Union-Intersection and Sample-Split Methods in Econometrics with Applications to MA and SURE Models *

Jean-Marie Dufour † and Olivier Torrès ‡

January 27, 1997
Published: 1998 §

*We thank Eugene Savin, the Editor David Giles and an anonymous referee for several useful comments. This work was supported by grants from the Social Sciences Research Council of Canada, the Natural Sciences and Engineering Council of Canada, and the Government of Québec (Fonds FCAR).

† CRDE and Département de Sciences Économiques, Université de Montréal, CP 6128, succursale Centre-ville, Montréal, Canada H3C 3J7. e-mail: `jean.marie.dufour@umontreal.ca`.
Web page: `http://www.fas.umontreal.ca/SCECO/Dufour`

‡ GREMARS, Université de Lille 3, BP 149, 59653 Villeneuve d'Ascq cedex, France.
e-mail: `torres@univ-lille3.fr`.

§ This article was published in the *Handbook of Applied Economic Statistics*, edited by David Giles and Aman Ullah, Marcel Dekker, New York, 1998, Chapter 14, 465-505. This version contains corrections of a few typos in the published article.

ABSTRACT

In this paper, we develop inference procedures (tests and confidence sets) for two apparently distinct classes of situations: first, problems of comparing or pooling information from several samples whose stochastic relationship is not specified; second, problems where the distributions of standard test statistics are difficult to assess (*e.g.*, because they involve unknown nuisance parameters), while it is possible to obtain more tractable distributional results for statistics based on appropriately chosen subsamples. A large number of econometric models lead to such situations, such as comparisons of regression equations when the relationship between the disturbances across equations is unknown or complicated: seemingly unrelated regression equations (SURE), regressions with moving average (MA) errors, etc. To deal with such problems, we propose a general approach which uses union-intersection techniques to combine tests (or confidence sets) based on different samples. In particular, we make a systematic use of Boole-Bonferroni inequalities to control the overall level of the procedure. This approach is easy to apply and transposable to a wide spectrum of models. In addition to being robust to various misspecifications of interest, the approach studied turns out to have surprisingly good power properties with respect to other available techniques (*e.g.*, various asymptotically motivated methods and other bounds procedures). Applications to inference in SURE and regressions with MA(q) errors are discussed in detail. In the latter case, we also present an extensive Monte Carlo study, demonstrating the advantages of the sample-split approach. Finally, the methods proposed are applied to a demand system for inputs, a multivariate return to schooling model, and a time series model of Canadian *per capita* GDP.

Contents

1. Introduction **1**

2. Hypothesis testing: general theory **4**
2.1. H_0 as the finite intersection of subhypotheses 4
2.2. H_0 as the union of subhypotheses 6
2.2.1. Intersection of confidence intervals: the sum of critical points rule 8
2.2.2. Intersection of two confidence ellipsoids 9
2.2.3. Intersection of two confidence boxes 10

3. Confidence set estimation **11**

4. Exact inference in SURE models **12**
4.1. The model and the procedures 12
4.2. Some examples 14
4.2.1. Testing restrictions in a system of demands for inputs 14
4.2.2. Testing restrictions on returns to schooling 15

5. Exact inference in linear regression models with MA(q) errors **16**
5.1. A test on the mean of a general MA(q) model 17
5.2. Exact inference in the context of a MA(1) process 17
5.2.1. An induced test on the mean 17
5.2.2. Alternative procedures 18
5.2.3. Simulations 21
5.2.4. An example: an induced test on the mean of the Canadian per capita GDP series 33

6. Concluding remarks **36**

List of Tables

1	Zinde-Walsh and Ullah's bounds	19
2	Size and critical values of 5% level asymptotic tests	21
3	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 1$	27
4	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = .5$	28
5	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 0$	29
6	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -0.5$	30
7	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -1$	31
8	Sample autocorrelations of the Canadian per capita GDP series	34
9	Induced and asymptotic tests. Model: $y_t = \beta + \varepsilon_t + \psi\varepsilon_{t-1}$	34
10	Induced and asymptotic tests. Model: $y_t = \beta_0 + \beta_1 t + \varepsilon_t + \psi\varepsilon_{t-1}$	35
11	Induced and asymptotic tests. Model: $y_t = \beta + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2}$	36

List of Figures

1	97.5% confidence ellipsoids and intervals in the Berndt example	16
2	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 1$	22
3	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 0.5$	23
4	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 0$	24
5	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -.5$	25
6	Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -1$	26
7	First differences of the Canadian per capita GDP	33

1. Introduction

Straightforward application of usual inference procedures (tests and confidence regions) in econometrics is often impossible. The problem usually comes from an insufficient specification of the probability distribution generating the data, as occurs for example when one makes assumptions only on the first few moments of an error distribution. However, the problem also arises in parametric models that specify the data generating process up to a finite vector of parameters. This is typically the case when the assumptions made on the distribution depart from those made in the standard linear regression framework, such as the absence of serial independence and homoskedasticity.

This paper treats in a unified way two apparently distinct categories of problems where distributional results are difficult to establish. The first one consists of comparing and pooling information about parameter estimates from samples whose stochastic relationship is totally unspecified. In such cases, it is not possible to write a usable joint likelihood function and standard finite sample or asymptotic methods are not applicable. The second one consists of making inferences in models for which the distributions of standard test and confidence set procedures are difficult to establish, *e.g.* because of the presence of nuisance parameters, but for which relevant test statistics based on appropriately selected subsamples are distributionally more tractable.

To illustrate the problems we will study, consider the case where we have $m \geq 1$ regression equations of the form:

$$\begin{aligned} y_{i,t} &= \boldsymbol{x}_{i,t}'\boldsymbol{\beta}_i + u_{i,t} :, \qquad t = t_i + 1, \, \ldots \, , t_i + N_i \, , \\ \boldsymbol{u}_i &= (u_{i,t_i+1}, \, \ldots \, , u_{i,t_i+N_i})' \sim N(0, \Omega_i), \qquad i = 1, 2, \, \ldots \, , m, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}_i$ is an unknown $k_i \times 1$ vector ($k_i < N_i$), $\boldsymbol{x}_{i,t}$ is a $k_i \times 1$ vector of fixed (or strictly exogenous) regressors and Ω_i is an unknown positive definite nonsingular $N_i \times N_i$ matrix, $i \in I = \{1, 2, \, \ldots \, , m\}$. This setup describes situations frequently met in econometrics. Special cases of interest include the following ones:

(a) models in which each equation expresses a similar relationship between analogous variables (*i.e.*, the coefficients β_i have the same economic interpretation), but each one corresponds to a different sample and the different samples may be dependent in a way that is difficult to specify (*e.g.*, this includes many panel data models);

(b) models with structural change: this situation is a special case of the previous one, where the different samples correspond to different subperiods;

(c) stacked regressions where each equation represents a different economic relation, possibly with different numbers of regressors which have different economic interpretations;

(d) time series models where the dependence between the m equations in (1) is induced by serial dependence.

For example, take $m = 2$. A model of type (a) could express the relation between the log of the wage and a variable measuring the level of education for two individuals. The coefficient β is then interpreted as the return to education [see Ashenfelter and Krueger (1992), for instance], and we may wish to test whether this return is the same for individuals 1 and 2. In models of type (b), we may wish to know whether the parameter linking variable y to variable x is the same over the whole period of observation. An example of a type (c) model could be two equations where $y_{1,t}$ and $y_{2,t}$ represent the consumption of two different goods and $\boldsymbol{x}_{1,t}$ and $\boldsymbol{x}_{2,t}$ are different vectors of explanatory variables. Model (c) is composed of two distinct relationships, but for some reason, we want to test the equality of the two coefficients. An important example of a type (d) model is a linear regression model with errors that follow a moving average (MA) process of order one, where the first equation contains the odd-numbered observations and the second equation the even-numbered observations.

The most common practice in such situations is to rely on asymptotic inference procedures. The lack of reliability of such methods is well documented in the literature. This feature of asymptotic tests has been established by Park and Mitchell (1980), Miyazaki and Griffiths (1984), Nankervis and Savin (1987) and DeJong, Nankervis, Savin, and Whiteman (1992) in the context of AR(1) models. Burnside and Eichenbaum (1994) provide evidence on the poor performance of GMM-based Wald test statistics. For more general theoretical results on the inaccuracy of asymptotic methods, the reader may consult Dufour (1997); see also Nelson, Startz, and Zivot (1996), Savin and Würtz (1996) and Wang and Zivot (1996). Furthermore, there are situations where usual asymptotic procedures do not apply. For instance, consider a model for panel data with time dependent errors: if no assumption is made on the dependence structure, it is not at all clear what should be done.

The main characteristic of model (1) is that the vector of dependent variables $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_m')'$ is in some way divided into m subsamples (different individuals and/or different subperiods), whose relationship is unknown. Because the joint distribution of the vector of errors $\boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2', \ldots, \boldsymbol{u}_m')'$ is not specified, usual inference methods based on the whole sample $\boldsymbol{y}$ are not applicable. This paper develops inference procedures which are valid in such contexts.

The general issues we shall consider can be described as follows. Given several data sets whose stochastic relationship is not specified (or difficult to model), but on which we can make inferences separately, we study the following problems: (I) how to combine separate tests for an hypothesis of interest bearing on the different data sets (more precisely, how to test the *intersection* of several related hypotheses pertaining to different data sets); for example, in model (1), we may wish to test whether the linear restrictions $\boldsymbol{C}_i\boldsymbol{\beta}_i = \boldsymbol{\gamma}_{i0}$, $i = 1, 2, \ldots, m$, hold *jointly*; (II) how to test cross-restrictions between the separate models (such as $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_m$, when $k_i = k, i = 1, 2, \ldots, m$), which involves testing the *union* of a large (possibly infinite) number of hypotheses of the preceding type (*e.g.*, $\boldsymbol{\beta}_i = \boldsymbol{\beta}_0, i = 1, 2, \ldots, m$, for some $\boldsymbol{\beta}_0$); (III) how to combine confidence sets (*e.g.*, confidence intervals or confidence ellipsoids) for a common parameter of interest and based on different data sets in order to obtain more accurate confidence sets. All these problems require procedures for *pooling* information obtained from separate, possibly non independent, samples and for making comparisons between them.

Besides being applicable to situations where the stochastic relationship between the different samples is completely unknown, the methods proposed will also be useful for inference on various models in which the distribution of a standard statistic based on the complete sample is quite difficult to establish (*e.g.*, because of nuisance parameters), while the distributional properties of test statistics can be considerably simplified by looking at properly chosen subsamples. This is the case, for example, in seemingly unrelated regressions (SURE) and linear regressions with MA errors.

The methods proposed here rely on a systematic exploitation of Boole-Bonferroni inequalities [see Alt (1982)] which allow one to bound the probability of the union (or intersection) of a finite set of events from their marginal probabilities, without any knowledge of their dependence structure. Although such techniques have been used in the simultaneous inference literature to build simultaneous confidence intervals, especially in standard linear regressions [see Miller (1981) and Savin (1984)], it does not appear they have been exploited for the class of problems studied here. In particular, for general problems of type I, we discuss the use of *induced tests* based on rejecting the null hypothesis when at least one of the several separate hypotheses is rejected by one of several separate tests, with the overall level of the procedure being controlled by Boole-Bonferroni inequalities. For problems of type II, we propose using *empty intersection tests* which reject the null hypothesis when the intersection of a number of separate confidence sets (or intervals) is empty. In the case of confidence intervals, this leads to simple rules that reject the null hypothesis when the distance between two parameter estimates based on separate samples is greater than the *sum* of the corresponding critical points. We also discuss how one can perform empty intersection tests based on confidence ellipsoids and confidence boxes. For problems of type III, we propose using the intersection of several separate confidence sets as a way of *pooling* the information in the different samples to gain efficiency. These common characteristics have led us to use the terminology *union-intersection* (UI) methods.

The techniques discussed in this paper for type I problems are akin to procedures proposed for combining test statistics [see Folks (1984)] and for *meta-analysis* [see Hedges and Olkin (1985)]. Meta-analysis tries to combine the evidence reported in different studies and articles on particular scientific questions: it has often been used to synthesize medical studies. However, these literatures have concentrated on situations where the separate samples can be treated as independent and do not deal with econometric problems. Conversely, these methods are practically ignored in the econometric literature. Note also that the techniques we propose for problems of types II and III can be viewed as extensions of the “union-intersection” method proposed by Roy (1953) [see Arnold (1981, pp. 363-364)] for testing linear hypotheses in multivariate linear regressions, in the sense that an *infinity* of relatively simple hypothesis tests are explicitly considered and combined. A central difference here comes from the fact that the “simple” null hypotheses we consider are themselves tested via induced tests (because we study quite distinct setups) and from the different nature of the models studied.

As pointed above, our methods have the advantage of being versatile and straightforward to implement, even when important pieces of information are missing. These also turn out to be easily applicable in various problems where the distributional properties of test statistics can be considerably simplified by looking at appropriately selected subsamples. We show in particular that this is the case for several inference problems in SURE models and linear regressions with MA errors. This provides original and rather striking examples of “sample split techniques" for simplifying distributional properties. For other recent illustrations of this general idea, the reader may consult Angrist and Krueger (1994), Dufour and Jasiak (1995) and Staiger and Stock (1993). In the first reference, the authors propose a sample-split technique to obtain IV estimators with improved properties, while the two other papers suggest similar methods to obtain more reliable tests in structural models.

The paper is organized as follows. Section 2 presents the general theory: in the context of a general statistical model, we derive procedures for testing null hypotheses of types I and II. In Section 3, we consider the problem of pooling confidence sets obtained from different data sets (type III problems). In Section 4, we apply our results to test the equality of linear combinations of parameters of different equations in a SURE model, an interesting setup where standard tests and confidence sets only have an asymptotic justification [for a review, see Srivastava and Giles (1987)]. In particular, we impose no restrictions on the contemporaneous covariance matrix, allowing for different variances and instantaneous cross-correlation. In section 5, we study inference for linear regression models with MA(q) errors. We show that our inference technique is very well suited for testing hypotheses on regression coefficients in the presence of MA errors. We study in detail the case of an MA(1) process and consider the problem of testing an hypothesis about the mean. We compare our procedure with some alternative tests. It appears much easier to implement than other commonly used procedures, since it does not require estimation of MA parameters. We also study the performance of our method by simulation. The results show that sample-split combined test procedures are reliable from the point of view of level control and enjoy surprisingly good power properties. We conclude in Section 6.

2. Hypothesis testing: general theory

In this section, we consider a general statistical model characterized by a sample space $\mathcal{Y}$ and a family $\mathcal{L} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ of probability distributions parameterized by θ, where Θ is the set of admissible values for θ. Let $\mathcal{L}_0$ be a subset of $\mathcal{L}$ and suppose we wish to test $H_0 : \mathsf{P}_\theta \in \mathcal{L}_0$ against $H_1 : \mathsf{P}_\theta \in \mathcal{L} \setminus \mathcal{L}_0$. If the model is identified, which will be assumed, this amounts to testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\theta \in \Theta_0 \Leftrightarrow \mathsf{P}_\theta \in \mathcal{L}_0$.

We consider here three classes of inference problems concerning θ. First, we study situations where Θ_0 can expressed as a finite intersection of subsets of Θ, *i.e.,* $\Theta_0 = \bigcap_{\gamma \in \Gamma} \Theta_{0\gamma}$, where Γ is an index set of the form $\Gamma = \{1, 2, \ldots, r\}$, and $\Theta_{0,\gamma} \subset \Theta, \gamma \in \Gamma$. Second, we examine null hypotheses which restrict θ to a subset Θ_0 of Θ, where Θ_0 can be written as $\Theta_0 = \bigcup_{\gamma \in \Gamma} \Theta_0(\gamma), \Theta_0(\gamma) \subset \Theta, \gamma \in \Gamma$. In this case, Γ is not constrained to be a finite set. Thirdly, we consider situations where the information about θ is available from different subsamples whose joint distribution is unknown. We then try to pool these pieces of information by combining inferences based on each subsample.

2.1. H_0 as the finite intersection of subhypotheses

The test procedure we present in this section is based on the fact that, although H_0 may not be easily testable, it can be expressed as the intersection of subhypotheses, $H_{0\gamma} : \theta \in \Theta_{0\gamma}$, each one of which can be tested by usual procedures. The decision rule is built from the logical equivalence that H_0 is wrong if and only if *any* of its components $H_{0\gamma}$ is wrong.

Assume that we can test $H_{0\gamma}$ using a statistic T_γ such that, for any $\theta \in \Theta_{0\gamma}$, $\mathsf{P}_\theta(\{\boldsymbol{y} \in \mathcal{Y} : T_\gamma(\boldsymbol{y}) \geq x\})$ is known, for all $x \in \mathbb{R}$, $\gamma \in \Gamma = \{1, 2, \ldots, r\}$. The relation between these statistics is unknown or difficult to establish [as it is the case in model (1)]. We wish to combine the information on the true probability distribution of the model, brought by each of those r statistics. Since H_0 is true if and only if all the $H_{0\gamma}$'s are individually true, a natural way of testing H_0 is to proceed as follows. Using the r statistics T_γ, we build r critical regions $W_\gamma(\alpha_\gamma) = T_\gamma^{-1}([t_\gamma(\alpha_\gamma), : \infty))$, where $t_\gamma(\alpha_\gamma)$ is chosen so that $\mathsf{P}_\theta[W_\gamma(\alpha_\gamma)] = \alpha_\gamma$ under $H_{0\gamma}$. We reject the null hypothesis H_0 when the vector of observations $\boldsymbol{y}$ lies in at least one of the $W_\gamma(\alpha_\gamma)$ regions, or equivalently if $T_\gamma(\boldsymbol{y}) \geq t_\gamma(\alpha_\gamma)$ for at least one γ. The rejection region corresponding to this decision rule is $\bigcup_{\gamma \in \Gamma} W_\gamma(\alpha_\gamma)$. Such a test is called an *induced* test of H_0; see Savin (1984). Its size is impossible or difficult to determine since the joint distribution of the statistics T_γ is generally unknown or intractable. It is however possible to choose the individual levels α_γ so that the induced test has level $\alpha \in (0, 1)$, for by sub-additivity:

$$\mathsf{P}_\theta\left[\bigcup_{\gamma \in \Gamma} W_\gamma(\alpha_\gamma)\right] \leq \sum_{\gamma \in \Gamma} \mathsf{P}_\theta\left[W_\gamma(\alpha_\gamma)\right] = \sum_{\gamma \in \Gamma} \alpha_\gamma,$$

for any $\theta \in \bigcap_{\gamma \in \Gamma} \Theta_{0\gamma} = \Theta_0$. Therefore, if we want the induced test to have level α, we only need to choose the α_γ's so that they sum to α (or less).

To our knowledge, there is no criterion for choosing the α_γ's in an optimal manner. Without such a rule, in most of our applications we will give the null hypotheses $H_{0\gamma}$ the same degree of protection against an erroneous rejection by taking $\alpha_\gamma = \alpha_0 = \alpha/r, \forall \gamma \in \Gamma$. However, there may exist situations where we wish to weigh the $H_{0\gamma}$'s in a different way. In particular, if for some reason we know that one of the decisions $d_{\gamma'}$ (say, accepting or rejecting $H_{0\gamma'}$) is less reliable than the other decisions, we are naturally led to give $d_{\gamma'}$ less impact on the final decision concerning the acceptance or rejection of H_0. In other words, we will choose $\alpha_{\gamma'} < \alpha_\gamma, \forall \gamma \neq \gamma'$.

In the case where we choose $\alpha_\gamma = \alpha_0 = \alpha/r, \forall\ \gamma \in \Gamma$, we reject $H_{0\gamma}$ at level α_0 when $\boldsymbol{y}$ is in $W_\gamma(\alpha_0)$. Assuming $F_{\gamma,\theta}(x)$ is continuous in x, this region of $\mathcal{Y}$ can be re-expressed as

$$\begin{aligned} W_\gamma(\alpha_0) &= \{\boldsymbol{y} \in \mathcal{Y} : 1 - F_{\gamma,\theta}[T_\gamma(\boldsymbol{y})] \leq 1 - F_{\gamma,\theta}[t_\gamma(\alpha_0)]\} \\ &= \{\boldsymbol{y} \in \mathcal{Y} : 1 - F_{\gamma,\theta}[T_\gamma(\boldsymbol{y})] \leq \alpha_0\}, \end{aligned}$$

for all $\theta \in \Theta_0$, where $F_{\gamma,\theta}(z) \equiv \mathsf{P}_\theta\left[T_\gamma^{-1}((-\infty,\, z])\right]$ is the probability distribution function of $T_\gamma(\boldsymbol{y})$. Then an equivalent alternative rejection criterion is to reject $H_{0\gamma}$ when the p-value $\Lambda_\gamma(\boldsymbol{y}) \equiv 1 - F_{\gamma,\theta}[T_\gamma(\boldsymbol{y})]$ is less than α_0. So according to this rule, a rejection of H_0 occurs whenever at least one of the r p-values Λ_γ is not greater than α_γ, and the critical region of this test procedure is

$$W(\alpha) = \left\{\boldsymbol{y} \in \mathcal{Y} : \min_{\gamma \in \Gamma} \Lambda_\gamma(\boldsymbol{y}) \leq \alpha_0\right\}.$$

If we assume that the statistics T_γ are identically distributed under the null hypothesis, then $F_{\gamma,\theta} =$

$F_\theta, \ \forall\, \theta \in \Theta_0$ and $t_\gamma(\alpha_0) = t(\alpha_0), \ \forall\, \gamma \in \Gamma$, hence (with probability one)

$$W(\alpha) = \left\{ \boldsymbol{y} \in \mathcal{Y} : \max_{\gamma \in \Gamma} T_\gamma(\boldsymbol{y}) \geq t(\alpha_0) \right\}.$$

This criterion is derived heuristically from the logical equivalence that H_0 is true if and only if all the $H_{0\gamma}$'s are true. It is similar to Tippett's (1931) procedure for combining inferences obtained from independent studies: using the fact that the r p-values are i.i.d. $\mathcal{U}_{[0,1]}$ under H_0, Tippett (1931) suggested the rule:

$$\text{reject } H_0 \text{ at level } \alpha \text{ if } \min\{\Lambda_\gamma : \gamma \in \Gamma\} \ \leq \ 1 - (1-\alpha)^{1/r}.$$

Such procedures were also proposed for meta-analysis [see Hedges and Olkin (1985)], but have seldom been used in econometrics. We show below that an extension of Tippett's procedure to the case where the p-values are not independent, can be fruitfully applied to several econometric problems. The analogy with meta-analysis comes from the fact that inference on the i-th equation in (1) is made "independently" from inference on any other equation, although the test statistics may not be stochastically independent. Since dependence of the test statistics is a common situation in econometrics, we do not assume the independence of the p-values, which leads one to use α/r instead of $1-(1-\alpha)^{1/r}$. When α is small, the difference between α/r and $1-(1-\alpha)^{1/r}$ is typically quite small. For some applications of this approach to independent test statistics, see Dufour (1990), McCabe (1988) and Phillips and McCabe (1988, 1989).

It is possible to demonstrate optimality properties for induced tests. For example, consider a test procedure which combines r p-values $\Lambda_1, \Lambda_2, \ldots, \Lambda_r$, so that it rejects H_0 when $S(\Lambda_1, \Lambda_2, \ldots, \Lambda_r) \leq s$, where S is some function from $\mathbb{R}^r$ into $\mathbb{R}$ and s a critical value such that $\mathsf{P}\left[S(\Lambda_1, \Lambda_2, \ldots, \Lambda_r) \leq s\right] \leq \alpha$ under H_0. Birnbaum (1954) showed that every monotone combined test procedure is admissible in the class of all combined test procedures.[1]. A combined test procedure S is monotone if S is a nondecreasing function, $i.e.$ if $x_i^* \leq x_i, i = 1, 2, \ldots, r \Rightarrow S(x_1^*, x_2^*, \ldots, x_r^*) \leq S(x_1, x_2, \ldots, x_r)$. In our case, $S(\Lambda_1, \Lambda_2, \ldots, \Lambda_r) = \min\{\Lambda_1, \Lambda_2, \ldots, \Lambda_r\}$, is clearly nondecreasing. For further discussion of admissibility issues in such contexts, see Folks (1984).

2.2. H_0 as the union of subhypotheses

Let us consider a null hypothesis of the form $H_0 : \theta \in \Theta_0$, where $\Theta_0 = \bigcup_{\gamma \in \Gamma} \Theta_0(\gamma)$. The solution to this testing problem is very similar to the one suggested above in the case of an intersection of subhypotheses. Once again it is based on the fact that H_0 is wrong if and only if each of its components is wrong. If each hypothesis $H_0(\gamma) : \theta \in \Theta_0(\gamma)$ can be tested using the rejection region $W_\gamma(\alpha_\gamma) = T^{-1}\big([t_\gamma(\alpha_\gamma), \infty)\big)$, satisfying $\mathsf{P}_\theta\big[W_\gamma(\alpha_\gamma)\big] = \alpha_\gamma, \forall\ \theta \in \Theta_0(\gamma)$, it would appear natural to consider the overall rejection region $W(\alpha_\gamma, \gamma \in \Gamma) = \bigcap_{\gamma \in \Gamma} W_\gamma(\alpha_\gamma)$ for a test of H_0.

[1]On admissibility of decision rules, see Lehmann (1986, section 1.8, p.17).

However, difficult problems arise when one wants to implement this procedure as described above. First, if Γ contains a finite number p of elements, we have

$$\mathsf{P}_\theta\big[W(\alpha_\gamma,\gamma\in\Gamma)\big] \;\geq\; 1-\sum_{\gamma=1}^{p}\Big[1-\mathsf{P}_\theta\big[W_\gamma(\alpha_\gamma)\big]\Big]\,,$$

which provides a $lower$ bound for the probability of making a type I error. Of course, this type of bound is of no use since we try to bound from above the probability of an erroneous rejection of H_0. Appropriate upper bounds for the probability of an intersection are difficult to obtain. Second, when Γ is infinite, it is impossible to build $W_\gamma(\alpha_\gamma)$ for every $\gamma\in\Gamma$.

It is however interesting to note that some null hypotheses can be written as the union of several hypotheses (possibly an infinite number of such hypotheses). It is then natural to construct an overall rejection region which is equivalent to the infinite intersection $\bigcap_{\gamma\in\Gamma}W_\gamma(\alpha_\gamma)$. For example, consider the hypothesis $H_0:\theta_1=\theta_2=\cdots=\theta_m$, where θ_i is a $q\times 1$ subvector of the initial parameter vector θ. We note that H_0 is true if and only if $\exists\theta_0\in\mathbb{R}^q$ such that $\theta_1=\theta_2=\cdots=\theta_m=\theta_0$, where θ_0 is the unknown true value of θ_i under the null. Defining $\Theta_0(\theta_0)\equiv\{\theta\in\Theta:\theta_1=\theta_2=\cdots=\theta_m=\theta_0\}$, we have $\Theta_0=\bigcup_{\theta_0\in\mathbb{R}^q}\Theta_0(\theta_0)$. H_0 can be expressed as an infinite union of subhypotheses $H_0(\theta_0):\theta\in\Theta_0(\theta_0)$. Therefore H_0 is true if and only if anyone of the $H_0(\theta_0)$'s is true.

Obviously, it is impossible to test every $H_0(\theta_0)$. Instead, we propose the following procedure. For each $i\in\{1,2,\,\dots\,,m\}$, we build a confidence region $C_i(\boldsymbol{y}_i,\alpha_i)$ for θ_i with level $1-\alpha_i$ using the sample $\boldsymbol{y}_i$, where the α_i's are chosen so that $\sum_{i=1}^m\alpha_i=\alpha$. This region satisfies

$$\mathsf{P}_\theta\left[A_i(\theta_i,\alpha_i)\right]\;\geq\;1-\alpha_i,\qquad\forall\,\theta\in\Theta,$$

where $A_i(\theta_i,\alpha_i)=\{\boldsymbol{y}\in\mathcal{Y}:C_i(\boldsymbol{y}_i,\alpha_i)\ni\theta_i\},\;\; i=1,2,\,\dots\,,m$, and $G\ni x$ means that the set "G contains x". In particular, if θ_0 is the true value of θ_i, we have

$$\mathsf{P}_\theta\left[A_i(\theta_0,\alpha_i)\right]\geq 1-\alpha_i,\quad\forall\,\theta\in\Theta_0.$$

Proposition 1 *A (conservative) α-level test of $H_0:\theta_1=\theta_2=\,\dots\,=\theta_m$ is given by the rejection region*

$$W(\alpha,m)=\Big\{\boldsymbol{y}\in\mathcal{Y}:\bigcap_{i=1}^{m}C_i(\boldsymbol{y}_i,\alpha_i)=\emptyset\Big\}\,.$$

where $\alpha_i, i=1,2,\,\dots\,,m$ satisfy $\sum_{i=1}^m\alpha_i\leq\alpha$.

Proof: We need to show that $\mathsf{P}_\theta\left[W(\alpha,m)\right]\;\leq\;\alpha,\forall\,\theta\in\Theta_0,\forall\,\alpha\in(0,\,1)$. Note that $\forall\;\theta_0\in\mathbb{R}^q$,

$$\bigcap_{i=1}^{m}C_i(\boldsymbol{y}_i,\alpha_i)=\emptyset\quad\Rightarrow\quad\exists j\in\{1,2,\,\dots\,,m\}:C_j(\boldsymbol{y}_j,\alpha_j)\not\ni\theta_0,$$

hence, using the Boole-Bonferroni inequality,

$$\begin{aligned}\mathsf{P}_\theta\left[W(\alpha, m)\right] &\leq \mathsf{P}_\theta\left[\bigcup_{i=1}^{m}\{\boldsymbol{y}\in\mathcal{Y} : C_i(\boldsymbol{y}_i, \alpha_i) \not\ni \theta_0\}\right] \\ &\leq \sum_{i=1}^{m}\mathsf{P}_\theta\left[\mathcal{Y}\setminus A_i(\theta_0, \alpha_i)\right] \leq \sum_{i=1}^{m}\alpha_i \leq \alpha, \quad \forall\,\theta\in\Theta. \qquad \square\end{aligned}$$

We shall call a critical region of the form of $W(\alpha, m)$ an *empty intersection* test. In our notation, $W(\alpha, m)$ does not depend directly upon α, but on how the α_i's are chosen to satisfy the constraint $\sum_{i=1}^{m}\alpha_i \leq \alpha$. For this procedure to be applicable, we need to have confidence regions $C_i(\boldsymbol{y}_i, \alpha_i)$ with levels $1-\alpha_i$. This is of course possible in model (1) as long as $\Omega_i = \sigma_i^2\, I_{N_i}, i \in \{1, 2, \ldots, m\}$. We describe three interesting special cases for which the procedure takes a simple and appealing form.

2.2.1. Intersection of confidence intervals: the sum of critical points rule

Consider a situation where $q = 1$. Typically, $C_i(\boldsymbol{y}_i, \alpha_i)$ has the form

$$C_i(\boldsymbol{y}_i, \alpha_i) = \left[\hat{\theta}_i - c_{iL}(\boldsymbol{y}_i, \alpha_i)\,, \hat{\theta}_i + c_{iU}(\boldsymbol{y}_i, \alpha_i)\right],$$

where $\hat{\theta}_i$ is some estimator of θ_i, $\sum_{i=1}^{m}\alpha_i \leq \alpha \in (0, 1)$, and $c_{iL}(\boldsymbol{y}_i, \alpha_i) > 0,\ c_{iU}(\boldsymbol{y}_i, \alpha_i) > 0$ for all possible values of $\boldsymbol{y}_i$. Furthermore, it is usually the case that $c_{iL}(\boldsymbol{y}_i, \alpha_i) = c_{iU}(\boldsymbol{y}_i, \alpha_i)$ but we shall not need this restriction here. It is easy to see that the following lemma holds.

Lemma 2.1 *The intersection of a finite number* m *of intervals* $I_i \equiv [L_i\,,\,U_i] \subset \mathbb{R}, i = 1, 2, \ldots, m$, *with non empty interiors is empty if and only if*

$$\min\{U_i : i = 1, 2, \ldots, m\} < \max\{L_i : i = 1, 2, \ldots, m\}.$$

Proof: Define $U_M \equiv \min\{U_i : i = 1, 2, \ldots, m\}$, $L_M \equiv \max\{L_i : i = 1, 2, \ldots, m\}$ and $\tilde{I} \equiv \{x \in \mathbb{R} : L_M \leq x \leq U_M\}$. Then

$$\tilde{I} \neq \emptyset \Leftrightarrow \exists x \text{ such that } L_i \leq x \leq U_i \text{ for } i = 1, 2, \ldots, m \Leftrightarrow \exists x \text{ such that } L_M \leq x \leq U_M$$

or equivalently,

$$\tilde{I} = \emptyset \quad \Leftrightarrow \quad \forall x, x \notin [L_M, U_M] \quad \Leftrightarrow \quad U_M < L_M. \qquad \square$$

From Lemma **2.1** and Proposition 1, we reject H_0 if and only if

$$\min\{\hat{\theta}_i + c_{iU}(\boldsymbol{y}_i, \alpha_i) : i = 1, 2, \ldots, m\} < \max\{\hat{\theta}_i - c_{iL}(\boldsymbol{y}_i, \alpha_i) : i = 1, 2, \ldots, m\}.$$

But this condition is equivalent to

$$\exists j,\, k \in \{1, 2, \, \ldots \, , m\} \text{ such that } \hat{\theta}_j + c_{jU}(\boldsymbol{y}_j, \alpha_j) < \hat{\theta}_k - c_{kL}(\boldsymbol{y}_k, \alpha_k)$$

or

$$\exists j,\, k \in \{1, 2, \, \ldots \, , m\} \text{ such that } \frac{|\hat{\theta}_k - \hat{\theta}_j|}{c_{jU}(\boldsymbol{y}_j, \alpha_j) + c_{kL}(\boldsymbol{y}_k, \alpha_k)} > 1.$$

Finally, we reject H_0 if and only if

$$\max_{j,\, k \in \{1,2,\, \ldots \,,m\}} \left[\frac{|\hat{\theta}_k - \hat{\theta}_j|}{c_{jU}(\boldsymbol{y}_j, \alpha_j) + c_{kL}(\boldsymbol{y}_k, \alpha_k)} \right] > 1.$$

In the case where $m = 2$, with $c_{jU}(\boldsymbol{y}_j, \alpha_j) = c_{jL}(\boldsymbol{y}_j, \alpha_j) = c_j(\boldsymbol{y}_j, \alpha_j)$, $j = 1, 2$, this criterion takes the simple form: *reject the null hypothesis when the distance between the two estimates is larger than the sum of the two corresponding "critical points"*. The rejection region is then:

$$W(\alpha, 2) = \{\boldsymbol{y} \in \mathcal{Y} : |\hat{\theta}_1 - \hat{\theta}_2| > c_1(\boldsymbol{y}_1, \alpha_1) + c_2(\boldsymbol{y}_2, \alpha_2)\}.$$

For $m > 2$, we reject the null hypothesis when at least one of the distances $|\hat{\theta}_k - \hat{\theta}_j|$ is larger than the sum $c_j(\boldsymbol{y}_j, \alpha_j) + c_k(\boldsymbol{y}_k, \alpha_k)$. We will now extend this procedure to multidimensional parameters and consider confidence ellipsoids.

2.2.2. Intersection of two confidence ellipsoids

Consider the null hypothesis $H_0 : \theta_1 = \theta_2$, where θ_i is a $q \times 1$ vector. As before, H_0 can be restated as $H_0 : \theta \in \{\theta \in \Theta : \exists \theta_0 \in \mathbb{R}^q : \theta_1 = \theta_2 = \theta_0\}$. Suppose that for $i = 1, 2$, we have a confidence ellipsoid $C_i(\boldsymbol{y}_i, \alpha_i)$ for θ_i, such that

$$C_i(\boldsymbol{y}_i, \alpha_i) \ni \theta_0 \quad \Leftrightarrow \quad \big(\hat{\theta}_i - \theta_0\big)' A_i \big(\hat{\theta}_i - \theta_0\big) \;\leq\; c_i(\alpha_i),$$

where A_i is a $q \times q$ positive definite matrix whose elements depend on $\boldsymbol{y}_i$, $\hat{\theta}_i$ is an estimator of θ_i, and $c_i(\alpha_i)$ is a constant such that

$$\mathsf{P}_\theta \left[\{\boldsymbol{y} \in \mathcal{Y} : C_i(\boldsymbol{y}_i, \alpha_i) \ni \theta\} \right] \geq 1 - \alpha_i, \quad \forall\, \theta \in \Theta.$$

Then there exists two $q \times q$ matrices P_1 and P_2 such that $P_1{}' A_1 P_1 = I_q$, $P_2{}'(P_1{}' A_2 P_1) P_2 = D$, where D is a diagonal non-negative definite $q \times q$ matrix, $|P_1| \neq 0$ and $P_2 P_2{}' = I_q$. It is easy to show that

$$\big(\hat{\theta}_1 - \theta_0\big)' A_1 \big(\hat{\theta}_1 - \theta_0\big) \;\leq\; c_1(\alpha_1) \quad \Leftrightarrow \quad (\hat{\gamma}_1 - \gamma)' \, (\hat{\gamma}_1 - \gamma) \;\leq\; c_1(\alpha_1),$$

$$\big(\hat{\theta}_2 - \theta_0\big)' A_2 \big(\hat{\theta}_2 - \theta_0\big) \;\leq\; c_2(\alpha_2) \quad \Leftrightarrow \quad (\hat{\gamma}_2 - \gamma)' \, D \, (\hat{\gamma}_2 - \gamma) \;\leq\; c_2(\alpha_2)$$

where $\gamma = {P_2}' {P_1}^{-1} \theta_0$ and $\hat{\gamma}_i = {P_2}' {P_1}^{-1} \hat{\theta}_i$, $i = 1, 2$. Setting

$$\begin{aligned} E_1(\alpha_1) &= \{\gamma \in \mathbb{R}^q : (\hat{\gamma}_1 - \gamma)' (\hat{\gamma}_1 - \gamma) \le c_1(\alpha_1)\}, \\ E_2(\alpha_2) &= \{\gamma \in \mathbb{R}^q : (\hat{\gamma}_2 - \gamma)' D (\hat{\gamma}_2 - \gamma) \le c_2(\alpha_2)\}, \end{aligned}$$

the rejection criterion $C_1(\boldsymbol{y}_1, \alpha_1) \bigcap C_2(\boldsymbol{y}_2, \alpha_2) = \emptyset$ of Proposition 1 is seen to be equivalent to $E_1(\alpha_1) \bigcap E_2(\alpha_2) = \emptyset$.

To determine whether the intersection of the two ellipsoids is empty, it is sufficient to find the set $E_2^*(\alpha_2)$ of solutions of the problem

$$\min_{\gamma \in E_2(\alpha_2)} \|\gamma - \hat{\gamma}_1\|^2 \tag{2}$$

and check whether there is at least one element of $E_2^*(\alpha_2)$ lying in $E_1(\alpha_1)$, in which case the two confidence ellipsoids have a non empty intersection and H_0 is accepted at level $\alpha_1 + \alpha_2$. This is justified by the following lemma.

Lemma 2.2 *Let $E_2^*(\alpha_2) \subset E_2(\alpha_2)$ be the set of the solutions of* (2). *Then*

$$E_1(\alpha_1) \cap E_2(\alpha_2) \neq \emptyset \quad \Longleftrightarrow \quad E_1(\alpha_1) \cap E_2^*(\alpha_2) \neq \emptyset.$$

Proof: ($\Leftarrow$) Let $E_1(\alpha_1) \cap E_2^*(\alpha_2) \neq \emptyset$. Since $E_2^*(\alpha_2) \subseteq E_2(\alpha_2)$, it follows trivially that $E_1(\alpha_1) \cap E_2(\alpha_2) \neq \emptyset$.
($\Rightarrow$) Let $E_1(\alpha_1) \cap E_2(\alpha_2) \neq \emptyset$. Then we can find $\tilde{\gamma}$ such that $\tilde{\gamma} \in E_1(\alpha_1)$ and $\tilde{\gamma} \in E_2(\alpha_2)$. In other words, $\tilde{\gamma}$ is an element of $E_2(\alpha_2)$ that satisfies the condition $\|\tilde{\gamma} - \hat{\gamma}_1\|^2 \le c_1(\alpha_1)$, which entails that $\min_{\gamma \in E_2(\alpha_2)} \|\gamma - \hat{\gamma}_1\|^2 \le c_1(\alpha_1)$. Now suppose $E_1(\alpha_1) \bigcap E_2^*(\alpha_2) = \emptyset$. This means that the following implication must hold:

$$\begin{aligned} \|\gamma_0 - \hat{\gamma}_1\|^2 \equiv \min_{\gamma \in E_2(\alpha_2)} \|\gamma - \hat{\gamma}_1\|^2 \quad &\Rightarrow \quad \gamma_0 \notin E_1(\alpha_1) \\ &\Rightarrow \quad \|\gamma_0 - \hat{\gamma}_1\|^2 > c_1(\alpha_1). \end{aligned}$$

Since $E_2^*(\alpha_2)$ is not empty, it follows that $\min_{\gamma \in E_2(\alpha_2)} \|\gamma - \hat{\gamma}_1\|^2 > c_1(\alpha_1)$, a contradiction. Thus we must have $E_1(\alpha_1) \bigcap E_2^*(\alpha_2) \neq \emptyset$. $\square$

Although any numerical calculus computer package is able to solve (**2.2**), we propose a simple two step procedure for deciding whether the intersection of two ellipsoids is empty. This method can prove useful for high dimensional problems. The two steps are the following:

(1) check whether $\hat{\theta}_1 \in C_2(\boldsymbol{y}_2, \alpha_2)$ or $\hat{\theta}_2 \in C_1(\boldsymbol{y}_1, \alpha_1)$. If one of these events is realized, then $C_1(\boldsymbol{y}_1, \alpha_1) \bigcap C_2(\boldsymbol{y}_2, \alpha_2)$ is not empty, and H_0 is accepted. Otherwise, go to the second stage of the procedure;

(2) since $\hat{\theta}_1 \notin C_2(\boldsymbol{y}_2, \alpha_2)$, it follows (by convexity) that $E_2^*(\alpha_2)$ is a subset of the boundary

$\partial E_2(\alpha_2)$ and

$$\min_{\gamma \in \partial E_2(\alpha_2)} \|\gamma - \hat{\gamma}_1\|^2 = \min_{\gamma \in E_2(\alpha_2)} \|\gamma - \hat{\gamma}_1\|^2 > 0, \tag{3}$$

so that we can check whether $E_1(\alpha_1) \bigcap E_2(\alpha_2) \neq \emptyset$ by checking if $E_1(\alpha_1) \bigcap \partial E_2(\alpha_2) \neq \emptyset$. If the latter condition is satisfied, H_0 is accepted; otherwise H_0 is rejected.

To be more specific, step (2) above simply requires one to find the vector $\tilde{\gamma}$ which minimizes $\|\gamma - \hat{\gamma}_1\|^2$ subject to the restriction $(\gamma - \hat{\gamma}_2)' D (\gamma - \hat{\gamma}_2) = c_2(\alpha_2)$, and then to reject H_0 when $\|\tilde{\gamma} - \hat{\gamma}_1\|^2 > c_1(\alpha_1)$.

2.2.3. Intersection of two confidence boxes

When the covariance matrices of the estimators $\hat{\theta}_i$ are unknown, one cannot typically build confidence ellipsoids. To illustrate such situations, consider the following example. Two published papers investigate econometric relationships of the form

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{u}_i\,, \quad i = 1, 2, \tag{4}$$

where $\boldsymbol{u}_i \sim N(0, \sigma_i^2 I_{N_i})$, and $\boldsymbol{\beta}_i$ is a $k \times 1$ vector of unknown parameters.

We wish to compare the two parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. However, only the standard errors of the coefficients are known (or reported), not their covariance matrices. Then it is not possible to use the previous procedure. But it is possible to use simultaneous inference techniques and build simultaneous confidence boxes (hyperrectangles). Various methods for building such confidence sets are described in Miller (1981) and Savin (1984). More precisely, let us build for each of the two regressions in (4) k simultaneous confidence intervals, denoted by $C_i^j(\boldsymbol{y}_i, \alpha_i^j)$ for the component β_i^j of $\boldsymbol{\beta}_i,\ j = 1, 2, \ldots, k,\ i = 1, 2$, such that

$$\mathsf{P}_\theta \left[\bigcap_{j=1}^{k} \{ \boldsymbol{y} \in \mathcal{Y} : C_i^j(\boldsymbol{y}_i, \alpha_i^j) \ni \beta_i^j \} \right] \geq 1 - \alpha_i, \quad \forall\, \theta \in \Theta, \quad i = 1, 2.$$

Then choosing the α_i's so that $\alpha_1 + \alpha_2 = \alpha$, and applying the result of Proposition 1, we reject $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ at level α when the intersection of the two hyperrectangles is empty.

Checking whether the intersection of the two boxes is empty is especially simple because one simply needs to see whether the confidence intervals for each component of $\boldsymbol{\beta}_i$ have an empty intersection (as in Section 2.2.1). Furthermore it is straightforward to extend this technique in order to compare more than two regressions. Similarly, although we proposed a test for the null hypothesis that all the parameter vectors $\boldsymbol{\beta}_i$ are equal (then imposing that in each equation has the same number of parameters), it is easy to extend this procedure in order to test the equality of *linear* transformations of $\boldsymbol{\beta}_i, i = 1, 2, \ldots, m$. Indeed, the method relies only on the ability to derive confidence regions for parameters which are restricted to be equal under the null. This is clearly possible whenever the paramaters of interest are of the form $\boldsymbol{R}_i \boldsymbol{\beta}_i$. The procedure is actually applicable to any function $h(\theta)$ of the parameter, provided we are able to build a confidence region for $h(\theta)$.

3. Confidence set estimation

In the previous section, we described a general method for testing hypotheses in several contexts. The main feature of the procedure is that a single final decision concerning a family of probability distributions is taken by combining several individual (partial) decisions on that family.

In many situations, we may wish to go a step further. For instance, consider again model (1) and the null hypothesis $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_m$. The results of Section 2 show how to test such an hypothesis. Suppose H_0 is taken for granted. It is then natural to ask what could be a valid confidence region for $\boldsymbol{\beta}$, the unknown common value of $\boldsymbol{\beta}_i, i = 1, 2, \ldots, m$. The main difficulty here comes from the fact that only the marginal distributions of the separate samples are specified, not their joint distribution. Suppose each of these marginal distributions can be used to build a confidence region for $\boldsymbol{\beta}$. The problem is then to find a way of pooling these pieces of information on the true value of $\boldsymbol{\beta}$ and derive a single confidence region which is based on the whole sample. This can be done as follows. Suppose each one of the separate observation vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m$ has a distribution which depends on $\theta \in \Theta$. Although the joint distribution of $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_m')'$ is unknown, we assume it is possible to build m separate confidence regions $C_i(\alpha_i, \boldsymbol{y}_i)$ for θ such that

$$\mathsf{P}_\theta \left[C_i(\alpha_i, \boldsymbol{y}_i) \ni \theta\right] \geq 1 - \alpha_i, \quad \forall\, \theta \in \Theta, \quad i = 1, 2, \ldots, m.$$

Then a natural way to exploit simultaneously these different pieces of information consists of taking the intersection $\bigcap_{i=1}^m C_i(\alpha_i, \boldsymbol{y}_i)$ of the different confidence regions. It is easy to see that

$$\mathsf{P}_\theta \left[\bigcap_{i=1}^m C_i(\alpha_i, \boldsymbol{y}_i) \ni \theta\right] \geq 1 - \sum_{i=1}^m \mathsf{P}_\theta \left[C_i(\alpha_i, \boldsymbol{y}_i) \not\ni \theta\right] \geq 1 - \sum_{i=1}^m \alpha_i\,, \quad \forall : \theta \in \Theta.$$

Thus selecting $\alpha_1, \alpha_2, \ldots, \alpha_m$ so that $\sum_{i=1}^m \alpha_i = \alpha$, we can get any desired confidence level. This procedure can be especially useful when one of the subsample yields a particularily accurate confidence set for θ.

In the next sections, we show how the procedures described in Sections 2 and 3 can be used to make inference in SURE and regressions with MA errors.

4. Exact inference in SURE models

4.1. The model and the procedures

In this section we consider the SURE-type model:

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{u}_i\,, \\ \boldsymbol{u}_i &\sim N(0\,, \sigma_{ii} I_{N_i}), \quad i = 1, 2, \ldots, m, \end{aligned} \tag{5}$$

where $\boldsymbol{X}_i$ is a $N_i \times k_i$ fixed matrix of rank $k_i < N_i$, $\boldsymbol{\beta}_i$ is a $k_i \times 1$ vector of unknown parameters, $\boldsymbol{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iN_i})'$, and $\mathsf{E}(u_{it} u_{is}) = 0, \forall t \neq s$. Note we do not impose any restriction on the relationship between the m equations, so that the above model is more general than the standard SURE model. The null hypotheses of interest are $H_0^{(1)}$: $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{0i}, i = 1, 2, \ldots, m$, and $H_0^{(2)}$: $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = \ldots = \boldsymbol{\lambda}_m$, where $\boldsymbol{\lambda}_i = \boldsymbol{R}_i\boldsymbol{\beta}_i$, $\boldsymbol{R}_i$ is a known $q_i \times k_i$ matrix with rank $q_i \leq k_i, i = 1, 2, \ldots, m$, and $\boldsymbol{\lambda}_{0i}$ is a known $q_i \times 1$ vector, $i = 1, 2, \ldots, m$.[2] An interesting special case of $H_0^{(1)}$ is $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_m = \boldsymbol{\beta}_0$, which is obtained by choosing $k_i = k$, $\boldsymbol{R}_i = I_k$, $\boldsymbol{\lambda}_{0i} = \boldsymbol{\beta}_0$, a known $k \times 1$ vector, in the above setup.

We will consider two versions of (5), depending on whether we make the assumption $A1 :: \boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2' \ldots, \boldsymbol{u}_m')' \sim N(0, \sigma^2 I_N)$, where $N = \sum_{i=1}^m N_i$. Under $A1$, there exists an optimal test of $H_0^{(1)}$ given by the critical region associated to the Fisher F statistic, based on the stacked model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, where $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2' \ldots, \boldsymbol{y}_m')', \boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_m')'$, $\boldsymbol{X} = \text{diag}(\boldsymbol{X}_i)_{i=1,2,\ldots,m}$, and

$$F = \frac{\left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0\right)' \left[s^2 \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1} \left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0\right)}{Q},$$

with $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2', \ldots, \boldsymbol{\lambda}_m')'$, $\hat{\boldsymbol{\lambda}} = \left(\hat{\boldsymbol{\lambda}}_1', \hat{\boldsymbol{\lambda}}_2', \ldots, \hat{\boldsymbol{\lambda}}_m'\right)'$, $\boldsymbol{\lambda}_0 = (\boldsymbol{\lambda}_{01}', \boldsymbol{\lambda}_{02}', \ldots, \boldsymbol{\lambda}_{0m}')'$, $\hat{\boldsymbol{\lambda}}_i = \boldsymbol{R}_i\hat{\boldsymbol{\beta}}_i$, $\hat{\boldsymbol{\lambda}}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_i, i = 1, 2, \ldots, m$, $\boldsymbol{R} = \text{diag}(\boldsymbol{R}_i)_{i=1,2,\ldots,m}$, $s^2 = \|(I_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}\|^2/(N-K), Q = \sum_{i=1}^m q_i$ and $K = \sum_{i=1}^m k_i$.

When we introduce heteroskedasticity in the model by allowing the variances to differ across equations, our procedure is still valid, but the Fisher procedure is not. As an alternative, one would typically use an asymptotic method based on a generalized least squares estimation and a critical region defined by a Wald, a Lagrange multiplier or a likelihood ratio statistic. But, as we already mentioned in the introduction, it is well known that these approximations are not reliable.

An induced test of $H_0^{(1)}$ consists in testing $H_{0i}^{(1)}$: $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{0i}$ at level α_i using the critical region $W_i(\alpha_i) = \{\boldsymbol{y} \in \mathcal{Y} : F_i > F(\alpha_i; q_i, N_i - k_i)\}$, where

$$F_i = \frac{(\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_{0i})'\left[s_i^2\boldsymbol{R}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{R}_i'\right]^{-1}(\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_{0i})}{q_i},$$

with $\hat{\boldsymbol{\lambda}}_i = \boldsymbol{R}_i\hat{\boldsymbol{\beta}}_i$,: $s_i^2 = \|(I_{N_i} - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i')\boldsymbol{y}_i\|^2/(N_i - k_i)$ and $F(\alpha_i; q_i, N_i - k_i)$ is the $1 - \alpha_i$ percentile of the Fisher distribution with $(q_i, N_i - k_i)$ degrees of freedom. The α_i's can be chosen so that $\sum_{i=1}^m \alpha_i = \alpha$. Then the level α critical region for an induced test of $H_0^{(1)}$ is $\bigcup_{i=1}^m W_i(\alpha_i)$.

If we wish to test $H_0^{(2)}$ at level α, we simply have to build m confidence regions at level $1 - \alpha_i$ for $\boldsymbol{\lambda}_i$ which are defined by

$$C_i(\boldsymbol{y}_i, \alpha_i) = \{\boldsymbol{x} \in \mathbb{R}^q : \left(\hat{\boldsymbol{\lambda}}_i - \boldsymbol{x}\right)'\left[s_i^2\boldsymbol{R}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{R}_i'\right]^{-1}\left(\hat{\boldsymbol{\lambda}}_i - \boldsymbol{x}\right) \leq q_i F(\alpha_i; q_i, N_i - k_i)\}$$

in the $\boldsymbol{\lambda}_i$ space, and reject $H_0^{(2)}$ whenever $\bigcap_{i=1}^m C_i(\boldsymbol{y}_i, \alpha_i) = \emptyset$, with $\sum_{i=1}^m \alpha_i = \alpha$.

Note that, under assumption $A1$, the induced procedure for a test of $H_0^{(1)}$can be improved by taking into account the independence of the regressions. In Section 2, we showed that the rejection region associated with an induced test of $H_0^{(1)}$ is $\bigcup_{i=1}^m W_i(\alpha_i)$, where $W_i(\alpha_i)$ is the critical region

[2] For $H_0^{(2)}$ we must have $q = q_i, \forall i = 1, 2, \ldots, m$, and $q \leq \min\{k_i : i = 1, 2, \ldots, m\}$.

for a test of $\boldsymbol{\beta}_i = \boldsymbol{\beta}_0$ at level α_i. Under $A1$, we have:

$$\mathsf{P}_\theta\left[\bigcup_{i=1}^{m} W_i(\alpha_i)\right] = 1 - \mathsf{P}_\theta\left[\bigcap_{i=1}^{m} \mathcal{Y} \setminus W_i(\alpha_i)\right] = 1 - \prod_{i=1}^{m} \mathsf{P}_\theta\left[\mathcal{Y} \setminus W_i(\alpha_i)\right].$$

Under $H_0^{(1)}$ we have $\mathsf{P}_\theta\left[\mathcal{Y} \setminus W_i(\alpha_i)\right] = 1-\alpha_i$, Thus by choosing the α_i's so that $\prod_{i=1}^{m}(1-\alpha_i) = \alpha$, we get a test of $H_0^{(1)}$ which has level α. If $\alpha_i = \alpha_0, i = 1, 2, \ldots, m$, we must have $\alpha_0 = 1 - (1 - \alpha)^{1/m}$.

Unfortunately, the independence assumption $A1$ is not helpful when we turn to the test of $H_0^{(2)}$. But the procedure of Section 2.2 remains valid. To see why it is difficult to exploit the independence of the regressions, consider the case where $m = 2$ and $k = 1$. We can build two confidence intervals $C_i(\boldsymbol{y}_i, \alpha_0) = \left[\hat{\beta}_i - c_i(\boldsymbol{y}_i, \alpha_0),\ \hat{\beta}_i + c_i(\boldsymbol{y}_i, \alpha_0)\right]$, with $\alpha_0 = \alpha/2$. According to our rejection criterion (see Section 2.2.1), we reject $H_0^{(2)}$ at level α when $|\hat{\beta}_1 - \hat{\beta}_2| > c_1(\boldsymbol{y}_1, \alpha_0) + c_2(\boldsymbol{y}_2, \alpha_0)$. It is quite difficult to find the size of this critical region.

Consider now model (5) where assumption $A1$ is not imposed. This model has $\frac{m(m+1)}{2} + \sum_{i=1}^{m} k_i$ parameters and $\sum_{i=1}^{m} N_i$ observations. In this case, no usual finite sample or asymptotic test procedure appears to be available for comparing the coefficients of the different regressions. But the induced test method allows one to test $H_0^{(1)}$ and $H_0^{(2)}$ relatively easily.

4.2. Some examples

We now present some illustrations of the procedure described in the previous sections.

4.2.1. Testing restrictions in a system of demands for inputs

The first example we consider is taken from Berndt (1991, p. 460-462). We consider the problem of testing restrictions on the parameters of a generalized Leontieff cost function. We assume that the production technology has constant returns to scale and incorporates only two inputs, capital (K) and labor (L), whose prices are P_K and P_L respectively. If we denote the output by Y and the total cost by C, the generalized Leontieff cost function is

$$C = Y \cdot (d_{KK}P_K + 2d_{KL}(P_K P_L)^{1/2} + d_{LL}P_L).$$

If the producer has a cost minimizing strategy, it can be shown that the demands for factors K and L are given by

$$K/Y = d_{KK} + d_{KL}(P_L/P_K)^{1/2}, \quad L/Y = d_{LL} + d_{KL}(P_K/P_L)^{1/2}.$$

A stochastic version of this model would consist in the two equation SURE model

$$k_t = a_k + b_k p_t^k + u_t^k, \quad l_t = a_l + b_l p_t^l + u_t^l,$$

where $\boldsymbol{u}^k$ and $\boldsymbol{u}^l$ are two Gaussian random vectors with zero mean and covariance matrices $\sigma_k^2 I_N$ and $\sigma_l^2 I_N$, respectively. $N = 25$ is the sample size for each variable of the model. A restriction imposed by the theory is $b_k = b_l$, which will be our null hypothesis. To test H_0, the procedure described in Section 2.2.1 is particularly well suited since we have no *a priori* information on the relation between the random variables u_t^k and u_s^l. Using the data provided in Berndt (1991), which are described in Berndt and Wood (1975), we performed separate tests of the following null hypotheses:

$$H_0 : b_k = b_l \qquad H_0^* : \begin{pmatrix} a_k \\ b_k \end{pmatrix} = \begin{pmatrix} a_l \\ b_l \end{pmatrix}$$

The results of the estimation are:

$$k_t = \underset{(.000125)}{0.0490} + \underset{(.000084)}{0.00342}\, p_t^k + \hat{u}_t^k\,, \qquad l_t = \underset{(.001621)}{-0.04464} + \underset{(.002350)}{0.28295}\, p_t^l \; + \; \hat{u}_t^l\,,$$

where the standard errors are given in parenthesis. In Figure 1, we show the two 97.5% level confidence ellipsoids required for testing H_0^*. It is straightforward to see that we can reject both null hypotheses at level 5% because none of the regions intersect. Similarly, the 97.5% level confidence intervals for b_k and b_l are respectively $(-0.01869, 0.02539)$ and $(0.1659, 0.3992)$, and so do not intersect.

Since no information on the joint distribution of u_t^k and u_t^l is available, usual GLS procedures cannot be applied in this context. However, suppose that we assume that $(u_1^k, u_2^k, \dots, u_{25}^k, u_1^l, u_2^l, \dots, u_{25}^l)'$ is a Gaussian random vector with variance matrix

$$\Omega = \begin{pmatrix} \sigma_1^2 I_{25} & \sigma_{12} I_{25} \\ \sigma_{12} I_{25} & \sigma_2^2 I_{25} \end{pmatrix},$$

as is usually done in SURE models. Using standard GLS techniques, the estimate of $(a_k, b_k, a_l, b_l)'$ is $(0.05100, 0.00235, -0.04886, 0.28804)'$ and the F statistics for testing H_0 and H_0^* at level are 27.61 and 938.37, respectively. Since the corresponding 5% asymptotic critical values are 4.05175 and 3.19958, the null hypotheses are both rejected. However, one may prefer the empty intersection test procedure, because it makes a weaker assumption on the error distribution. Moreover, GLS-based tests only have an asymptotic justification.

4.2.2. Testing restrictions on returns to schooling

This example is taken from Ashenfelter and Zimmerman (1993). The study considers the following SURE model:

$$Y_{1j} = \theta_1 X_{1j} + \lambda_2 X_{2j} + w_{1j} \qquad Y_{2j} = \lambda_1 X_{1j} + \theta_2 X_{2j} + w_{2j}\,,$$

where Y_{ij} and X_{ij} represent the log wage and the schooling of the i-th brother in the j-th family. These equations are the reduced form of a structural model which expresses the relationship between

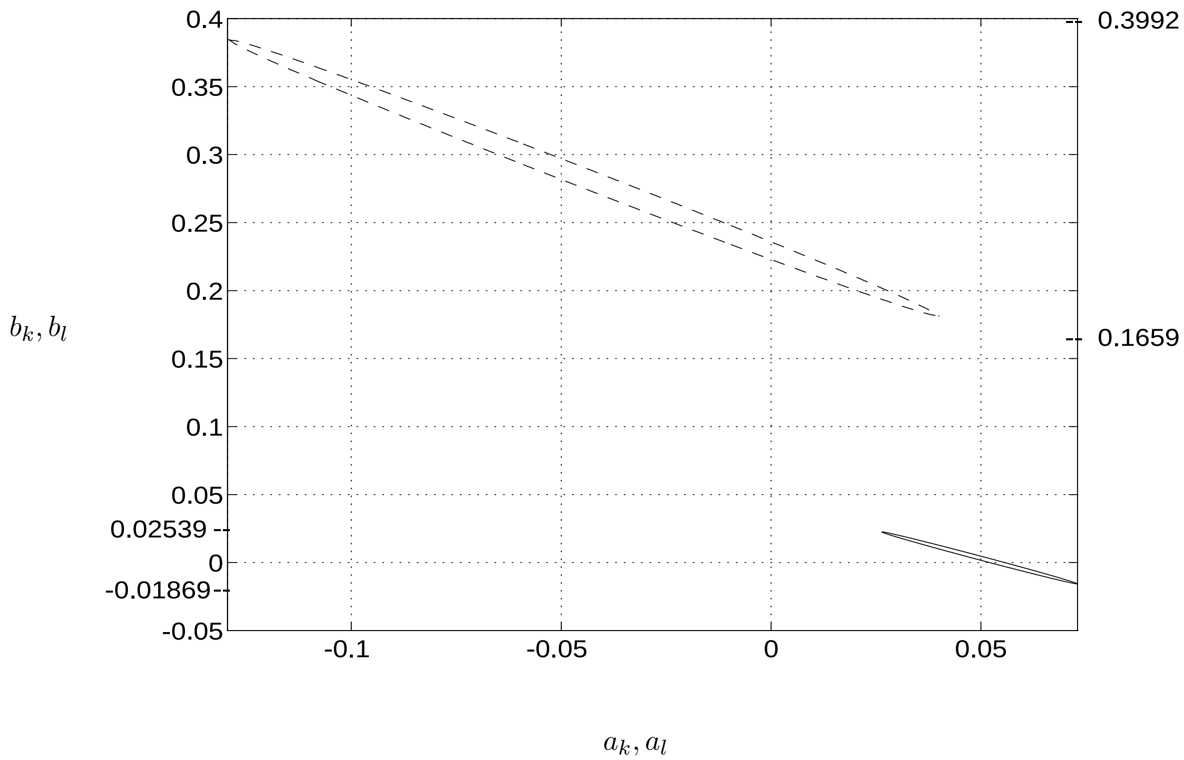

FIGURE 1. 97.5% confidence ellipsoids and intervals in the Berndt example. Confidence ellipsoid for $(a_k, b_k)'$: — ; confidence ellipsoid for $(a_l, b_l)'$: $- - -$; the confidence intervals for b_k and b_l appear on the left and right vertical axes respectively.

the wage and years of schooling:

$$Y_{1j} = \beta_1 X_{1j} + F_j + v_{1j} , \qquad Y_{2j} = \beta_2 X_{2j} + F_j + v_{2j} ,$$

$$F_j = \lambda_1 X_{1j} + \lambda_2 X_{2j} + \xi_j ,$$

where F is a family specific component. We must have $\theta_i = \beta_i + \lambda_i, i = 1, 2$.

The structural model has been estimated over a sample of 143 pairs of brothers. The estimates reported by Ashenfelter and Zimmerman (1993, Table 3) are given below, with standard errors in parentheses:

$$\underset{(0.015)}{\hat{\theta}_1 = 0.052} , \quad \underset{(0.020)}{\hat{\lambda}_1 = 0.018} , \quad \underset{(0.019)}{\hat{\theta}_2 = 0.068} , \quad \underset{(0.015)}{\hat{\lambda}_2 = 0.006} .$$

A natural hypothesis to test here is $H_0 : (\beta_1, \lambda_1)' = (\beta_2, \lambda_2)'$. This can easily be tested from the estimated structural model, since H_0 is equivalent to $H_0^* : (\theta_1, \lambda_1)' = (\theta_2, \lambda_2)'$. Here, we will use the hyperrectangle technique, because Ashenfelter and Zimmerman (1993) do not provide the full estimated covariance matrix for each regression. We first find a confidence interval with level $1-(\alpha/4)$ for each one of the mean parameters in the structural model, and check whether the two rectangles so obtained overlap, in which case we accept the null hypothesis. This is done for $\alpha = 5\%$. Each one of the following events $[0.0140, 0.0900] \ni \theta_1$, $[-0.0326, 0.0686] \ni \lambda_1$, $[0.0199, 0.1161] \ni \theta_2$, $[-0.0320, 0.0440] \ni \lambda_2$ occurs with probability 0.9875. We accept the null hypothesis at level 5%, since the two boxes $[0.0140, 0.0900] \times [-0.0326, 0.0686]$ and $[0.0199, 0.1161] \times [-0.0320, 0.0440]$ have a non-empty intersection, which is $[0.0199, 0.0900] \times [-0.0320, 0.0440]$.

5. Exact inference in linear regression models with MA(q) errors

In this section, we show that the procedures developed in Section 2 can be useful for inference in some dynamic models.

5.1. A test on the mean of a general MA(q) model

In this section, we consider models of the form

$$\begin{aligned} & y_t = m_t + u_t\,, \quad u_t = \Psi(B)\varepsilon_t\,, \quad t \in \mathrm{T} = \{1, 2, \ldots, T\}, \\ & \boldsymbol{\varepsilon} \equiv (\varepsilon_{1-q}, \varepsilon_{2-q}, \ldots, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_T)' \sim N\big[0, \sigma^2 \boldsymbol{I}_{T+q}\big]\,, \end{aligned} \tag{6}$$

where $\Psi(z) = \psi_0 + \psi_1 z + \psi_2 z^2 + \cdots + \psi_q z^q$, $\psi_0 \equiv 1$, $m_t = \sum_{k=1}^{K} x_{tk} b_k = \boldsymbol{x}_t' \boldsymbol{b}$, $\boldsymbol{b} \equiv (b_1, b_2, \ldots, b_K)'$ is a vector of unknown coefficients and $\boldsymbol{x}_t \equiv (x_{t1}, x_{t2}, \ldots, x_{tK})'$, $t = 1, 2, \ldots, T$, are vectors of fixed (or stricly exogenous) variables. In model (6), $\boldsymbol{y} \sim N(\boldsymbol{m}, \Omega)$, where $\boldsymbol{m} = (\mathsf{E}\, m_1, \mathsf{E}\, m_2, \ldots, \mathsf{E}\, m_T)'$ and $\Omega = (\omega_{t,s})_{t,s=1,2,\ldots,T}$, with

$$\omega_{t,s} = \begin{cases} \sigma^2 \sum_{i=|t-s|}^{q} \psi_i \psi_{i-|t-s|}\,, & \text{if} \quad |t-s| \le q, \\ 0, \quad \text{if} \quad |t-s| > q. & \end{cases} \tag{7}$$

(7) shows the key feature of model (6): observations distant by more than q periods from each other are mutually independent. Then, we are naturally led to consider model (6) for subsamples obtained as follows. Define subsets of T, $J_i \equiv \{i, i+(q+1), i+2(q+1), \ldots, i+n_i(q+1)\}$, where $n_i \equiv I\left[(T-i)/(q+1)\right]$ ($I[x]$ denotes the integer part of x), $i = 1, 2, \ldots, q+1$, and consider the $q+1$ equations

$$y_t = m_t + u_t\,, \quad t \in J_i\,, \quad \boldsymbol{u} \sim N\big[0, \sigma_u^2 \boldsymbol{I}_{n_i+1}\big]\,, \quad i = 1, 2, \ldots, q+1. \tag{8}$$

(8) belongs to the class of model (1). In each equation, the error term satisfies the assumptions of the linear regression model, so that it is possible to apply usual inference procedures to test restrictions on $\boldsymbol{b}$, $H_0 : \boldsymbol{b} \in \Phi$. This null hypothesis can be seen as the intersection of $q+1$ hypotheses $H_{0,i}$, each of which restricts the mean of the i-th subsample to be in Φ, $i = 1, 2, \ldots, q+1$. The

methods presented in Sections 2 and 3 are perfectly suited to such situations. We build $q+1$ critical regions with level $\alpha/(q+1)$ to test each one of the hypotheses $H_{0,i}$, and reject the null hypothesis at level α if the vector of observations belongs to the union of these regions. Note we did not make any assumption on the roots of $\Psi(z)$. In particular, we did not restrict the MA process $\{\Psi(B)\varepsilon_t : t \in \mathrm{T}\}$ to be invertible.

In the next subsection we apply the procedure to a MA(1) process with a constant and provide comparisons with some alternative procedures such as asymptotic tests and bounds tests.

5.2. Exact inference in the context of a MA(1) process

5.2.1. An induced test on the mean

Consider the model described by (6), with $q = 1$, $K = 1$ and $x_t = 1, \ \forall\, t \in \mathrm{T}$:

$$y_t = \beta + \varepsilon_t + \psi\varepsilon_{t-1}\,, \quad \varepsilon_t \overset{ind.}{\sim} N(0, \sigma^2), \quad t \in \mathrm{T}. \tag{9}$$

The vector of parameters is $\theta = (\beta, \psi, \sigma^2)'$. The null hypothesis we consider is $H_0 : \theta \in \Theta_0$, $\Theta_0 = \{\boldsymbol{\theta} \in \Theta : \beta = 0\}$. According to our procedure, assuming T is even, we form two subsamples of size $T/2$, $(y_t, t \in J_i)$, where $J_1 = \{1, 3, 5, \ \dots \ , T-1\}$ and $J_2 = \{2, 4, 6, \ \dots \ , T\}$. For each subsample, we make inference on β from the regression equation

$$y_t = \beta + u_t\,, \quad t \in J_i\,, \quad \boldsymbol{u}_i = (u_t : t \in J_i)' \sim N(0, \sigma_u^2 I_{T/2})\,, \quad i = 1, 2. \tag{10}$$

A natural critical region with level $\alpha/2$ for testing $\beta = 0$ is then given by

$$W_i(\alpha/2) = \{\boldsymbol{y} \in \mathcal{Y} : |\hat{\beta}_i|/[\hat{V}(\hat{\beta}_i)]^{1/2} \ > \ t(T/2-1; \alpha/4)\},$$

where $\hat{\beta}_i$ is the OLS estimator of β and $\hat{V}(\hat{\beta}_i)$ the usual unbiased estimator of the variance of $\hat{\beta}_i$ from regression (10) using sample $(y_t : t \in J_i)$; $t(T/2-1; \alpha/4)$ is the upper $1-(\alpha/4)$ percentile of the Student's t distribution with $(T/2)-1$ degrees of freedom. We reject $H_0 : \beta = 0$ at level α if $\boldsymbol{y} \in W_1(\alpha/2) \cup W_2(\alpha/2)$.

5.2.2. Alternative procedures

We compared this procedure with two alternatives. The first one consists in testing H_0 using bounds proposed by Hillier and King (1987), Zinde-Walsh and Ullah (1987), Vinod (1976) and Kiviet (1980); see also Vinod and Ullah (1981, ch. 4). The latter are based on standard least squares based tests statistics for testing $\beta = 0$ obtained from the complete sample, such as the t-statistic or its absolute value. Since the distributions of the latter depend on the unknow value of the moving average parameter ψ, one finds instead bounds $t^l(\alpha)$ and $t^u(\alpha)$ which do not depend on the parameter vector θ and such that

$$\mathsf{P}_\theta[T(\boldsymbol{y}) > t^l(\alpha)] \ \geq \ \alpha \ \geq \ \mathsf{P}_\theta[T(\boldsymbol{y}) > t^u(\alpha)],$$

for all $\theta \in \Theta_0,\ \alpha \in (0,1)$. Then the decision rule that consists in rejecting H_0 when $T(\boldsymbol{y}) > t^u(\alpha)$ and accepting H_0 when $T(\boldsymbol{y}) < t^l(\alpha)$ has level α. An inconvenient feature of such procedures is that it may be unconclusive (when $T(\boldsymbol{y}) \in [t^l(\alpha), t^u(\alpha)]$). Obviously, to avoid losses of power, the bounds should be as tight as possible.

In all the above references on bounds tests, the bounds are derived assuming that the MA parameter is known, so that they depend on it, even under the null hypothesis. Therefore we will denote by $t^l_\theta(\alpha)$ and $t^u_\theta(\alpha)$ the lower and upper bounds on $t_\theta(\alpha)$. But as ψ is unknown, we have to find the supremum, $t^u(\alpha)$, over the set $\{t^u_\theta(\alpha) : t^u_\theta(\alpha) :\geq: t_\theta(\alpha),\ \forall\ \theta \in \Theta_0\}$, to make sure that the test based on the rejection region

$$W(\alpha) = \{\boldsymbol{y} \in \mathcal{Y} : T(\boldsymbol{y})\ >\ t^u(\alpha)\}$$

satisfies the level constraint,

$$\sup_{\theta \in \Theta_0} \mathsf{P}_\theta\,[W(\alpha)] :\leq: \alpha.$$

Since the moving average parameter is not restricted by H_0, the set of admissible values for ψ is $\mathbb{R}$. The upper bound is then likely to be quite large.

In the context of model (9), $T(\boldsymbol{y})$ is typically the usual t statistic, its square or its absolute value. Since under H_0, its distribution only depends on ψ (and the sample size), we write t_ψ, t^u_ψ and t^l_ψ instead of t_θ, t^u_θ and t^l_θ, respectively.

Here, we only use the bounds of Zinde-Walsh and Ullah (1987) and Kiviet (1980), denoted by $t^u_{Z,\psi}(\alpha)$ and $t^u_{K,\psi}(\alpha)$, because they are respectively tighter than those of Hillier and King (1987) and Vinod (1976). The supremum $t^u_K(\alpha)$ of $t^u_{K,\psi}(\alpha)$ for $\psi \in \mathbb{R}$ is difficult to establish, but Kiviet (1980, Table 6, p. 357), gives the values of the bounds for $\psi \in \{.2, .3, .5, .9\}$, and it can be seen that $t^u_{K,.9}(\alpha)\ \geq\ t^u_{K,\psi}(\alpha)$, for $\psi \in \{.2, .3, .5, .9\}$. We note that these bounds increase with ψ, and we suspect that the supremum is arbitrarily large, possibly infinite when $\psi = 1$. Nevertheless, we will use $t^u_{K,.9}(\alpha)$ as the relevant upper bound in our simulations. Zinde-Walsh and Ullah (1987) derived bounds on the Fisher statistic (or on the square of the t statistic in our case). $t^u_{Z,\psi}(\alpha)$ is proportional to the ratio $\lambda_{max}(\psi)/\lambda_{min}(\psi)$ of the highest and lowest eigenvalues of the covariance matrix of $\boldsymbol{y}$:

$$t^u_{Z,\psi}(\alpha) = [t_0(\alpha)]^2 \lambda_{max}(\psi)/\lambda_{min}(\psi).$$

We need to make here a remark about the accuracy of Zinde-Walsh and Ullah's bound. Their test rejects H_0 at level α when $[T(\boldsymbol{y})]^2\ >\ \sup_{\psi \in \mathbb{R}} t^u_{Z,\psi}(\alpha) \equiv t^u_Z(\alpha)$. The critical value $t^u_Z(\alpha)$ is not easy to determine analytically, so instead of finding the maximum of $t^u_{Z,\psi}(\alpha)$ on $\mathbb{R}$,we reckoned $t^u_{Z,\psi}(0.05)$ for some values of ψ in the interval $[-1, 2]$. We found a maximum at $\psi = 1$, and a minimum at $\psi = -1$, for every sample size we considered. Although $t^u_{Z,1}(0.05)\ \leq\ t^u_Z(0.05)$, we used this value as the upper bound. Doing so gives more power to the Zinde-Walsh – Ullah test than it really has, because it may reject H_0 more often than it would do if we used $t^u_Z(0.05)$. Despite this fact, $t^u_{Z,1}(0.05)$ is so large (see Table 1) that the power of the test is zero everywhere on the set of alternatives we considered, for any sample size and for any ψ (see Section 5.2.3).

The second alternative consists of using asymptotic tests. In this category, we considered three

TABLE 1. Zinde-Walsh and Ullah's bounds

Sample size T	25	50	75	100
$t^u_{Z,1}(0.05)$	1 164.1972	4 254.3396	9 291.4222	16 274.6855

commonly used tests. The first category includes tests based on a GLS estimation of (9). In the first step, one finds a consistent estimator $\hat{\Omega}$ of Ω and $\hat{P}$ such that $\hat{P}'\hat{P} = \hat{\Omega}^{-1}$. In the second step, we multiply both sides of (9) by $\hat{P}$ and apply OLS to that transformed model. In the last step, we test H_0 using the standard F test. We examine two estimation procedures that lead to a consistent estimator of β, resulting in two test statistics. The first one is detailed in Fomby, Hill, and Johnson (1984, p. 220-221). We denote it by GLS-MM because in the first step of GLS, we estimate the MA parameter ψ by the method of moments. ψ is estimated by minimizing the distance (in the sense of the Euclidean norm on $\mathbb{R}$) between the sample and true first order autocorrelations. The second estimation procedure uses exact maximum likelihood in the first step of GLS and will be denoted by GLS-ML.[3]

The third test we consider is motivated by a central limit theorem [see Brockwell and Davis (1991, p. 219) for instance] which establishes the following property: if a process, with mean β, has an infinite order MA representation with IID error terms and MA coefficients $\psi_i, i = \ldots, -2, -1, 0, 1, 2, \ldots$, satisfying the two following conditions

$$\sum_{i=-\infty}^{\infty} |\psi_i| < \infty, \quad \sum_{i=-\infty}^{\infty} \psi_i \neq 0,$$

then the sample mean of the process is asymptotically normally distributed, with mean β and variance $T^{-1}\sum_{k=-\infty}^{\infty}\gamma(k)$, where $\gamma(k)$ is the autocovariance at lag k. Note that the last condition on the ψ_i's is not satisfied for the MA(1) process (9) with $\psi = -1$, but as ψ is unknown, we might not be aware of this fact or ignore it. Then a natural way of testing H_0 is to estimate β by the sample mean $\bar{Y}_T$ and the asymptotic variance by the consistent estimator proposed in Newey and West (1987):

$$\hat{\phi}_T(p) = \frac{1}{T}\left[r_T(0) + 2\sum_{k=1}^{p}\left(1 - \frac{k}{p+1}\right) r_T(k)\right],$$

where $r_T(k)$ is the sample autocovariance at lag k. Then, if H_0 is true, the statistic

$$\xi_T^{NW} = \frac{T\bar{Y}_T^2}{\hat{\phi}_T(p)}$$

has an asymptotic χ^2 distribution with 1 degree of freedom. We will denote this procedure by NW.[4]

[3]For further discussion of ML estimation in this context, see Tunnicliffe Wilson (1989) ans Laskar and King (1995).

[4]Of course, the list of the methods considered in the present simulation is not exhaustive. For example, possible variants of the NW method include the covariance matrix estimators proposed by Wooldridge (1989). Bayesian methods [see Kennedy and Simons (1991)] and marginal likelihood methods [see King (1996)] could also be used in this context.

TABLE 2. Size and critical values of 5% level asymptotic tests

Sample size	$T = 25$				$T = 50$			
	ψ^*	Size (%)	ACV	CCV	ψ^*	Size (%)	ACV	CCV
GLS-MM	-0.5	19.22	4.25968	30.664	-0.5	18.59	4.0384	59.555
GLS-ML	-0.5	27.87	4.25968	37.979	-0.5	15.06	4.0384	14.615
NW	1	15.03	3.840	8.459	1	10.25	3.840	5.789
Sample size	$T = 75$				$T = 100$			
	ψ^*	Size (%)	ACV	CCV	ψ^*	Size (%)	ACV	CCV
GLS-MM	-0.5	16.98	3.97024	64.502	-0.5	14.98	3.9371	38.789
GLS-ML	-0.5	10.13	3.97024	6.396	-0.5	7.84	3.9371	4.983
NW	1	8.82	3.840	5.243	1	8.08	3.840	4.907

Before presenting the results of our simulations, we wish to insist on a very important condition one has to impose when comparing the relative performance of two tests. In the Neyman-Person approach to the problem of testing a null hypothesis H_0 against an alternative H_1, it is meaningless to say that a test A has a higher power than a test B, if the two tests do not have the same level. A test of $H_0 : \theta \in \Theta$ with critical region W has *level* α if $\sup_{\theta \in \Theta_0} \mathsf{P}_\theta(W) \leq \alpha$, and it has *size* α if $\sup_{\theta \in \Theta_0} \mathsf{P}_\theta(W) = \alpha$; see Lehmann (1986, Section 3.1). When the probability of rejection $\mathsf{P}_\theta(W)$ under the null hypothesis ($\theta \in \Theta_0$) is not constant, *controlling the level* of a test requires one to ensure that $\mathsf{P}_\theta(W) \leq \alpha$ for all $\theta \in \Theta_0$, and *controlling its size* involves ensuring that the maximum (or supremum) of the rejection probabilities over $\theta \in \Theta_0$ is equal to α. Of course, this may lead to a difficult search over the parameter space. When the distribution of a test statistic depends on a nuisance parameter (the unknown value of ψ, in the present case), correcting the size of a test requires one to find a critical value such that the maximum probability of rejection under the null hypothesis (irrespective of ψ) is equal to α. A way of doing this is to detect the value of ψ, ψ^* say, for which the discrepancy between the level and the size is maximum. For that value, we simulate S times the test statistic, $T_{\psi^*}(\boldsymbol{y})$.[5] We then take the observation of rank $((95 \times S)/100) + 1$ of the statistic as our corrected 5%-level critical value: we reject H_0 at level 5% when $T_{\psi^*}(\boldsymbol{y})$ is larger than or equal to that value. Table 2 reports ψ^*, the size (in %), the 5% asymptotic critical value (*ACV*), and the 5% corrected critical value (*CCV*), for each sample size T, and each of the three asymptotic procedures.

But space and time limitations have precluded us from including all proposed methods in our simulations.

[5]For all asymptotic test procedures, we set $S = 10\,000$.

5.2.3. Simulations

In our simulations, we proceeded as follows. For $\psi \in \{-1, -.5, 0, .5, 1\}$ and $T \in \{25, 50, 75, 100\}$, we considered a grid of β values around $\beta_0 = 0$. In each case, 1000 independent samples $(y_t, t = 1, 2, \ldots, T)$ were generated and the following statistics were computed: (1) the t statistic based on the whole sample; (2) the t statistic based on the two subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$ containing the odd and even observations respectively; (3) the GLS-MM and GLS-ML based F statistics; (4) the ξ_n^{NW} statistic. Using these statistics, the following tests were implemented at level 5% and the corresponding rejection frequencies were computed: (a) Zinde-Walsh and Ullah's bounds test; (b) Kiviet's bounds test;[6] (c) GLS-MM asymptotic test (corrected and uncorrected for size); (d) GLS-ML asymptotic test (corrected and uncorrected for size); (e) NW asymptotic test (corrected and uncorrected for size); (f) the induced test which combines the standard t-tests based on the two subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$; (g) the separate tests based on the subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$. The results are presented in Figures 2 to 6 and Tables 3 to 7.

As it became clear in the description of the induced test, when applying such a procedure to model (9), one is led to split the sample in two, and make two tests at level $\alpha/2$. At first sight, the procedure displays features that may seem quite unattractive. First, it splits the available sample in two, and second it combines two tests whose levels are only $\alpha/2$ (instead of α). From these two remarks, one may expect the procedure to lack power. But we should keep in mind that, since the two "sub-tests" have level α/2, the resulting induced test has level certainly greater than $\alpha/2$ (although not greater then α). Furthermore, this test actually uses the information contained in the whole sample. Then it becomes less clear whether the induced test procedure automatically leads to a loss of power relatively to other alternative tests. Two questions arises from these remarks: (1) is combining preferable to not combining ?, *i.e.*, should our decision at level α rely on an induced test procedure or on a test based on one of the subsample only ? (2) how does the induced test compare with the procedures mentionned in Section 5.2.2 ?

Figures 2 to 6 answer the first question. They show that the power of the induced test (solid line) is generally higher than that of an α-level test based on one of the two subsamples (dashed lines). In other words, combining is preferable to not combining. When it is not the case [when the true value of the MA parameter is unity, $\psi = 1$, see Figures 2(a) to 2(d)], the power loss from using the induced test is very small, so that one would usually prefer the sample-split procedure that uses all the observations.

Tables 3 to 7 report the estimated probability of a rejection of $H_0 : \beta = 0$ for different sample sizes ($T \in \{25, 50, 75, 100\}$) and true values for β $(\beta \in \{-1, -.8, -.5, -.2, 0, .2, .5, .8, 1\})$, for each one of the test procedures of Section 5.2.2. If we first consider bounds tests, we note that the Kiviet test is dominated by the induced test, except for $\psi = .5$ and $\psi = 1$. We already mentioned in Section 5.2.2 that the bound which has been used here, namely $t^u_{K,.9}(0.05)$, is not appropriate because we do not know whether this value satisfies the level constraint:

$$\sup_{\theta \in \Theta_0} \mathsf{P}_\theta \left[\{ \boldsymbol{y} \in \mathcal{Y} : T(\boldsymbol{y}) > t^u_{K,.9}(0.05) \} \right] \leq 0.05.$$

[6] Because Kiviet (1980) does not provide the upper bound for $T = 75$ and $T = 100$, we did not investigate the behaviour of Kiviet's test for these values of T.

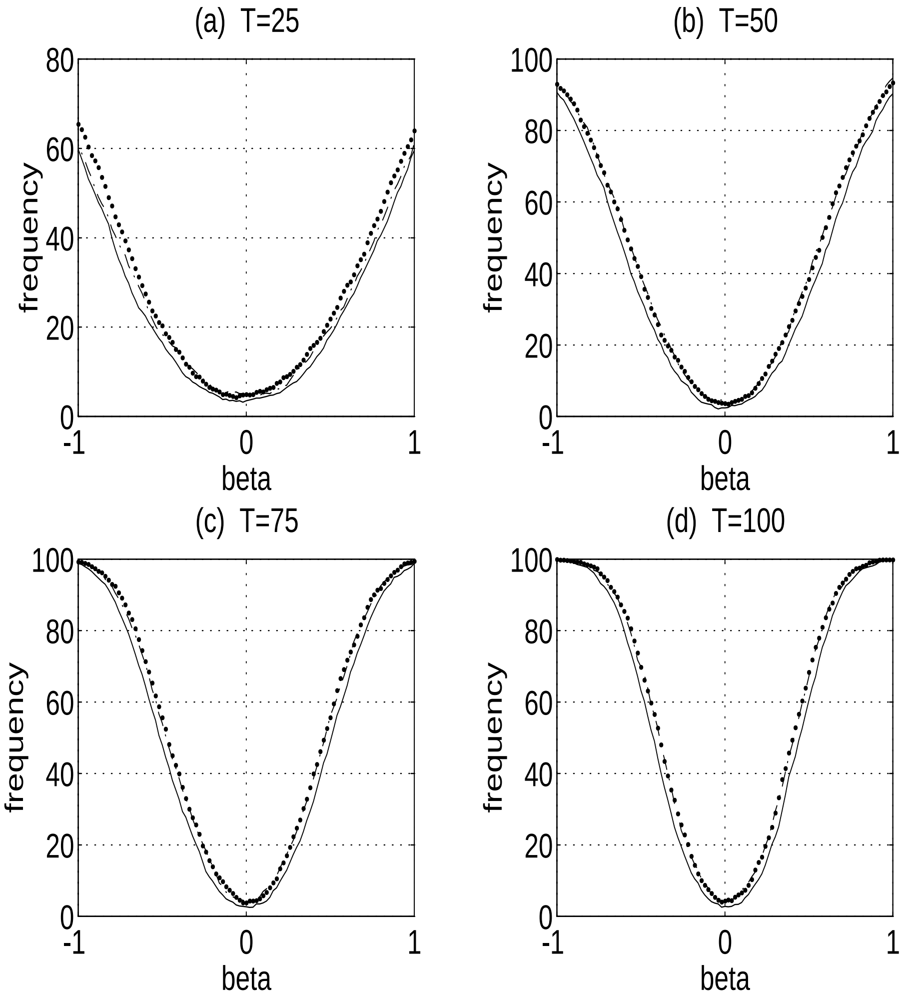

FIGURE 2. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 1$.
$T = 25, 50, 75, 100$.
Induced test (—); tests based on subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$ ($\cdots$ and $- \cdot -$).

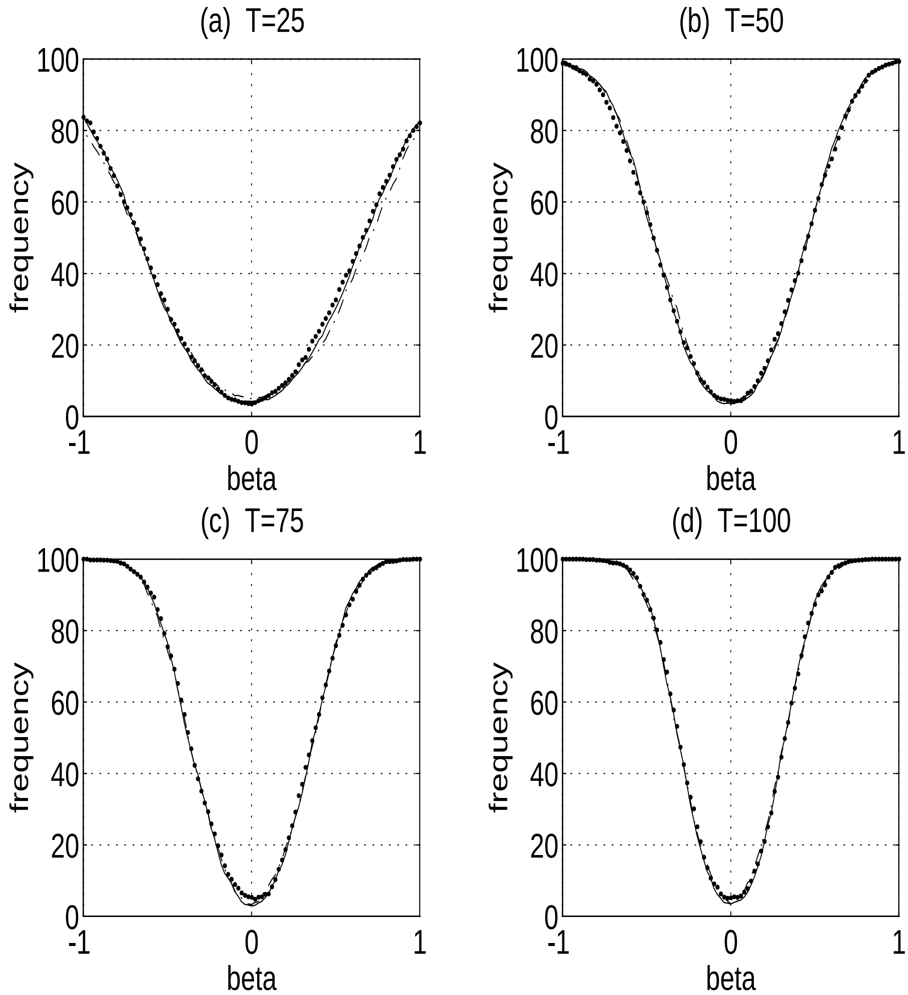

FIGURE 3. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 0.5$.
$T = 25, 50, 75, 100.$
Induced test (—), tests based on subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$ ($\cdots$ and $- \cdot -$).

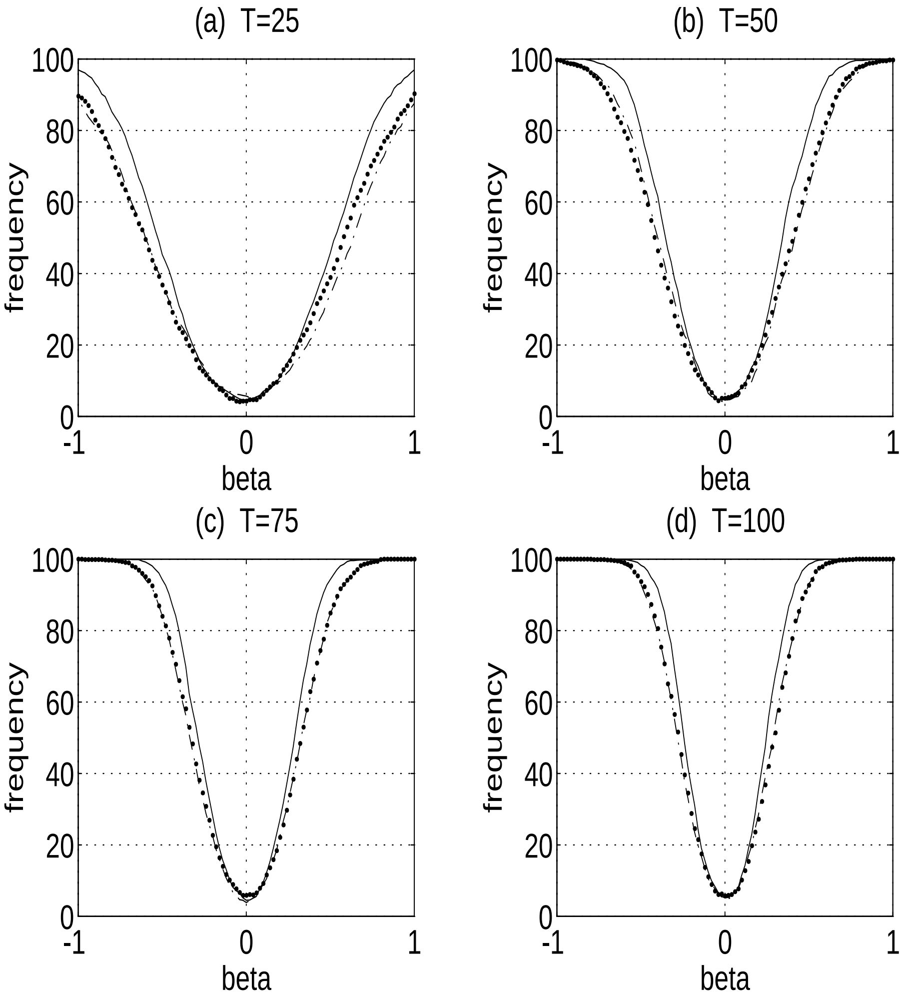

FIGURE 4. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 0$.
$T = 25, 50, 75, 100$.
Induced test (—), tests based on subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$ ($\cdots$ and $- \cdot -$).

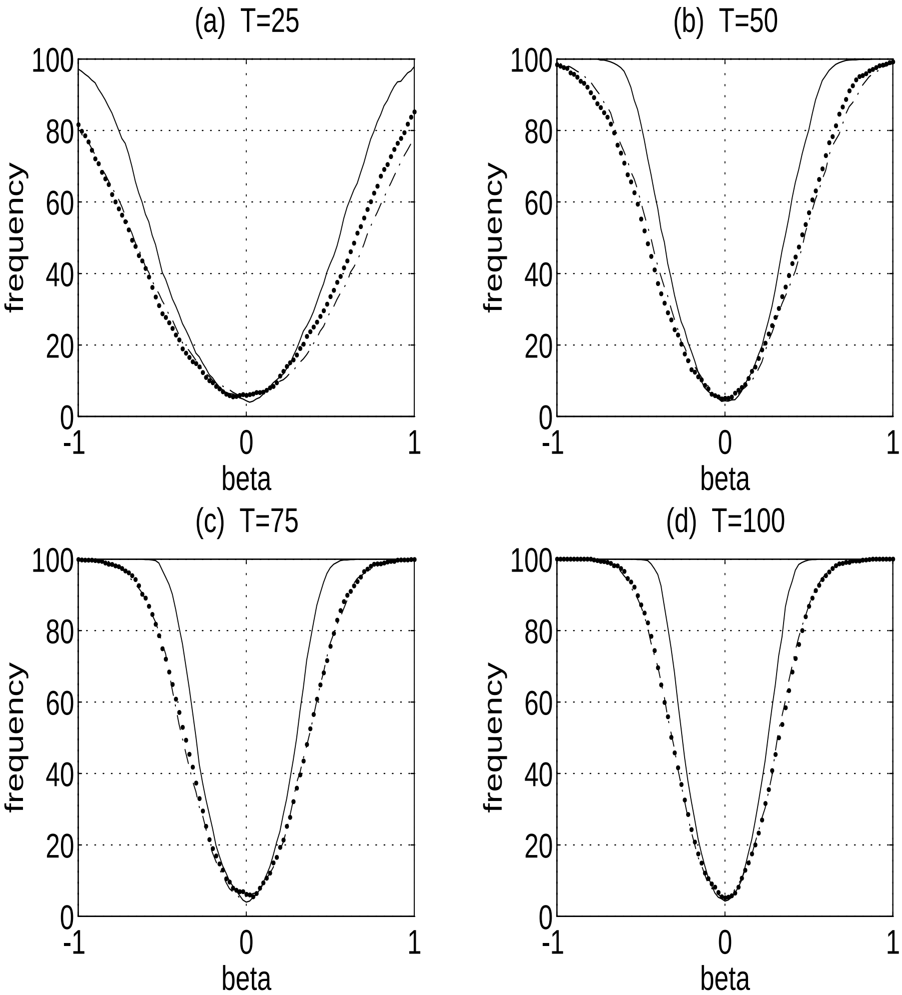

FIGURE 5. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -.5$.
$T = 25, 50, 75, 100$.
Induced test (—), tests based on subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$ ($\cdots$ and $- \cdot -$).

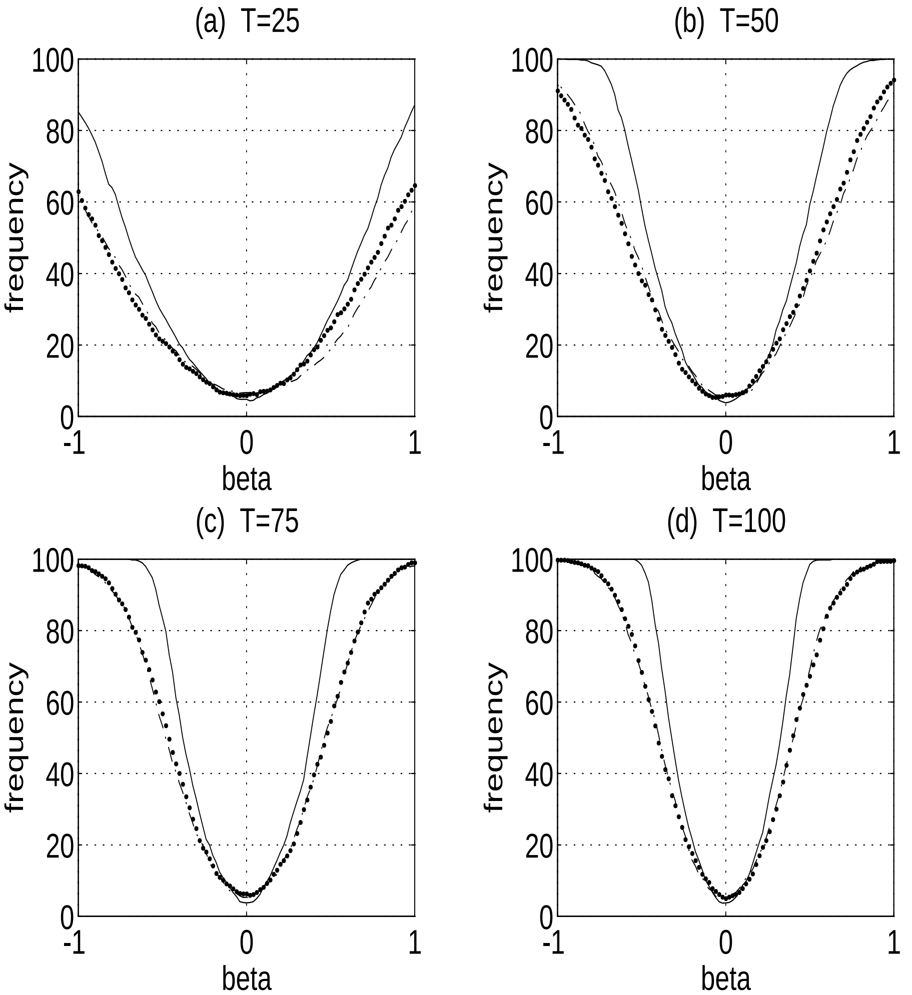

FIGURE 6. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -1$.
$T = 25, 50, 75, 100$.
Induced test (—), tests based on subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$ ($\cdots$ and $- \cdot -$).

TABLE 3. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 1$

	Sample size $T = 25$					Sample size $T = 75$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	0.80	3.40	56.00	66.70	59.80	97.70	0.70	99.10	98.80
-0.8	0.20	1.10	37.80	46.60	40.00	84.90	0.10	91.50	89.70
-0.5	0.00	0.20	17.00	20.10	16.60	40.00	0.00	52.00	48.00
-0.2	0.00	0.00	6.10	6.20	5.20	6.60	0.00	14.30	9.90
-0.1	0.00	0.00	4.50	5.00	3.50	2.20	0.00	7.00	4.40
0.0	0.00	0.00	4.30	4.30	3.50	1.10	0.00	4.30	2.70
0.1	0.00	0.00	5.20	5.10	4.30	2.10	0.00	6.50	3.80
0.2	0.00	0.00	6.50	7.60	5.30	5.80	0.00	11.80	9.80
0.5	0.10	0.60	18.10	21.20	18.10	38.60	0.00	55.20	48.90
0.8	0.10	1.80	38.80	46.40	40.50	83.60	0.20	91.10	89.40
1.0	0.30	3.90	55.00	64.90	59.60	96.70	0.80	98.90	98.80

	Sample size $T = 50$					Sample size $T = 100$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	46.00	0.20	92.40	93.30	90.70	99.50	25.60	99.80	99.70
-0.8	21.50	0.00	77.30	78.30	72.30	97.00	6.70	97.70	96.90
-0.5	3.50	0.00	40.10	40.20	32.90	61.30	0.20	69.10	63.20
-0.2	0.20	0.00	9.60	10.00	6.80	11.90	0.00	17.40	12.20
-0.1	0.00	0.00	4.90	4.50	3.50	4.60	0.00	7.90	5.00
0.0	0.10	0.00	3.90	3.90	2.40	3.40	0.00	4.40	2.80
0.1	0.10	0.00	4.30	4.70	3.50	5.60	0.00	7.40	3.90
0.2	0.20	0.00	8.20	9.20	6.60	11.60	0.00	15.10	10.30
0.5	3.70	0.00	34.90	40.60	33.70	62.80	0.30	67.00	60.40
0.8	21.00	0.10	74.00	78.60	72.70	96.50	4.60	97.30	96.60
1.0	45.10	0.40	89.70	95.00	90.30	99.90	22.90	99.90	99.80

TABLE 4. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = .5$

β	Sample size $T = 25$					Sample size $T = 75$			
	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	4.80	10.70	79.40	85.30	83.90	100.00	8.50	100.00	100.00
-0.8	2.50	4.70	61.20	67.30	66.30	98.80	1.00	99.50	99.40
-0.5	0.60	1.00	27.20	30.30	29.00	65.50	0.10	77.50	76.70
-0.2	0.00	0.00	6.70	6.80	7.00	11.30	0.00	19.40	18.10
-0.1	0.00	0.00	4.60	3.90	4.60	2.80	0.00	8.40	7.00
0.0	0.00	0.00	4.30	3.50	4.20	0.80	0.00	4.30	2.90
0.1	0.00	0.00	5.60	4.70	4.70	3.10	0.00	7.80	6.40
0.2	0.10	0.10	7.80	8.10	8.30	10.40	0.00	19.30	17.10
0.5	0.50	0.60	27.80	30.90	30.20	64.70	0.10	77.70	75.60
0.8	1.60	4.60	60.50	66.80	64.70	97.80	1.10	99.50	99.40
1.0	5.00	12.10	78.30	84.80	82.80	100.00	7.20	100.00	100.00

β	Sample size $T = 50$					Sample size $T = 100$			
	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	83.00	2.40	99.30	99.80	99.40	100.00	65.10	100.00	100.00
-0.8	50.80	0.20	93.80	94.30	94.00	99.80	26.80	99.90	99.90
-0.5	10.80	0.00	58.30	59.20	55.80	86.80	1.10	90.90	89.00
-0.2	0.40	0.00	12.70	11.50	11.90	20.50	0.00	24.70	22.50
-0.1	0.10	0.00	6.50	4.10	5.60	6.00	0.00	10.90	7.90
0.0	0.10	0.00	3.30	2.90	3.60	3.40	0.00	4.30	3.50
0.1	0.20	0.00	5.70	4.30	5.20	7.40	0.00	8.70	6.70
0.2	0.40	0.00	13.00	12.20	11.90	19.10	0.00	22.30	20.70
0.5	10.40	0.00	59.70	59.30	57.40	88.10	1.00	89.70	88.10
0.8	50.60	0.50	94.80	96.00	94.70	100.00	25.20	100.00	99.90
1.0	82.10	3.50	99.40	99.70	99.50	100.00	63.30	100.00	100.00

TABLE 5. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = 0$

	Sample size $T = 25$					Sample size $T = 75$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	35.30	45.80	97.80	96.20	97.00	100.00	70.30	100.00	100.00
-0.8	19.30	23.10	89.30	81.70	85.20	100.00	28.20	100.00	100.00
-0.5	6.20	5.10	48.40	31.70	45.30	96.90	0.80	98.50	94.30
-0.2	2.50	0.90	8.90	2.80	9.90	25.80	0.00	33.70	28.20
-0.1	1.20	0.30	3.30	0.90	6.60	7.20	0.00	11.80	10.40
0.0	1.00	0.20	3.10	0.40	4.70	1.40	0.00	2.70	4.50
0.1	1.40	0.60	5.10	0.80	7.00	6.90	0.00	10.60	9.50
0.2	1.90	0.70	9.40	3.00	11.00	23.40	0.00	32.10	27.10
0.5	7.00	5.40	48.30	31.80	45.70	96.10	0.80	98.00	94.30
0.8	20.80	23.80	87.50	80.50	85.90	100.00	26.70	100.00	100.00
1.0	37.70	45.30	98.00	97.20	97.00	100.00	70.70	100.00	100.00

	Sample size $T = 50$					Sample size $T = 100$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	99.70	40.80	100.00	100.00	100.00	100.00	99.80	100.00	100.00
-0.8	95.00	13.20	100.00	99.90	99.60	100.00	93.20	100.00	100.00
-0.5	45.10	0.50	88.60	73.30	80.20	99.80	20.80	99.60	98.30
-0.2	2.70	0.00	21.40	6.90	19.40	42.50	0.00	47.50	36.30
-0.1	0.50	0.00	7.50	1.40	6.50	12.30	0.00	14.60	12.50
0.0	0.30	0.00	2.40	0.10	4.70	3.90	0.00	3.30	5.20
0.1	0.90	0.00	6.90	1.60	8.00	12.60	0.00	12.70	11.50
0.2	2.60	0.00	20.80	7.10	18.30	43.50	0.20	43.60	35.40
0.5	43.80	0.60	89.00	72.90	80.20	100.00	17.40	99.60	98.60
0.8	95.10	13.30	99.90	99.70	99.60	100.00	94.00	100.00	100.00
1.0	99.60	41.30	100.00	100.00	100.00	100.00	99.90	100.00	100.00

TABLE 6. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -0.5$

	Sample size $T = 25$					Sample size $T = 75$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	93.50	91.70	100.00	97.50	97.20	100.00	100.00	100.00	100.00
-0.8	82.80	76.30	99.80	77.30	84.90	100.00	98.50	100.00	100.00
-0.5	49.80	41.70	82.70	10.10	40.30	100.00	56.70	100.00	96.90
-0.2	18.00	18.20	6.30	0.00	10.80	82.90	14.70	65.10	24.60
-0.1	8.00	8.60	1.40	0.00	6.40	29.60	9.30	10.80	9.90
0.0	4.50	4.30	0.10	0.00	4.40	3.60	5.30	0.40	4.00
0.1	10.60	9.70	1.90	0.00	6.20	29.20	9.60	12.20	9.40
0.2	18.90	19.50	8.70	0.00	11.10	81.50	15.30	67.10	23.80
0.5	50.00	41.80	81.40	10.50	42.80	100.00	58.00	100.00	97.60
0.8	82.10	76.00	99.80	76.60	84.40	100.00	98.10	100.00	100.00
1.0	93.10	92.40	100.00	97.90	97.90	100.00	100.00	100.00	100.00

	Sample size $T = 50$					Sample size $T = 100$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	100.00	97.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00
-0.8	100.00	86.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00
-0.5	98.60	38.90	100.00	66.10	81.80	100.00	97.00	100.00	99.90
-0.2	33.50	16.80	39.70	0.10	18.30	94.90	20.60	84.60	32.30
-0.1	10.30	10.30	4.60	0.00	7.00	45.50	9.40	18.60	11.00
0.0	2.90	5.00	0.10	0.00	4.40	5.00	4.40	0.40	4.30
0.1	9.50	9.10	5.60	0.00	7.10	44.40	9.30	16.80	10.10
0.2	33.90	15.80	37.20	0.00	17.30	95.30	19.70	82.70	32.00
0.5	99.00	41.30	99.90	67.00	80.20	100.00	97.40	100.00	99.80
0.8	100.00	87.70	100.00	99.90	99.90	100.00	100.00	100.00	100.00
1.0	100.00	98.10	100.00	100.00	100.00	100.00	100.00	100.00	100.00

TABLE 7. Rejection frequencies of $H_0 : \beta = 0$ in model (9) with $\psi = -1$

	Sample size $T = 25$					Sample size $T = 75$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	99.50	93.10	100.00	79.90	85.20	100.00	100.00	100.00	100.00
-0.8	98.20	80.10	100.00	34.80	64.00	100.00	99.20	100.00	100.00
-0.5	89.80	58.30	89.10	0.00	28.70	100.00	82.30	100.00	83.60
-0.2	49.90	40.80	0.10	0.00	9.00	100.00	51.60	84.60	17.10
-0.1	2.20	4.30	0.00	0.00	5.90	98.80	47.00	0.10	8.10
0.0	0.00	0.00	0.00	0.00	4.80	0.70	0.00	0.00	3.80
0.1	2.30	2.80	0.00	0.00	6.10	99.10	47.00	0.10	8.10
0.2	52.40	41.40	0.40	0.00	9.30	100.00	51.40	87.90	17.90
0.5	90.70	59.40	88.10	0.40	28.50	100.00	82.00	100.00	85.50
0.8	98.10	81.30	100.00	34.30	64.70	100.00	99.30	100.00	100.00
1.0	99.60	93.10	100.00	79.40	87.20	100.00	100.00	100.00	100.00

	Sample size $T = 50$					Sample size $T = 100$			
β	GLS-ML	GLS-MM	NW	Kiv.	Ind. test	GLS-ML	GLS-MM	NW	Ind. test
-1.0	100.00	98.30	100.00	100.00	100.00	100.00	100.00	100.00	100.00
-0.8	100.00	92.60	100.00	99.20	99.00	100.00	100.00	100.00	100.00
-0.5	100.00	69.30	100.00	13.10	58.20	100.00	99.20	100.00	98.30
-0.2	96.60	46.00	29.10	0.00	12.10	100.00	61.20	99.70	21.70
-0.1	75.00	42.10	0.00	0.00	6.20	100.00	48.60	0.50	8.50
0.0	0.00	0.00	0.00	0.00	3.80	1.10	0.00	0.00	3.80
0.1	76.00	41.90	0.00	0.00	6.50	100.00	48.90	0.40	8.70
0.2	96.40	45.80	28.20	0.00	10.80	100.00	61.90	99.50	20.70
0.5	100.00	69.30	100.00	12.10	59.20	100.00	99.00	100.00	98.50
0.8	100.00	92.30	100.00	99.50	98.70	100.00	100.00	100.00	100.00
1.0	100.00	98.30	100.00	100.00	100.00	100.00	100.00	100.00	100.00

In other words, a critical region based on Kiviet's bounds has an unknown level. Moreover, what makes the induced test more attractive relatively to Kiviet's test is that it avoids the calculation of a bound that changes with the sample size. Finally, because Zinde-Walsh and Ullah's upper bounds are so large (see Table 1), the power of their test is zero for all ψ. These are not reported in Tables 3-7.

The most surprising result which emerges from our Monte Carlo study can be seen in Tables 3, 4 and 5. Once the asymptotic critical values used for the GLS-MM and GLS-ML tests have been corrected so that the corresponding critical regions have the desired level, our procedure becomes more powerful than these alternatives for many plausible values of ψ. The difference between estimated power functions grows as ψ increases, but diminishes when the sample size T gets larger. The GLS-MM method seems to be the worst of all the asymptotic procedures studied here, whereas GLS-ML appears to benefit from the asymptotic efficiency property of maximum likelihood estimators. But for non negative values of ψ, the sample size has to be $T = 100$ for the GLS-ML test to have a probability of correctly rejecting the null as high as the induced test. The GLS-MM test is still dominated for some negative values of ψ $(\psi = -.5)$, irrespective of the sample size. Only when ψ is close to -1 does this procedure become admissible.

While the two commonly used asymptotic inference procedures, GLS-MM and GLS-ML, cannot be recommended on the ground of our Monte Carlo study, the conclusion is less negative for the NW method. Except for small sample sizes $(T = 25)$ and large values of the MA parameter $(\psi = 1, .5)$, it does better than the induced test procedure. This result is somewhat unexpected because the Newey-West estimator of $\mathsf{V}(\bar{Y}_T)$ does not take into account the autocovariance structure of the process. However, although the induced test is conservative, it is more powerful than NW test for alternatives close to the null hypothesis when ψ is negative. Furthermore, it is important to remember that the NW test suffers from level distortions (see Table 2) that are not easy to correct in practice.

5.2.4. An example: an induced test on the mean of the Canadian per capita GDP series

We now apply our procedure to test the nullity of the mean of a process that has a MA(1) representation. Our series is the first difference of the Canadian per capita GDP, denominated in real 1980 Purchasing Power Parity-adjusted US dollars, observed yearly from 1901 to 1987. It is taken from Bernard and Durlauf (1995). Figure 7 plots the series.

Using standard Box-Jenkins procedure (autocorrelation and partial autocorrelation functions), we identified a MA(1) process for the series (see Table 8).

We then consider a model like (9). ML estimation of (9) gives $\hat{\beta} = 136.1810$ and $\hat{\psi} = 0.4211$ with estimated variances 945.1919 and 0.0095, respectively. The estimated $\mathsf{Cov}(\hat{\beta}, \hat{\psi})$ is 0.0834 and the sample variance of the residuals is 40117.5725.

To implement an induced test for the nullity of the mean parameter, β, at level 5%, we split the sample in two parts, $\{y_t : t \in J_i\}, i = 1, 2$, which contain respectively the odd and the even observations, and make two 2.5% tests of $\beta = 0$, using the statistics $t_i = \sqrt{n_i}\, |\bar{y}_i|/s_i$, where $\bar{y}_i = \sum_{j \in J_i} y_j/n_i$, $s_i^2 = \left(\sum_{j \in J_i}(y_j - \bar{y}_i)^2\right)/(n_i - 1)$, and n_i is the size of subsample i, $i = 1, 2$. We reject the null hypothesis when $t_1 > t(\alpha/4, \nu_1)$ or $t_2 > t(\alpha/4, \nu_2)$, where $t(\alpha/4, \nu)$ is the

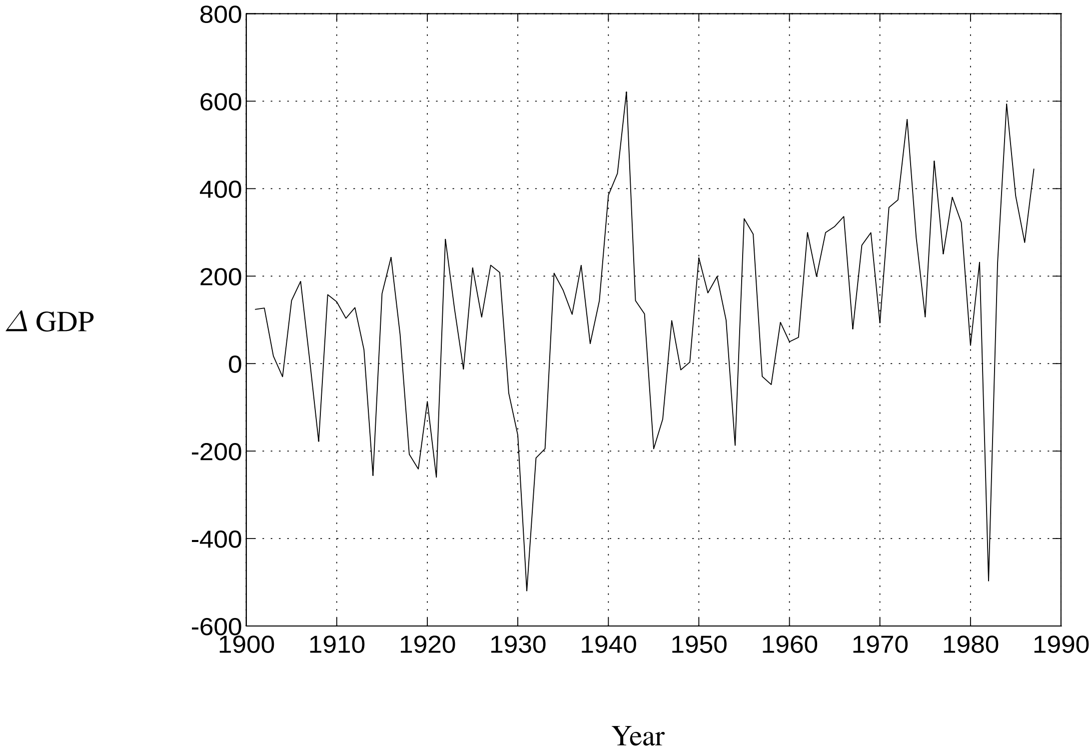

FIGURE 7. First differences of the Canadian per capita GDP.
Source: Bernard and Durlauf (1995)

TABLE 8. Sample autocorrelations of the Canadian per capita GDP series

Lag	1	2	3	4	5	6	7	8	9	10	11	12
Autocorrelation	.41	.19	.10	-.04	.05	.07	.12	.04	-.04	.09	.08	.20
Standard Error	.11	.12	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13
Ljung-Box Q-statistic	15.4	18.8	19.8	19.9	20.2	20.6	22.1	22.3	22.5	23.3	24.0	28.3

TABLE 9. Induced and asymptotic tests. Model: $y_t = \beta + \varepsilon_t + \psi\varepsilon_{t-1}$

	$i = 1$	$i = 2$	GLS-MM	GLS-ML
$\hat{\beta}$	127.6836	125.4406	122.2522	123.6574
$t(\nu)$	$4.1892\ (43)$	$3.5076\ (42)$	$7.5112\ (86)$	$6.9345\ (86)$
p-value	0.00014	0.00109	0.00000	0.00000
$\hat{\psi}$	$--$	$--$	0.5298	0.4221

$1 - (\alpha/4)$ percentile of the Student's t distribution with ν degrees of freedom. We also perform both GLS-MM and GLS-ML asymptotic tests. Our results are reported in Table 9. $\hat{\beta}$ is the two step estimator of β, $\hat{\psi}$ the estimator of ψ that has been obtained in the first step to estimate the error covariance matrix, and t the test statistic, whose distribution will be approximated by a Student's t distribution with 86 degrees of freedom.

Both subtests reject the null hypothesis at level 2.5%. Hence the induced test rejects the nullity of the mean at level 5%. The two asymptotic tests also reject the null hypothesis, if we admit that the asymptotic critical value is a good approximation when the sample size is 87. Our findings are consistent with the results of the Monte Carlo study of Section 5.2.3. For similar sample sizes ($T = 75$ or $T = 100$) we found that the GLS-MM test produces larger values of the test statistic than the GLS-ML test does. This is what we have here with $T = 87$.

If we decide to include a linear trend in the mean of the MA(1) process, our induced test procedure still applies. The per capita GDP series now admits the following representation

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t + \psi\varepsilon_{t-1}\,, \qquad \varepsilon_t \overset{ind}{\sim} N(0, \sigma^2)$$

for $t = 1, 2, \ldots, T$. We consider the three hypotheses:

$$H_0^{(0)} : \beta_0 = 0\,, \qquad H_0^{(1)} : \beta_1 = 0\,, \qquad H_0 : \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

For each hypothesis, we perform the induced test as well as the asymptotic tests. Results appear in Table 10.

We note that only one of the subtests rejects the presence of a linear trend. However, according to our decision rule, this is enough to reject $H_0^{(1)}$. Both GLS-MM and GLS-ML unambiguously reject this hypothesis. But we know from our simulations that the asymptotic tests tend to reject the null too often when it is true. For the parameters $\beta_j, : j = 1, 2$, we also report two confidence intervals I_1^j and I_2^j, each with level 97.5%, based on the two subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$. The intersection $I_1^j \bigcap I_2^j$ gives the set of values $\gamma \in \mathbb{R}$ such that the hypothesis $H_0^j(\gamma) : \beta_j = \gamma$ is not

TABLE 10. Induced and asymptotic tests. Model: $y_t = \beta_0 + \beta_1 t + \varepsilon_t + \psi\varepsilon_{t-1}$

	$i=1$	$i=2$	GLS-MM	GLS-ML
$\hat{\beta}_0$	-37.0695	-6.2981	-22.5258	-22.5578
$\hat{\beta}_1$	3.7444	2.9941	3.3507	3.3554
$t_0(\nu)$	-0.6832 (42)	-0.0903 (41)	-0.6726 (85)	-0.6577 (85)
p-value $\left(H_0^{(0)}\right)$	0.49823	0.92849	0.50303	0.51251
$t_1(\nu)$	3.5058 (42)	2.1674 (41)	5.0392 (85)	4.9319 (85)
p-value $\left(H_0^{(1)}\right)$	0.00110	0.03606	0.00000	0.00000
$F(\nu_1, \nu_2)$	17.2244 $(2, 42)$	9.0421 $(2, 41)$	37.9250 $(2, 85)$	39.3875 $(2, 85)$
p-value (H_0)	0.00000	0.00056	0.00000	0.00000
$\hat{\psi}$	$--$	$--$	0.3536	0.3253

t_j, $j = 1, 2$, and F denote the Student's t and Fisher's F statistics used for testing $H_0^{(j)}$, $j = 1, 2$, and H_0, respectively.

rejected at level 5% by the induced test. These intervals are:

$$I_1^0 = [-163.2002, 89.0612], \quad I_2^0 = [-168.5787, 155.9825],$$
$$I_1^1 = [1.2616, 6.2272], \quad I_2^1 = [-0.2201, 6.2083],$$

yielding the following 95% confidence intervals for β_0 and β_1 :

$$\beta_0 \in [-163.2002, 89.0612], \quad \beta_1 \in [1.2616, 6.2083].$$

These entail that β_0 is not significantly different from 0, while β_1 is.

When we apply the induced test procedure, we implicitely assume that we have correctly identified a MA(1) process. An interesting issue is to look at what we get if, instead of the true MA(1) representation, we use a MA(2) model to build our test statistics. In this case, we split the sample in three parts, make three tests at level 5/3% and reject the null $H_0 : \beta = 0$ when our sample falls in one of the three critical regions. The results are in Table 11.

We first note that as two of the subsample based tests reject H_0 at level 5/3%, we reject the null at level 5%. We also note that both asymptotic tests reject the null hypothesis. But we know that using 5% critical values obtained from asymptotic distributions leads to a probability of making a type I error larger than 5%. Therefore, although asymptotic tests and tests based on the sample split yield the same decision of rejecting H_0, we put more confidence in the sample split procedure.

TABLE 11. Induced and asymptotic tests. Model: $y_t = \beta + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2}$

	$i = 1$	$i = 2$	$i = 3$	GLS-MM	GLS-ML
$\hat{\beta}$	122.4644	130.7687	126.4918	128.9405	129.7032
$t(\nu)$	2.6573 (28)	4.1328 (28)	2.9150 (28)	3.6828 (86)	3.2812 (86)
p-value	0.01286	0.00029	0.00692	0.00040	0.00149
$\hat{\psi}_1$	$--$	$--$	$--$	0.4096	0.3931
$\hat{\psi}_2$	$--$	$--$	$--$	0.2354	0.1037

t denotes the statistics used for testing $H_0 : \beta = 0$.

6. Concluding remarks

In this paper, we proposed a set of inference methods for comparing and pooling information obtained from different data sets, which simply use separate tests (or confidence sets) based on the different data sets. The methods described are based on a systematic exploitation of Boole-Bonferroni inequalities and can yield exact tests and confidence sets without the need to specify at all the relationship between the data sets, even with small sample sizes. As a result, they are quite versatile and usually easy to implement. The general problems studied include: (I) combining separate tests based on different data sets for an hypothesis of interest (more precisely, for the intersection of similar hypotheses), to obtain more powerful tests; (II) comparing parameters estimated from the different data sets (*e.g.*, to test their equality); (III) combining confidence regions based on different samples to obtain a more accurate confidence set. For problem I, we were led to consider Bonferroni-type induced tests; for problem II, we proposed *empty intersection* tests; and for problem III, we suggested taking the intersection of separate confidence sets with appropriate levels.

We also showed that the methods proposed can be quite useful in various models where usual inference procedures based on a complete sample involve difficult distributional problems (*e.g.*, because of nuisance parameters), but for which distributional properties of test statistics computed on appropriately selected subsamples are simpler. This leads to an interesting form of *sample-split* method. One first splits the sample into several subsets of observations from which separate inferences (tests or confidence sets) are obtained. Then these results are recombined, using the general union-intersection (UI) methods already described, to obtain a single inference which uses the full sample. The way the data set is split depends on the model considered. In some situations the structure naturally suggests the division. This is for example true when the model contains several equations. In other cases, the division is based on more elaborate arguments, as in moving average models.

The union-intersection/sample-split methods proposed can be applied to a wide spectrum of econometric situations and models. We discussed and illustrated their applications in two cases where only asymptotic methods are typically available, namely inference in SURE models and linear regressions with MA errors. In the latter case, we also presented an extensive Monte-Carlo study comparing the sample-split union-intersection (SS-UI) method for testing an hypothesis about a mean with other available approaches. Two main conclusions emerged from these results: first, they provided further evidence on the size distortions associated with usual asymptotic procedures; second, they showed that SS-UI tests not only have the predicted levels, but enjoy good power properties. In view of the fact that these methods involve splitting the sample and lead to conservative procedures, hence leading one to expect a power loss, this is indeed quite remarkable. Our results show that the Bonferroni-based recombination of the evidence obtained from the different subsamples apparently makes up for the loss. For another application of SS-UI approach to autoregressive and other types of dynamic models, the reader may consult Dufour and Torrès (1997).

Before closing, it is worthwhile noting a few other points. First, the UI (or UI-SS) procedures are often simpler to implement than usual asymptotic procedures. For SURE models and linear regressions with MA errors, they only require critical values from standard distributions. For MA models, they avoid the task of estimating MA coefficients. Second, they offer some extra robustness to model specification, as illustrated by SURE models where no assumption on the relationship between the different equations is needed. Thirdly, although we stressed here the derivation of finite sample methods, there is nothing that forbids the application of UI (or UI-SS) methods to situations where only asymptotically justified tests or confidence sets are available from the separate data sets. In such cases, the methods are applied in exactly the same way. This feature may be especially attractive for gaining robustness to model specification. We think all these properties make this UI-SS approach an attractive and potentially quite useful addition to the methods available to applied econometricians.

References

ALT, F. B. (1982): "Bonferroni Inequalities and Intervals," in *Encyclopedia of Statistical Sciences, Volume 1*, ed. by N. L. Johnson, S. Kotz, and C. Read, vol. 1, pp. 294–301. John Wiley & Sons, New York.

ANGRIST, J. D., AND A. B. KRUEGER (1994): "Split Sample Instrumental Variables," Technical Working Paper 150, N.B.E.R., Cambridge, MA.

ARNOLD, S. F. (1981): *The Theory of Linear Models and Multivariate Analysis*. John Wiley & Sons, New York.

ASHENFELTER, O., AND A. KRUEGER (1992): "Estimates of the Economic Return to Schooling from a New Sample of Twins," Discussion Paper 304, Industrial Relations Section, Princeton University.

ASHENFELTER, O., AND D. J. ZIMMERMAN (1993): "Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons, and Brothers," Discussion Paper 318, Industrial Relations Section, Princeton University.

BERNARD, A. B., AND S. N. DURLAUF (1995): "Convergence in International Output," *Journal of Applied Econometrics*, 10, 97–108.

BERNDT, E. R. (1991): *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley, Reading (MA).

BERNDT, E. R., AND D. O. WOOD (1975): "Technology, Prices and the Derived Demand for Energy," *Review of Economics and Statistics*, 53, 259–268.

BIRNBAUM, A. (1954): "Combining Independent Tests of Significance," *Journal of the American Statistical Association*, 49, 559–575.

BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*. Springer-Verlag, New York, second edn.

BURNSIDE, C., AND M. EICHENBAUM (1994): "Small Sample Properties of Generalized Method of Moments Based Wald Tests," Discussion Paper 155, National Bureau of Economic Research, Cambridge, MA.

DEJONG, D. N., J. C. NANKERVIS, N. E. SAVIN, AND C. H. WHITEMAN (1992): "The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors," *Journal of Econometrics*, 53, 323–343.

DUFOUR, J.-M. (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475–494.

——— (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

DUFOUR, J.-M., AND J. JASIAK (1995): "Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors," Discussion paper, C.R.D.E., Université de Montréal, 38 pages.

DUFOUR, J.-M., AND O. TORRÈS (1997): "Markovian Processes, Two-Sided Autoregressions and Exact Inference for Stationary and Nonstationary Autoregressive Processes," Discussion paper, C.R.D.E., Université de Montréal, 30 pages.

FOLKS, J. L. (1984): "Combination of Independent Tests," in *Handbook of Statistics 4: Nonparametric Methods*, ed. by P. R. Krishnaiah, and P. K. Sen, pp. 113–121. North-Holland, Amsterdam.

FOMBY, T. B., R. C. HILL, AND S. R. JOHNSON (1984): *Advanced Econometric Methods*. Springer-Verlag, New York.

HEDGES, L. V., AND I. OLKIN (1985): *Statistical Methods for Meta-Analysis*. Academic Press, San Diego, California.

HILLIER, G. H., AND M. L. KING (1987): "Linear Regressions with Correlated Errors: Bounds on Coefficient Estimates and $t-$values," in *Specification Analysis in the Linear Model: In Honour of Donald Cochrane*, ed. by M. L. King, and D. E. A. Giles, chap. 4, pp. 74–80. Routledge & Kegan Paul.

KENNEDY, P., AND D. SIMONS (1991): "Fighting the Teflon Factor: Comparing Classical and Bayesian Estimators for Autocorrelated Errors," *Journal of Econometrics*, 48, 15–27.

KING, M. L. (1996): "Hypothesis Testing in the Presence of Nuisance Parameters," *Journal of Statistical Planning and Inference*, 50, 103–120.

KIVIET, J. F. (1980): "Effects of ARMA Errors on Tests for Regression Coefficients: Comments on Vinod's Article; Improved and Additional Results," *Journal of the American Statistical Association*, 75, 353–358.

LASKAR, M. R., AND M. L. KING (1995): "Parameter Orthogonality and Likelihood Functions," in *Proceedings of the 1995 Econometrics Conference at Monash*, ed. by C. S. Forbes, P. Kofman, and T. R. L. Fry, pp. 253–289, Melbourne (Australia). Monash University.

LEHMANN, E. L. (1986): *Testing Statistical Hypotheses, 2nd edition*. John Wiley & Sons, New York.

MCCABE, B. P. M. (1988): "A Multiple Decision Theory Analysis of Structural Change in Regression," *Econometric Theory*, 4, 499–508.

MILLER, JR., R. G. (1981): *Simultaneous Statistical Inference*. Springer-Verlag, New York, second edn.

MIYAZAKI, S., AND W. E. GRIFFITHS (1984): "The Properties of Some Covariance Matrix Estimators in Linear Models with Autocorrelated Errors," *Economics Letters*, 14, 351–356.

NANKERVIS, J. C., AND N. E. SAVIN (1987): "Finite Sample Distributions of t and F Statistics in an AR(1) Model with an Exogenous Variable," *Econometric Theory*, 3, 387–408.

NELSON, C. R., R. STARTZ, AND E. ZIVOT (1996): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," Discussion paper, Department of Economics, University of Washington.

NEWEY, W. K., AND K. D. WEST (1987): "Hypothesis Testing with Efficient Method of Moments Estimators," *International Economic Review*, 28, 777–787.

PARK, R. E., AND B. M. MITCHELL (1980): "Estimating the Autocorrelated Error Model with Trended Data," *Journal of Econometrics*, 13, 185–201.

PHILLIPS, G. D. A., AND B. P. M. MCCABE (1988): "Some Applications of Basu's Independence Theorems in Econometrics," *Statistica Neerlendica*, 42, 37–46.

——— (1989): "A Sequential Approach to Testing Econometric Models," *Empirical Economics*, 14, 151–165.

ROY, S. N. (1953): "On a Heuristic Method of Test Construction and its Use in Multivariate Analysis," *Annals of Mathematical Statistics*, 24, 220–238.

SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches, and M. D. Intrilligator, chap. 14, pp. 827–879. North-Holland, Amsterdam.

SAVIN, N. E., AND A. WÜRTZ (1996): "The Effect of Nuisance Parameters on the Power of LM Tests in Logit and Probit Models," Discussion paper, Department of Economics, University of Iowa, Iowa City, IA.

SRIVASTAVA, V. K., AND D. E. GILES (1987): *Seemingly Unrelated Regression Equations Models. Estimation and inference*. Marcel Dekker, New York.

STAIGER, D., AND J. H. STOCK (1993): "Instrumental Variables Regression with Weak Instruments," Technical Working Paper 151, N.B.E.R., Cambridge, MA.

TIPPETT, L. H. (1931): *The Methods of Statistics*. Williams and Norgate, London.

TUNNICLIFFE WILSON, G. (1989): "On the Use of Marginal Likelihood in Time Series Model Estimation," *Journal of the Royal Statistical Society, Series B*, 51, 15–27.

VINOD, H. D. (1976): "Effects of ARMA Errors on the Significance Tests for Regression Coefficients," *Journal of the American Statistical Association*, 71, 929–933.

VINOD, H. D., AND A. ULLAH (1981): *Recent Advances in Regression Methods*. Marcel Dekker, New York.

WANG, J., AND E. ZIVOT (1996): "Inference on a Structural Parameters in Instrumental Variable Regression with Weakly Correlated Instruments," Discussion paper, Department of Economics, University of Washington.

WOOLDRIDGE, J. M. (1989): "A Computationally Simple Heteroskedasticity and Serial Correlation Robust Standard Error for the Linear Regression Model," *Economics Letters*, 31, 239–243.

ZINDE-WALSH, V., AND A. ULLAH (1987): "On the Robustness of Tests of Linear Restrictions in Regression Models with Elliptical Error Distributions," in *Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Volume III, Time Series and Econometric Modelling*, ed. by I. B. MacNeill, and G. J. Umphrey, pp. 235–251. D. Reidel, Dordrecht, The Netherlands.