pi = \lim_{n\to\infty} \frac{n^{1/2}(E(\mathcal{X}_{fn}(\beta_*, \lambda_*, h_*)) - E(\mathcal{X}_{ng}(\theta_*, \gamma_*, l_*)))(1+\epsilon_n/2)}{((1+\epsilon_n)\sigma_*^2 + \epsilon_n^2(\sigma_{g*}^2 + \sigma_{f*}^2))^{1/2}}.$$

Note that the mean Π coincides with that derived by Schennach and Wilhelm (2017) (hence it is always finite regardless as to whether $\sigma_*^2 > 0$ or $\sigma_*^2 \to 0$) because under assumption A5 the estimation effect of the infinite dimensional parameters is asymptotically negligible. Note also that if $\pi \to \infty$ the test is consistent.

3. Choosing the regularization parameter

We briefly discuss how to choose the regularization parameter $\widehat{\epsilon}_n$, which can be data dependent or a deterministic sequence ϵ_n. The latter can be a finite positive constant or more generally can vary with n as long it satisfies the requirement $n^{1/4}\epsilon_n \to \infty$ (see A6 in the supplemental Appendix). For such a choice the proposed test is asymptotically valid, but caution should be warranted because it leaves open the possibility of a user choosing a value that produces a desired outcome. For the data dependent choice one could start with the procedure suggested by Schennach and Wilhelm (2017), which is optimal (in the sense of simultaneously minimizing the size distortion and potential power loss) in the absence of infinite dimensional parameters, and then perform a grid search around the chosen $\widehat{\epsilon}_n$. In the supplemental Appendix we provide a rigorous theoretical justification of the Schennach and Wilhelm (2017) data dependent regularization parameter choice for possibly nonstationary strong mixing processes without infinite dimensional parameters, which is of independent interest, and provide some simulation evidence about the finite sample performance of the proposed test under different choices of $\widehat{\epsilon}_n$, which seems to suggest that a simple grid search about the Schennach and Wilhelm (2017) choice of $\widehat{\epsilon}_n$ delivers a model selection procedure with good finite sample size and power properties.

References

Andrews, D., Pollard, D., 1994. An introduction to functional central limit theorems for dependent stochastic processes. International Statistical Review 62, 119–132.

Bravo, F., Chu, B., Jacho-Chavez, D., 2017. Semiparametric estimation of moment conditions models with weakly dependent data. Journal of Nonparametric Statistics 29, 108–136.

Ekström, M., 2014. A general central limit theorem for strong mixing sequences. Statistics and Probability Letters 94, 236–238.

Hall, A., Pelletier, D., 2011. On non-nested testing in models estimated by generalized methods of moments. Econometric Theory 27, 443–456.

Hansen, L., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factors models. Journal of Finance 52, 557–590.

Hansen, L., Marinacci, M., 2016. Ambiguity aversion and model misspecification: An economic perspective. Statistical Science 31, 511–515.

Kitamura, Y., 2000. Comparing misspecified dynamic econometric models using nonparametric likelihood. Technical Report. University of Pennsylvania.

Liao, Z., Shi, X., 2020. A uniform Vuong test for semi/nonparametric models. Quantitative Economics 11, 983–1017.

Schennach, S., Wilhelm, D., 2017. A simple parametric model selection test. Journal of the American Statistical Association 112, 1663–1674.

Vuong, Q., 1989. Likelihood ratio tests for model selection and non nested hypothesis. Econometrica 57, 307–333.