Markovian processes, two-sided autoregressions and finite-sample inference for stationary and nonstationary autoregressive processes *

Jean-Marie Dufour †
Université de Montréal

Olivier Torrès ‡
Université de Lille 3

First version: September 1995
Revised: November 1997, February 1999, January 2000
This version: August 2000
Compiled: April 11, 2001, 10:17pm

A shorter version of this paper has been published in the *Journal of Econometrics*, 99 (2000), 255-289.

*The authors thank Frédéric Jouneau, Pierre Perron, Eugene Savin, an anonymous referee and the Editor Peter Robinson for several useful comments. An earlier version of this paper was presented at the Econometric Society European Meeting (Toulouse, August 1997). This work was supported by the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec).

† Centre de recherche et développement en économique (C.R.D.E.), Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Département de sciences économiques, Université de Montréal. Mailing address: C.R.D.E, Université de Montréal, C.P. 6128 succursale Centre Ville, Montréal, Québec, Canada H3C 3J7. TEL: (514) 343 2400; FAX: (514) 343 5831; e-mail: jean.marie.dufour@umontreal.ca .
Web page: http://www.fas.umontreal.ca/SCECO/Dufour .

‡ GREMARS, UFR de mathématiques, sciences économiques et sociales, Université de Lille 3, BP 149, 59 653 Villeneuve d'Ascq cedex, France. *e-mail*: torres@univ-lille3.fr

ABSTRACT

In this paper, we develop finite-sample inference procedures for stationary and nonstationary autoregressive (AR) models. The method is based on special properties of Markov processes and a split-sample technique. The results on Markovian processes (intercalary independence and truncation) only require the existence of conditional densities. They are proved for possibly nonstationary and/or non-Gaussian multivariate Markov processes. In the context of a linear regression model with AR(1) errors, we show how these results can be used to simplify the distributional properties of the model by conditioning a subset of the data on the remaining observations. This transformation leads to a new model which has the form of a two-sided autoregression to which standard classical linear regression inference techniques can be applied. We show how to derive tests and confidence sets for the mean and/or autoregressive parameters of the model. We also develop a test on the order of an autoregression. We show that a combination of subsample-based inferences can improve the performance of the procedure. An application to U.S. domestic investment data illustrates the method.

Keywords: time series; Markov process; autoregressive process; autocorrelation; dynamic model; distributed-lag model; two-sided autoregression; intercalary independence; exact test; finite-sample test; Ogawara-Hannan; investment.

JEL classification numbers: C22, C32, C12, C5, E2, E22.

RÉSUMÉ

Dans cet article, nous proposons des procédures d'inférence valides à distance finie pour des modèles autorégressifs (AR) stationnaires et non-stationnaires. La méthode suggérée est fondée sur des propriétés particulières des processus markoviens combinées à une technique de subdivision d'échantillon. Les résultats sur les processus de Markov (indépendance intercalaire, troncature) ne requièrent que l'existence de densités conditionnelles. Nous démontrons les propriétés requises pour des processus markoviens multivariés possiblement non-stationnaires et non-gaussiens. Pour le cas des modèles de régression linéaires avec erreurs autorégressives d'ordre un, nous montrons comment utiliser ces résultats afin de simplifier les propriétés distributionnelles du modèle en considérant la distribution conditionnelle d'une partie des observations étant donné le reste. Cette transformation conduit à un nouveau modèle qui a la forme d'une autorégression bilatérale à laquelle on peut appliquer les techniques usuelles d'analyse des modèles de régression linéaires. Nous montrons comment obtenir des tests et régions de confiance pour la moyenne et les paramètres autorégressifs du modèle. Nous proposons aussi un test pour l'ordre d'une autorégression. Nous montrons qu'une technique de combinaison de tests obtenus à partir de plusieurs sous-écantillons peut améliorer la performance de la procédure. Enfin la méthode est appliquée à un modèle de l'investissement aux États-Unis.

Mots-clés: séries chronologiques; processus de Markov; processus autorégressif; autocorrélation; modèle dynamique; modèle à retards échelonnés; autorégression bilatérale; indépendance intercalaire; test exact; Ogawara-Hannan; investissement.

Classification JEL: C22, C32, C12, C5, E2, E22.

Contents

List of Definitions, Propositions and Theorems

List of Tables

List of Figures

1. Introduction

The presence of nuisance parameters is a crucial problem when making inference on the parameters of a dynamic model. Typically test statistics have distributions which depend on those nuisance parameters so that they are difficult to interpret. A first approach to solve this difficulty consists in finding consistent estimates of the nuisance parameters which are then substituted for these parameters in the distribution of the statistic considered. However it is well known that such approximations can be arbitrarily bad; see Park and Mitchell (1980), Miyazaki and Griffiths (1984) and DeJong, Nankervis, Savin, and Whiteman (1992) for examples in the context of AR processes, Burnside and Eichenbaum (1994, 1996) about Wald-type tests based on GMM estimators, Dufour (1997) for a more general treatment of asymptotic approximation failures in the case of Wald statistics, Savin and Würtz (1996) for a similar discussion in the case of logit models, and Maasoumi (1992) for some general criticisms. Consequently, when hypothesis testing is the main objective, such a procedure offers no guarantee that the level constraint in the sense of Neyman-Pearson [see Lehmann (1986, p. 69) and Gouriéroux and Monfort (1989, p. 14)] be satisfied. This also makes comparisons between testing procedures difficult.

A second approach consists in using bounds which typically lead to conservative tests. Suppose the true critical value for our test statistic is unknown, but that it is possible to find bounds on this value, most importantly a bound yielding a critical region whose probability under the null hypothesis is not larger (but could be smaller) than the stated level of the test. For some examples of such methods in time series models, see Vinod (1976), Kiviet (1980) and Hillier and King (1987). In these cases, the bounds appear to increase without limit when the nuisance parameters approach some boundary (*e.g.*, the stationarity frontier in the case of ARMA processes) and/or with the sample size so they become useless [see Dufour and Torrès (1998)]. For regression models with AR(1) disturbances, procedures which do not display this unattractive feature were proposed in Dufour (1990); for further examples of such techniques, see also Dufour (1989), Dufour and Kiviet (1996, 1998), Campbell and Dufour (1997), Dufour, Hallin, and Mizera (1998), and Kiviet and Dufour (1997). However, these methods appear difficult to extend to more complex dynamic models such as AR(p) processes, $p \geq 2$.

In this paper, we propose an exact inference procedure for the parameters of Markov processes. It is based on extending old but little known results stated by Ogawara (1951) for univariate stationary Gaussian AR(p) process. Note Ogawara's article does not contain the proof of the result, and such a demonstration does not appear to be available elsewhere. The procedure has been extended by Hannan (1956) to multivariate, stationary, Gaussian processes admitting a VAR(1) representation. In the two latter references, procedures are developed for making inference on the autocorrelation parameters of pure AR processes. Hannan (1955a, 1955b) also showed this method can be applied to test a hypothesis on the coefficients of a linear regression model with stationary AR(1) errors.

In this paper, we generalize and improve these results in several directions. First, the initial results of Ogawara (1951) are extended to a larger class of processes, which includes multivariate, possibly non-normal, integrated or explosive processes. In particular, for general Markov processes of order p, it is shown that the variables separated by lags of p periods are mutually independent conditional on the intercalary observations (*intercalary independence*), a rather surprising property which is certainly of interest by itself. Second, we consider a more general class of models and hypotheses which includes as special cases all the models previously treated in the earlier literature [Ogawara (1951), Hannan (1955a, 1955b, 1956) and Krishnaiah and Murthy (1966)]. In particular, although this procedure was originally designed to make inference on the mean of a dynamic model, we show it is also suitable for inference on the nuisance parameters, such as autoregressive coefficients. Furthermore, we develop a procedure for constructing confidence regions. Third, we propose a way of resolving the information loss due to the application of the Ogawara-Hannan procedure. Fourth, we provide simulations results to evaluate the performance of our method.

Our procedure involves several steps. First, the sample is split into several subsets of observations. Next, on conditioning the original model on one of these subsamples, a transformed model having the form of a two-sided autoregression is obtained, *i.e.*, the dependent variable is regressed on its own leads and lags. This transformed model has simpler distributional properties and allows one to apply standard fixed regressor techniques. This is repeated for each subsample. Then a pooling method described in Dufour and Torrès (1998) is used to combine the results of subsample-based inferences and obtain a single answer based on the whole sample.

The procedures are quite easy to implement, for they only require applying standard test procedures (Student, Fisher, χ^2) to a transformed model. This means that there is no need to establish special critical points. The method is flexible enough to be easily adaptable to a wide variety of dynamic and econometric models. In particular, we show it can easily be adapted to various setups, such as: (1) integrated and explosive processes; (2) multidimensional processes (VAR models); (3) various models with more general dynamic structures

The paper is organized as follows. In Section 2, we motivate and expose the procedures developed in this paper in the context a simple AR(1) model with a deterministic trend. In particular, we demonstrate how to use a number of general results on Markov processes which are exposed in Section 3. In Section 4, we discuss in detail how these results can be applied to obtain finite sample inference procedures in the context of an AR(1) process. In Section 5, we consider a more general model by introducing a drift function in the AR(1) model. In particular, we explicitly show how one can obtain an exact test on the mean parameters *and* the autoregressive coefficients. We also derive an exact test for the order of an autoregression. In Section 6, we propose a method for improving the performance of Ogawara's procedure and we present simulation results. We conclude in Section 7. The proofs appear in the Appendix.

2. An introductory example

As an example of the procedures presented in this paper, consider the following AR(1) model

$$Y_t = m_t + \lambda Y_{t-1} + u_t\,, \quad \text{with } m_t = b_0 + b_1 t\,, \quad t = 1, 2, \ldots, T, \tag{2.1}$$

where $u_1, \ldots, u_T$ are independent and identically distributed $(i.i.d.)$ according to a $N(0, \sigma^2)$ distribution [henceforth denoted $u_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $t = 1, \ldots, T$].

Because such a model is recognized for describing well the dynamic behavior of many economic time series, a large part of the econometrics literature has been devoted to estimating it and making inferences on its parameters. One of the most investigated issue consists in testing the unit root hypothesis $H_0 : \lambda = 1$. Most of the (now) standard test procedures proposed in the literature use an OLS estimate $\hat{\lambda}_T$ of λ to form a statistic which is usually a normalized version of $\hat{\lambda}_T - 1$. The non-standard asymptotic distribution of this statistic is used to define a critical region for H_0. As mentioned in Section 1, the lack of reliability of such procedures is well documented. We propose here a simple approach which avoids the use of asymptotic approximations and provides tests and confidence regions having the stated level. Although the procedure presented in this paper goes much beyond this single issue, we illustrate it in the context of the simple AR(1) model (2.1) where we wish to test $H_0 : \lambda = 1$. For the sake of simplicity, we assume the sample size is odd, so that T may be written $T = 2n + 1$, for some strictly positive integer n.

The method may then be described as follows. The results of this paper entail the following properties: (1) conditionally on $\Phi_{\text{odd}} = (Y_1, Y_3, \ldots, Y_{2n+1})'$, the remaining observations $Y_2, Y_4, \ldots, Y_{2n}$ are mutually independent [see Theorem **3.1**]; (2) the conditional distribution of Y_{2t} given Φ_{odd} is identical to the conditional distribution of Y_{2t} conditional on (Y_{2t-1}, Y_{2t+1}) [see Theorem **3.2**]. In particular, for any $t = 1, 2, \ldots, n$, the mean of this conditional distribution is $\mathsf{E}(Y_{2t}|Y_{2t-1}, Y_{2t+1}) = \beta_1 m_{2t} + \beta_2 m_{2t+1} + \beta(Y_{2t-1} + Y_{2t+1})$, so that we may write

$$Y_t = \beta_1 m_t + \beta_2 m_{t+1} + \beta(Y_{t-1} + Y_{t+1}) + \eta_t$$

or, using the expression of m_t,

$$Y_t = a_0 + a_1 t + \beta(Y_{t-1} + Y_{t+1}) + \eta_t \tag{2.2}$$

for $t = 2, 4, \ldots, 2n$. The coefficients a_0, a_1 and β can be shown to be the following transformations of the initial parameters:

$$a_0 = b_0 \left(\frac{1-\lambda}{1+\lambda^2}\right) - b_1 \left(\frac{\lambda}{1+\lambda^2}\right), \quad a_1 = b_1 \left(\frac{1-\lambda}{1+\lambda^2}\right), \quad \beta = \frac{\lambda}{1+\lambda^2}. \tag{2.3}$$

Further, the error terms $\eta_2, \eta_4, \ldots, \eta_{2n}$, are $i.i.d.$ $N(0, \sigma_1^2)$, conditionally on Φ_{odd}. Now, it is interesting to note that (2.2) enjoys all the properties of a standard linear regression model with Gaussian $i.i.d.$ errors. Therefore, any linear hypothesis on its coefficients may be tested with usual procedures. In particular, $H_0 : \lambda = 1$ in (2.1) may be reformulated as a linear restriction on the parameters of (2.2), namely $H_0^{(a)} : (a_1 = 0 \text{ and } \beta = 1/2)$. A simple Fisher procedure gives a critical region with any required level α for $H_0^{(a)}$.

To illustrate the procedure, we propose the following numerical example. Following Dufour and Kiviet (1998), a model similar to (2.1) describes the dynamics of the (logarithm of) US gross private domestic investment in non-residential structures over the period 1952:I to 1969:IV [see Berndt (1991, p. 278) for a detailed description of the data]. The model is

$$Y_t = b_0 + b_1 t/100 + \lambda Y_{t-1} + m_t + u_t. \tag{2.4}$$

When the latter is estimated by OLS, we obtain: $\hat{\lambda} = 0.92143$, $\hat{b}_0 = 0.87197$, $\hat{b}_1 = 0.06986$ with unbiased error variance estimator $s^2 = 4.92300 \times 10^{-4}$. $\hat{\lambda}$ being close to 1, one may wish to test for the presence of a unit root in the AR polynomial. According to the discussion above, one would estimate the transformed model similar to (2.2)

$$Y_t = a_0 + a_1 t/100 + \beta(Y_{t+1} + Y_{t-1}) + \eta_t\,, \quad t = 2, 4, \ldots, 70\,, \tag{2.5}$$

where a_0, a_1 and β are given by (2.3), and test $H_0^{(a)}$: ($a_1 = 0$ and $\beta = 1/2$). Rewriting the null hypothesis under the form $H_0^{(a)} : R\delta - r = 0$ where

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \delta = (a_0,\, a_1,\, \beta)',\ r = (0,\, 1/2)',$$

the unit root hypothesis may then be tested at any level α by forming the statistic $F_1 = (R\widehat{\delta}_1 - r)'[R\widehat{V}(\widehat{\delta}_1)R']^{-1}(R\widehat{\delta}_1 - r)$ and using the critical region $F_1 \geq F(2, 31; 1 - \alpha)$. Here $\widehat{\delta}_1$ denote the vector of the OLS estimates of the components of $(a_0,\, a_1,\, \beta)'$ in (2.5), $\widehat{V}(\widehat{\delta}_1)$ is the usual ("unbiased") estimator of the variance-covariance matrix of $\widehat{\delta}_1$, and $F(2, 31; 1 - \alpha)$ is the $(1 - \alpha)$ quantile of the Fisher distribution with (2, 31) degrees of freedom. Computations yield the following results:

$$\widehat{\delta}_1 = \begin{pmatrix} 0.17965 \\ 0.00883 \\ 0.49191 \end{pmatrix},\ \widehat{V}(\widehat{\delta}_1) = \begin{pmatrix} 0.21226 & 0.01775 & -0.00968 \\ \cdot & 0.00163 & -0.00081 \\ \cdot & \cdot & 0.00044 \end{pmatrix},\ F_1 = 0.211179\,.$$

The p-value associated with F_1 is 0.8107842 so that $H_0^{(a)}$ is accepted at any level less than 81.07842%.

In our example, the transformed model (2.2) [or (2.5)] uses $Y_2, Y_4, \ldots, Y_{2n}$, as dependent variables and $Y_1, Y_3, \ldots, Y_{2n+1}$, as the conditioning variables. Obviously, the results we used for writing (2.2) may also be applied when $\Phi_{\text{even}} = (Y_2, Y_4, \ldots, Y_{2n})'$ are the conditioning variables. Another transformed model is then

$$Y_t = a_0 + a_1 t + \beta(Y_{t-1} + Y_{t+1}) + \nu_t\,, \quad t = 3, 5, \ldots, 2n - 1\,. \tag{2.6}$$

The error terms $\nu_3, \nu_5, , \ldots, \nu_{2n-1}$ are independent $N(0, \sigma_1^2)$ conditionally on Φ_{even} and (2.6) produces another critical region with level α for $H_0^{(a)}$. Back to the US gross private domestic investment in non-residential structures example, OLS estimation of

$$Y_t = a_0 + a_1 t/100 + \beta(Y_{t+1} + Y_{t-1}) + \nu_t\,, \quad t = 3, 5, \ldots, 71\,, \tag{2.7}$$

yields

$$\widehat{\delta}_2 = \begin{pmatrix} -0.49970 \\ -0.03342 \\ 0.52265 \end{pmatrix},\ \hat{V}(\widehat{\delta}_2) = \begin{pmatrix} 0.31227 & 0.02587 & -0.01423 \\ \cdot & 0.00234 & -0.00118 \\ \cdot & \cdot & 0.00065 \end{pmatrix},\ F_2 = 0.563799$$

with a p-value of 0.5747689 for F_2.

The tests based on F_1 and F_2 both agree on accepting $H_0^{(a)}$ at level 5% so that we would be tempted to accept the null at the same level of 5%. However, the decision rule which consists of accepting the null hypothesis when $m \geq 2$ tests accept it each at level α has a level which is larger to α. Such a method is the well known *induced test* procedure [see Savin (1984)] which combines several results from separate (although not necessarily independent) inferences. A sufficient condition ensuring it has level α is that each one of the m tests which are combined has level α/m [see Savin (1984) and Dufour and Torrès (1998) for further details]. In model (2.1), we accept $H_0 : \lambda = 1$ at level α whenever the tests based on (2.2) and (2.6) both accept (at level $\alpha/2$) the hypothesis $H_0^{(a)} : a_1 = 0$ and $\beta = 1/2$. In terms of p-values, this criterion can be reformulated as follows: we reject $H_0 : \lambda = 1$ at level α when the minimum of the p-values obtained from (2.2) and (2.6) is smaller than $\alpha/2$. When applied to the US investment data, it is easy to see that the null hypothesis of a unit root is accepted at level 5% for instance.

The procedure just described is very simple as it only requires standard tabulated distributions. Its steps can be summarized as follows. The initial model expresses the conditional mean of a Markov process, typically $Y_t = \mathsf{E}(Y_t|Y_{t-1}) + u_t$. By using properties of such processes, we are able to transform the initial model by first splitting the sample into two subsets of variables, and then writing the conditional mean of the variables in the first subset given some of the variables in the second subset. This leads to several transformed models such as $Y_t = \mathsf{E}(Y_t|Y_{t-1}, Y_{t+1}) + u_{it}$, $t \in J_i$, $i = 1, 2$, for instance, where J_1 and J_2 are collections of indices defining the two subsets of variables. The testing procedure exploits the fact that, due to some properties of Markov processes, these transformed models are standard linear regressions for which usual inference techniques apply.

In the next section, we present extensions of the theoretical results of Ogawara (1951) and Hannan (1956). These results establish the properties of Markov processes on which the inference procedures proposed rely.

3. Results on Markov processes

3.1. Notation

Let $\{X_t : t \in \mathbb{T}\}$ be a stochastic process on a probability space $(\Omega,\ \mathcal{F},\ \mathsf{P})$ with trajectories in $\mathbb{R}^m$, *i.e.* $X(\omega, t) \equiv \big(X_1(\omega, t), X_2(\omega, t),\ \ldots\ , X_m(\omega, t)\big)'$, $m \geq 1,\ t \in \mathbb{T}$, where $\mathbb{T}$ is an interval of the integers $\mathbb{Z}$. The symbol "$\equiv$" means "equal by definition". We assume that for all $t \in \mathbb{T}$, the probability law of X_t has density f_{X_t} with respect to the Lebesgue measure on $\mathcal{R}^m$ (the Borel σ-algebra of subsets of $\mathbb{R}^m$). For any random vector Φ of conditioning variables, we denote $f_{X_t|\Phi}(x|\phi)$ the conditional density of X_t given $\Phi = \phi$, evaluated at $x \in \mathbb{R}^m$.

It will be useful to introduce the following notations. Let p and n be two positive integers $(p \geq 1,\ n \geq 1)$. We consider the stochastic process $\{X_t : t \in \mathbb{T}\}$ and define:

$$B_{t,p} \equiv (X_{t-1}, X_{t-2},\ \ldots\ , X_{t-p}) = \big(X_{t-\tau} : 1 \leq \tau \leq p\big)\,,\ b_{t,p} \equiv \big(x_{t-\tau} : 1 \leq \tau \leq p\big)\,,$$

for $p+1 \leq t \leq (n+1)(p+1)$, and

$$A_{t,p} \equiv \big(B_{s(p+1),p} : t \leq s \leq n+1\big), a_{t,p} \equiv \big(b_{s(p+1),p} : t \leq s \leq n+1\big)\,, \text{ for } 1 \leq t \leq n+1\,,$$

where we assume the set $\mathbb{T}$ contains 1 and $(n+1)(p+1)-1$. In other words, $B_{t,p}$ denotes the set of p variables immediately preceding X_t, and $A_{t,p}$ is a collection of $B_{s,p}$ sets. We can give the following illustration of the way we split the variables in $\{X_t : t \in \mathbb{T}\}$:

$$\underbrace{X_{t(p+1)-p}, X_{t(p+1)-p+1}, \ldots, X_{t(p+1)-1}}_{B_{t(p+1),p}}, X_{t(p+1)}, \underbrace{X_{(t+1)(p+1)-p}, X_{(t+1)(p+1)-p+1}, \ldots, X_{(t+1)(p+1)-1}}_{B_{(t+1)(p+1),p}}$$

The following notation will provide a convenient shortcut: for any $t \in \mathbb{T}$, we set

$$X_t, \ldots_{(r)} \ldots, X_{t+kr} \equiv (X_t, X_{t+r}, X_{t+2r}, \ldots, X_{t+kr})$$

for any positive integers r and k such that $t + kr \in \mathbb{T}$. With this notation, we may now give the main definition.

Let $\{X_t : t \in \mathbb{T}\}$ be a stochastic process and p a positive integer. We say that $\{X_t : t \in \mathbb{T}\}$ is a Markov process of order p on $\mathbb{T}$ (or $\{X_t : t \in \mathbb{T}\}$ is Markovian of order p on $\mathbb{T}$) if it satisfies condition $M(p)$ defined as follows:

$$M(p) : f_{X_t|X_{t-k},\ldots_{(1)}\ldots,X_{t-1}} = f_{X_t|X_{t-p},\ldots_{(1)}\ldots,X_{t-1}}, \forall k \in \mathbb{N}, \forall t \in \mathbb{T}, \text{ with } t-k \in \mathbb{T} \text{ and } k \geq p. \tag{3.1}$$

Note that, for $\mathbb{T} = \mathbb{Z}$ and $p = 1$, we have the standard definition of a Markov process.

Let X and Y be two random vectors of dimension q and r, respectively. Whenever the relevant moments exist, the affine regression of X on Y is the random vector of size q, denoted $\mathsf{E}_{\mathsf{L}}(X|Y)$, whose i-th component is the orthogonal projection of X_i on the space spanned by the affine functions of Y (an affine function of Y is a linear combination of the elements of Y plus possibly a constant). If W is another random vector, $X \perp Y | W$ means that the residuals from the affine regressions of X and Y on W are uncorrelated, *i.e.* $\mathsf{E}\big[X - \mathsf{E}_{\mathsf{L}}(X|W)\big][Y - \mathsf{E}_{\mathsf{L}}(Y|W)]'\big] = 0$.

3.2. Intercalary independence and truncation properties

The procedures presented in Ogawara (1951) and Hannan (1955a, 1955b, 1956) exploit special properties of Markov processes (intercalary independence, truncation), which we now study in detail and generalize. The propositions below will be used to build a transformed model that satisfies the assumptions of the classical linear model on which standard inference techniques can be applied. Further they provide interesting insights on the structure of Markovian processes, and thus have interest by themselves. The intercalary independence property was apparently first given without proof by Ogawara (1951) for univariate Markov processes, while the truncation property was used implicitly by him (again without proof) in the context of univariate autoregressive stationary Gaussian processes. Ogawara (1951) notes that these results have been stated without proof in Linnik (1949). However no proof is given by Ogawara (1951) nor (apparently) by any other author. In this section, we demonstrate and extend these results to multivariate Markov processes of order p, allowing for non-stationarity and non-normality. In order to keep things as simple as possible, we shall assume that the time index set $\mathbb{T}$ contains the positive integers $\mathbb{N} : \mathbb{T} \supseteq \mathbb{N} = \{1, 2, \dots\}$.

The first result we state (*intercalary independence* for Markov processes of order p) is an extension of Theorems 1 and 2 of Ogawara (1951). The proofs are given in the Appendix.

Theorem 3.1 INTERCALARY INDEPENDENCE. *Let $\{X_t : t \in \mathbb{T}\}$ be a stochastic process satisfying condition $M(p)$, with $\mathbb{T} \supseteq \mathbb{N}$. Then for any positive integer n, $X_{p+1}, X_{2(p+1)}, \dots, X_{n(p+1)}$ are mutually independent, conditionally on $A_{1,p}$.*

Consider a dynamic model of the form

$$X_t = g_{1,t}(Z_t, X_\nu, \dots_{(1)} \dots, X_\tau) + \varepsilon_t\,, \quad t = 1, 2, \dots, T \equiv n(p+1) + p \tag{3.2}$$

where $\{X_t : t \in \overline{T}\}$ is an m-dimensional Markov process of order p on $\overline{T} \equiv \{1, 2, ..., n(p+1)+p\}$, and $1 \le \nu \le \tau \le t-1$. If $\{X_t : t \in \overline{T}\}$ is Markovian of order p, we have $t-1 \ge \tau \ge \nu \ge t-p$. Z_t is a vector of fixed exogenous variables, $\varepsilon_{i,t} \overset{i.i.d.}{\sim} N(0, \sigma_i^2)$, $i = 1, 2, \dots, m$, and $g_{1,t}$ is a deterministic function in $\mathbb{R}^m$. If we condition (3.2) on $A_{1,p}$, we obtain a conditional model

$$X_{t(p+1)} = g_{2,t}(Z_{t(p+1)}, A_{1,p}) + \eta_{t(p+1)}\,, \quad t = 1, 2, \dots, n\,, \tag{3.3}$$

in which, according to Theorem **3.1**, the endogenous variables are independent and

$$\mathsf{E}(\eta_{t(p+1)} | A_{1,p}) = 0\,, \quad t = 1, 2, \dots, n\,.$$

We achieve the independence at the expense of a larger number of variables in the conditional mean of $X_{t(p+1)}$ ($A_{1,p}$ instead of $X_\nu, \dots_{(1)} \dots, X_\tau$). However, by the following theorem, we can restrict ourselves to consider a more parsimonious model which is distributionally equivalent to (3.3).

Theorem 3.2 TRUNCATION PROPERTY. *Let $\{X_t : t \in \mathbb{T}\}$ be a stochastic process satisfying condition $M(p)$ with $\mathbb{T} \supseteq \mathbb{N}$. Then*

$$f_{X_{t(p+1)}|A_{1,p}} = f_{X_{t(p+1)}|B_{(t+1)(p+1),p}, B_{t(p+1),p}}$$

for any $t = 1, 2, \dots, n$, $\forall n \in \mathbb{N}$.

Note only the Markov property of the process is needed to establish these results. In particular, stationarity and/or normality are not required. The above theorem extends a result stated without proof by Ogawara (1951) in the context of a univariate, stationary, Gaussian Markov process of order p. For completeness, we state the latter as a corollary.

Corollary 3.3 INTERCALARY INDEPENDENCE FOR GAUSSIAN PROCESSES. *Let $\{X_t : t \in \mathbb{Z}\}$ be a (multidimensional) Gaussian Markov process of order p $(p \ge 1)$. Then Theorems* **3.1** *and* **3.2** *hold for $\{X_t : t \in \mathbb{Z}\}$.*

To see the latter corollary, we simply note that for any t, $f_{X_t|X_{-\infty}, \dots_{(1)} \dots, X_{t-1}} = f_{X_t|X_{t-p}, \dots_{(1)} \dots, X_{t-1}} \Rightarrow f_{X_t|X_{t-s}, \dots_{(1)} \dots, X_{t-1}} = f_{X_t|X_{t-p}, \dots_{(1)} \dots, X_{t-1}}$, for any $s \ge p$. Theorems **3.1** and **3.2** extend the results used by Ogawara to a larger class of processes. Theorem **3.2** shows that, if $\{X_t : t \in \mathbb{T}\}$ is Markovian of order p, variables other than those in $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$ do not appear in the conditional density of $X_{t(p+1)}$ given $A_{1,p}$. For example in (3.3), this suggests we can limit ourselves to consider a simpler equivalent model where $X_{t(p+1)}$ only depends on the adjacent variables $B_{(t+1)(p+1),p}$ and $B_{t(p+1),p}$, instead of the complete set $A_{1,p}$:

$$X_{t(p+1)} = g_t(Z_{t(p+1)}, B_{(t+1)(p+1),p}, B_{t(p+1),p}) + \eta_{t(p+1)}\,, \quad t = 1, 2, \ldots, n, \tag{3.4}$$

where the $X_{t(p+1)}$'s are (conditionally) independent. The function $g_t(\,\cdot\,)$ in (3.4) may be interpreted as the "best approximation" (projection) of $X_{t(p+1)}$ on the space spanned by (possibly nonlinear) functions of the variables in $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$. Corollary **3.5** below gives a sufficient condition for such "projections" to be invariant with respect to t, *e.g.*, to have $g_t(\,\cdot\,) = g(\,\cdot\,)$, for all $t = 1, 2, \ldots, n$. We first need to introduce the following definition.

Definition 3.4 CONDITIONAL STRICT STATIONARITY OF ORDER *p*. *Let* $\{X_t : t \in \mathbb{T}\}$ *be a stochastic process on* $\mathbb{T} \supseteq \mathbb{N}$. *We say that* $\{X_t : t \in \mathbb{T}\}$ *is conditionally strictly stationary of order* p [*denoted CSS*(p)] *if there exists a strictly positive integer p such that*

$$f_{X_t|X_{t-p},\cdots_{(1)}\cdots,X_{t-1}}(\,\cdot\,|\,\cdot\,) = f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}}(\,\cdot\,|\,\cdot\,)$$

for all $s \in \mathbb{T}$ *and* $t \in \mathbb{T}$ *such that* $s - p \in \mathbb{T}$ *and* $t - p \in \mathbb{T}$.

Corollary 3.5 TRUNCATION PROPERTY FOR CSS MARKOV PROCESSES. *Let* $\{X_t : t \in \mathbb{T}\}$ *be a* $CSS(p)$ *process satisfying condition* $M(p)$ *with* $\mathbb{T} \supseteq \mathbb{N}$. *Then*

$$f_{X_{t(p+1)}|B_{(t+1)(p+1),p},B_{t(p+1),p}}(\,\cdot\,|\,\cdot\,) = f_{X_{s(p+1)}|B_{(s+1)(p+1),p},B_{s(p+1),p}}(\,\cdot\,|\,\cdot\,)\,,\ \forall t \geq 1,\ \forall s \geq 1\,.$$

To see the latter property, we note that $B_{\tau,p} = X_{\tau-p}, \cdots_{(1)} \cdots, X_{\tau-1}$. Then writing the conditional density as

$$f_{X_{t(p+1)}|B_{(t+1)(p+1),p},B_{t(p+1),p}} = \frac{\prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|B_{\tau,p}}}{\int \prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|B_{\tau,p}} dx_{t(p+1)}}$$

[see the proof of Theorem **3.2**, equation (A.4) in the Appendix], the CSS(p) property of $\{X_t : t \in \mathbb{T}\}$ yields the result. The CSS(p) condition is entailed by strict stationarity. Furthermore, any random process that admits an AR(p) representation with $i.i.d.$ errors is Markovian of order p and CSS(p). This will be important for our purpose, since (3.4) can be rewritten as

$$X_{t(p+1)} = g(Z_{t(p+1)}, B_{(t+1)(p+1),p}, B_{t(p+1),p}) + \eta_{t(p+1)}\,, \quad t = 1, 2, \ldots, n, \tag{3.5}$$

where g no longer depends on t, which makes statistical inference much easier. Furthermore, for g affine, (3.5) is the classical linear regression model.

We now give two other propositions that will be especially useful when the process $\{X_t : t \in \mathbb{T}\}$ has an AR representation.

Theorem 3.6 TRUNCATION PROPERTY FOR AR PROCESSES. *Let* $\{X_t : t \in \mathbb{T}\}$ *be a Markov process of order* p *on* $\mathbb{T} \supseteq \mathbb{N}$. *Then for any integer* $q \geq p$, *we have* $f_{X_t|B_{t+1+q,q},B_{t,q}} = f_{X_t|B_{t+1+p,p},B_{t,p}}$, $\forall t \geq q+1$.

Corollary 3.7 PROJECTION TRUNCATION FOR AR PROCESSES. *Let* $\{X_t : t \in \mathbb{T}\}$ *be a Markov process of order* p *on* $\mathbb{T}$, *whose elements have finite second moments. Then, for any* q *such that* $q \geq p$, *we have* $\mathsf{E_L}(X_t|B_{t+1+q,q}, B_{t,q}) = \mathsf{E_L}(X_t|B_{t+1+p,p}, B_{t,p})$.

In the context of random processes which satisfy only second order properties analogous to those of Markov processes, results similar to intercalary independence and truncation hold. These are given in Theorems **3.8** and **3.9**.

Theorem 3.8 INTERCALARY ORTHOGONALITY. *Let* $\{X_t : t \in \mathbb{T}\}$ *be a random process with finite second moments such that*

$$X_t \perp (X_1, \ldots_{(1)} \ldots, X_{t-p-1}) | B_{t,p} .$$

Then

$$X_{t(p+1)} \perp X_{s(p+1)} | A_{1,p} \, , \; \forall t \geq 1 \, , \quad \forall s \geq 1 \, , \; t \neq s \, .$$

Theorem 3.9 INTERCALARY ORTHOGONAL TRUNCATION. *Let* $\{X_t : t \in \mathbb{N}\}$ *be a random process with finite second moments such that*

$$X_t \perp (X_1, \ldots_{(1)} \ldots, X_{t-p-1}) | B_{t,p} .$$

Then for all $t \geq 1$, *we have*

$$X_{t(p+1)} \perp B_{s(p+1),p} | \left[B_{(t+1)(p+1),p} \, , B_{t(p+1),p} \right] \, , \; \forall t \geq 1 \, , \; \forall s \geq 1 \, , \; s \neq t \text{ and } s \neq t+1 \, .$$

In the next section, we apply the above results to derive exact inference procedures for the parameters of the original model (3.4). We start with AR(1) processes. We then consider a Markov process of order 1 admitting a more general dynamic representation, which includes the classical linear regression model with AR(1) errors as a special case. In a subsequent section, we shall derive an exact inference procedure in the context of Markov processes of order p.

4. Exact inference for AR(1) models

In the previous section, we showed how to use Theorems **3.1** and **3.2** to derive a time invariant transformed model (3.5) from the initial model (3.2). If we wish to make inference on the parameters of (3.2) via those of (3.5), we must establish in a more explicit way the relationship between the two models. We can transform (3.2) into (3.5) by using two sorts of projections. Let $\{Y_t : t \in \overline{T}\}$ be a Markov process of order p on $\overline{T} \equiv \{1, 2, ..., n(p+1) + p\}$. The first kind of projection is suggested by the results of Section 3. It is the projection of $Y_{t(p+1)}$ on the space generated by the functions of the variables in $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$ (or the conditioning of $Y_{t(p+1)}$ upon $B_{t(p+1),p}$ and

$B_{(t+1)(p+1),p}$). Unless normality is assumed, this projection is likely to be nonlinear and difficult to establish. Moreover, if $\{Y_t : t \in \overline{T}\}$ is not CSS(p), we have no guarantee that this projection will be identical for all t.

The second type of projection is the affine regression of $Y_{t(p+1)}$ on $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$. The resulting model is linear by construction and the relation between the initial and transformed parameters is likely to be simple enough for inference. A sufficient condition (although not necessary, as we will see in the case of AR(1) processes) for this relation to be time invariant is weak stationarity of the process $\{Y_t : t \in \overline{T}\}$. However, our objective is to make exact inference and we will need to specify the probability distribution of $\{Y_t : t \in \overline{T}\}$. We will then assume that $\{Y_t : t \in \overline{T}\}$ is a Gaussian process. In that case, the two projections coincide.

In this section, we show how the results of the previous section can be applied to obtain exact tests and confidence regions on the parameters of an AR(1) model.

4.1. Model transformation

Suppose the scalar process $\{Y_t : t \in \overline{T}\}$, where $\overline{T} \equiv \{1, 2, \ldots, T\}$ and $T = 2n+1$ for some integer n, admits the following representation:

$$Y_t = \phi Y_{t-1} + \varepsilon_t\,, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)\,, \quad t \in \overline{T}\,, \tag{4.1}$$

with Y_0 given and $\phi \in \mathbb{R}$. If we assume the ε_t's are normally distributed, then $\{Y_t : t \in \overline{T}\}$ is a CSS(1) Markov process of order 1 on $\overline{T}$. We are now ready to apply the results of Section 3. The conditional distribution of Y_{2t} given (Y_{2t+1}, Y_{2t-1}) is normal, for all $t = 1, 2, \ldots, n$. Its mean is given by the affine regression of Y_{2t} on (Y_{2t+1}, Y_{2t-1}) and takes the form

$$\mathsf{E}_{\mathsf{L}}(Y_{2t}|Y_{2t+1}, Y_{2t-1}) = a + \beta_1 Y_{2t+1} + \beta_2 Y_{2t-1}\,, \quad t = 1, 2, \ldots, n\,.$$

The following theorem shows that if $|\phi| < 1$, then $\beta_1 = \beta_2 \equiv \beta$.

Theorem 4.1 REGRESSION SYMMETRY FOR WEAKLY STATIONARY PROCESSES. *Let $\{X_t : t \in \overline{T}\}$ be a weakly stationary univariate regular stochastic process. For all strictly positive integers p, the coefficients of X_{t+k} and X_{t-k} in the affine regression of X_t on $(B_{t+p+1,p}, B_{t,p})$ are equal, $1 \le k \le p$, for all $t \ge p+1$.*

Expressions for β and a are derived in the Appendix where it is shown that $\beta = \phi/(1+\phi^2)$ and $a = 0$. The variance of the residuals from the regression is $\sigma_\varepsilon^2/(1+\phi^2)$. These expressions are valid for any $\phi \in \mathbb{R}$. Starting from (4.1), the equivalent of the transformed model (3.5) is

$$Y_{2t} = \beta Y_{2t}^* + \eta_{2t}\,, \quad t = 1, 2, \ldots, n\,, \quad \boldsymbol{\eta}|(Y_{2t}^*\,,\, t = 1, 2, \ldots, n) \sim N\left[0,\, \frac{\sigma_\varepsilon^2}{1+\phi^2} I_n\right]\,, \tag{4.2}$$

where $Y_{2t}^* \equiv Y_{2t+1} + Y_{2t-1}\,,\ t = 1, 2, \ldots, n$, $\boldsymbol{\eta} = (\eta_2, \eta_4, \ldots \eta_{2n})'$ and I_n is the $n \times n$ identity matrix. (4.2) is a Gaussian linear regression model from which we can easily estimate β and make exact inference on it. In particular, using the usual critical region $W(\alpha) \equiv \{|t(\beta_0)| > t_{1-\alpha/2}(n -$

$1)\}$, with $t(\beta_0) \equiv (\hat{\beta} - \beta_0)/\hat{V}(\hat{\beta})^{1/2}$ where $\hat{\beta}$ and $V(\hat{\beta})$ are the usual OLS estimators of β and $V(\hat{\beta})$, we can test any hypothesis of the form $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$. This test has exact level α.

4.2. Exact tests on ϕ

Since $(1+\phi^2)\beta = \phi$, the relation between the "initial" parameter ϕ and the "transformed" parameter β is given by $\beta\phi^2 - \phi + \beta = 0$. In order to make inference on ϕ using model (4.2), we need to examine the roots of the polynomial $q(x) = \beta x^2 - x + \beta = 0$. Since ϕ is assumed to lie in $\mathbb{R}$, we discard complex roots, obtained with $|\beta| > 1/2$. If we also exclude the trivial case $\beta = 0$ which yields $\phi = 0$, the roots of $q(x)$ are $x_1 = \left(1 + \Delta_q^{1/2}\right)/2\beta$, $x_2 = \left(1 - \Delta_q^{1/2}\right)/2\beta$, where $\Delta_q = 1 - 4\beta^2$. Since $x_1 x_2 = 1$, we have $sign(x_1) = sign(x_2)$ and $x_i > 1 \iff x_j < 1,\ i, j = 1, 2,\ i \neq j$. Hence, with $\beta \neq 0$ and $|\beta| \leq 1/2$, two values of ϕ only are identified in (4.2). These values are 1 and -1 which are respectively equivalent to $\beta = 1/2$ and $\beta = -1/2$. In other words, given an *a priori* value for β, we can decide whether the process is integrated ($|\phi| = 1$), but, if not, we cannot distinguish a stationary process ($|\phi| < 1$) from an explosive process ($|\phi| > 1$). However this identification problem can be avoided by excluding explosive processes. This should not be a too restrictive practice if we admit that macroeconomic time series are usually integrated or stationary. The case where $\beta = 0$ corresponds to a white noise process, *i.e.* $\phi = 0$.

From the point of view of hypothesis testing, we have established the equivalence of each one of the null hypotheses $H_{01} : \phi = 0,\ H_{02} : \phi = 1$, and $H_{03} : \phi = -1$, with $H_{01}^* : \beta = 0,\ H_{02}^* : \beta = 1/2$, and $H_{03}^* : \beta = -1/2$, respectively. For these *a priori* values of ϕ, we have derived an exact test procedure. For other values of ϕ, we can still consider the test of $H_0^* : \beta - \beta_0 = 0$ which corresponds to the test of $H_0 : \phi \in \{x_0, x_0^{-1}\}$, where x_0 is the first root of $q(x)$, evaluated at $\beta = \beta_0$.

4.3. Exact confidence sets for ϕ

It is easy to build an exact confidence interval at level $1-\alpha$ for the parameter β in (4.2). Suppose the random variables c_1 and c_2 satisfy $c_1 \leq c_2$ with probability one and $\mathsf{P}(\{c_1 \leq \beta\} \cap \{\beta \leq c_2\}) = 1 - \alpha$. Since the events $\{c_1\phi^2 - \phi + c_1 \leq 0\} \cap \{c_2\phi^2 - \phi + c_2 \geq 0\}$ and $\{c_1 \leq \beta\} \cap \{\beta \leq c_2\}$ are identical, the set $\{\phi : c_1\phi^2 - \phi + c_1 \leq 0 \text{ and } c_2\phi^2 - \phi + c_2 \geq 0\}$ is a confidence region for ϕ with level $1 - \alpha$. To characterize this region in the space of the parameter ϕ, we need to find the roots of the polynomials $q_i(x) = c_i x^2 - x + c_i\ ,\ i = 1, 2$, when c_1 and c_2 are treated as constants. We can then distinguish the following cases.

1. If $|c_1| < 1/2$, the polynomial $q_1(x)$ has two distinct real roots denoted x_1 and x_2, and we can assume that $x_1 < x_2$. If $-1/2 < c_1 < 0$, then $q_1(x) \leq 0$ if and only if $x \in (-\infty, x_1] \cup [x_2, \infty)$. If $0 < c_1 < 1/2$, $q_1(x) \leq 0$ if and only if $x \in [x_1, x_2]$. If $c_1 = 0$, $q_1(x) \leq 0$ if and only if $x \in [0, \infty)$.

2. If $|c_1| = 1/2$, $q_1(x)$ has only one root. In this case, when $c_1 = 1/2$, $q_1(x) \leq 0$ if and only if $x = 1$, and when $c_1 = -1/2$, $q_1(x) \leq 0$ if and only if $x = -1$.

Table 1: Confidence regions for the autocorrelation parameter of an AR(1) process

c_1 \ c_2	$(-\infty,-1/2)$	$-1/2$	$(-1/2,0)$	0	$(0,1/2)$	$1/2$	$(1/2,\infty)$
$(-\infty,-1/2)$	$\varnothing$	$\{1\}$	$[z_1,z_2]$	$(-\infty,0]$	$(-\infty,z_1]\cup[z_2,\infty)$	$\mathbb{R}$	$\mathbb{R}$
$-1/2$		$\{1\}$	$[z_1,z_2]$	$(-\infty,0]$	$(-\infty,z_1]\cup[z_2,\infty)$	$\mathbb{R}$	$\mathbb{R}$
$(-1/2,0)$			$\big((-\infty,x_1]\cup[x_2,\infty)\big) \cap [z_1,z_2]$	$\big((-\infty,x_1]\cup[x_2,\infty)\big) \cap (-\infty,0]$	$\big((-\infty,x_1]\cup[x_2,\infty)\big) \cap \big((-\infty,z_1]\cup[z_2,\infty)\big)$	$(-\infty,x_1] \cup [x_2,\infty)$	$(-\infty,x_1] \cup [x_2,\infty)$
0				$\{0\}$	$[0,\infty) \cap \big((-\infty,z_1]\cup[z_2,\infty)\big)$	$[0,\infty)$	$[0,\infty)$
$(0,1/2)$					$[x_1,x_2] \cap \big((-\infty,z_1]\cup[z_2,\infty)\big)$	$[x_1,x_2]$	$[x_1,x_2]$
$1/2$						$\{-1\}$	$\{-1\}$
$(1/2,\infty)$							$\varnothing$

Note 1 _ $x_i,\ i = 1,\ 2,$ are the roots of $q_1(x),$ and $z_i,\ i = 1,\ 2,$ the roots of $q_2(x)$.
Note 2 _ Empty cells come from the inequality $c_1 \le c_2$.

3. If $|c_1| > 1/2$, $q_1(x)$ always takes the same sign on $\mathbb{R}$. If $c_1 < -1/2$, $q_1(x) \le 0$ for all $x \in \mathbb{R}$; if $c_1 > 1/2$, no real value of x satisfies $q_1(x) \le 0$.

Similarly, we determine the regions of $\mathbb{R}$ on which $q_2(x) \ge 0$. The different possibilities are summarized in Table 1.

5. Extension of the AR(1) model

In this section, we extend the procedures described in the previous section by considering more general processes. Let $\{Y_t : t \in \overline{T}\}$, where $\overline{T} \equiv \{1, 2, \ldots, T = n(p+1) + p\}$, be a random process with the following representation:

$$\begin{aligned} &\Lambda(B)Y_t = m_t + \varepsilon_t\,, \quad \varepsilon_t \overset{ind}{\sim} (0, \sigma_\varepsilon^2)\,, \quad t = 1, 2, \ldots, T\,, \\ &\Lambda(B) \equiv 1 - \textstyle\sum_{i=1}^{p} \lambda_i B^i\,, \quad \text{with } Y_0, Y_{-1}, \ldots, Y_{-p+1} \text{ fixed,} \end{aligned} \tag{5.1}$$

where B is the backward shift operator, m_t is an exogenous component, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$, and $\varepsilon_t \overset{ind}{\sim} (0, \sigma_\varepsilon^2)\,,\ t = 1, 2, \ldots, T$, means the ε_t's are independent with common mean 0 and variance σ_ε^2. Taking expectations on both sides, we obtain $\Lambda(B)M_t = m_t$, where $M_t \equiv \mathsf{E}(Y_t)$. Define the process $\{X_t \equiv Y_t - M_t : t \in \overline{T}\}$. Clearly, $\{X_t : t \in \overline{T}\}$ satisfies

$$\Lambda(B)X_t = \varepsilon_t\,, \quad \varepsilon_t \overset{ind}{\sim} (0, \sigma_\varepsilon^2)\,, \quad t = 1, 2, \ldots, T\,, \tag{5.2}$$

i.e. $\{X_t : t \in \overline{T}\}$ is a zero mean process which admits an AR(p) representation, where the disturbances ε_t , $t \in \overline{T}$, are independent with common mean zero and variance σ_ε^2. Consider now the case where $p = 1$. We have

$$Y_t = m_t + \lambda Y_{t-1} + \varepsilon_t\,, \quad \varepsilon_t \overset{ind}{\sim} (0, \sigma_\varepsilon^2)\,, \quad t = 1, 2, \ldots, T\,.$$

This representation includes as particular cases a wide range of models frequently used in econometrics. In particular: (1) if $m_t = 0$, $\forall t \in \overline{T}$, and $\lambda = 1$, we have the random walk model; (2) if $m_t = b_0$, $\forall t \in \overline{T}$, and $\lambda = 1$, we have a random walk with drift; (3) if $m_t = b(t) \equiv \sum_{i=1}^{r} b_i t^i$, $\forall t \in \overline{T}$, the process contains a deterministic polynomial trend. In what follows, we assume m_t has the form $m_t = \sum_{k=0}^{K} b_k Z_{k,t}$, where $Z_0, Z_1, \ldots, Z_K$ are exogenous variables.

Since $\{X_t : t \in \overline{T}\}$ has an AR(1) representation, application of the procedure described in Section 4 is straightforward. The projection is $\mathsf{E_L}[X_{2t}|(X_{2t+1}, X_{2t-1})] = \beta(X_{2t+1} + X_{2t-1})$ with $\beta = \lambda/\left(1 + \lambda^2\right)$ and we consider the following transformed model:

$$X_{2t} = \beta X_{2t}^* + \eta_{2t}\,, \quad t = 1, 2, \ldots, n\,, \quad \boldsymbol{\eta} \sim (\mathbf{0}, \sigma_\eta^2 I_n) \tag{5.3}$$

where $X_{2t}^* \equiv X_{2t+1} + X_{2t-1}$, $\sigma_\eta^2 \equiv \sigma_\varepsilon^2/\left(1 + \lambda^2\right)$ and $\boldsymbol{\eta} = (\eta_2, \eta_4, \ldots, \eta_{2n})'$. (5.3) can be written

$$Y_{2t} = M_{2t} - \beta(M_{2t+1} + M_{2t-1}) + \beta Y_{2t}^* + \eta_{2t}$$

with $Y_{2t}^* = Y_{2t+1} + Y_{2t-1}$. Now, with $m_t = M_t - \lambda M_{t-1}$ and $\beta = \lambda/\left(1 + \lambda^2\right)$, (5.3) becomes

$$Y_{2t} = \beta_1 m_{2t} + \beta_2 m_{2t+1} + \beta Y_{2t}^* + \eta_{2t}\,, \quad t = 1, 2, \ldots, n\,,$$

in which $\beta_1 \equiv 1/\left(1 + \lambda^2\right)$, $\beta_2 \equiv -\beta$. Finally, since $m_t = \sum_{k=0}^{K} b_k Z_{k,t}$, the transformed model is

$$Y_{2t} = \beta Y_{2t}^* + \sum_{k=0}^{K} \theta_{1k} Z_{k,2t} + \sum_{k=0}^{K} \theta_{2k} Z_{k,2t+1} + \eta_{2t} \tag{5.4}$$

where $\theta_{1k} \equiv b_k/\left(1 + \lambda^2\right)$ and $\theta_{2k} \equiv -\lambda b_k/\left(1 + \lambda^2\right)$. Using the matrix notation, (5.4) is equivalent to

$$\nu_t = \mathbf{Z}_t^{*\prime} \boldsymbol{\delta} + \eta_t^*\,, \quad t = 1, 2, \ldots, n\,, \tag{5.5}$$

with $\nu_t \equiv Y_{2t}$, $\mathbf{Z}_t^* \equiv (Z_{2t}, Z_{2t+1}, Y_{2t}^*)'$, $\boldsymbol{\delta} \equiv (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \beta)'$, $\boldsymbol{\theta}_i \equiv (\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,K})'$, $i = 1, 2$. If we assume that $\boldsymbol{\eta}$ is normally distributed, we can perform exact tests on λ and/or $b_k, k = 0, 1, \ldots, K$. This is done in the next section.

5.1. Exact confidence sets and tests on b_k

As we showed, the parameters of (5.5) must satisfy $\theta_{2k} = -b_k \beta, k = 0, 1, \ldots, K$. The hypothesis $b_k = \bar{b}_0$ is therefore equivalent to $\theta_{2k} + \bar{b}_0 \beta = 0$ which can be tested in (5.5) by a standard F procedure. Furthermore it is well known that the set of all values $\bar{b}_0$ such that the hypothesis $H_0 : \theta_{2k} + \bar{b}_0 \beta = 0$ is not rejected at level α forms a confidence region for b_k at level $1 - \alpha$.

Using the same relation between the transformed parameters θ_{2k} and β and the initial parameters $b_k, k = 0, 1, \ldots, K$, any linear hypothesis of the form $R\mathbf{b} - \mathbf{r} = \mathbf{0}$, where R is a known $q \times (K+1)$ matrix with rank q, $\mathbf{r}$ is a known $q \times 1$ vector and $\mathbf{b} = (b_0, b_1, \ldots, b_K)'$, can be tested at level α. To see how to exploit the relation between the two sets of parameters, note that

$$R\mathbf{b} - \mathbf{r} = \mathbf{0} \iff R\boldsymbol{\theta}_2 + \mathbf{r}\beta = \mathbf{0} \iff R^*\boldsymbol{\delta} = \mathbf{0}$$

where $R^* \equiv \big(\mathbf{0}, R,\, \mathbf{r}\big)$ so that a test of $R\mathbf{b} - \mathbf{r} = \mathbf{0}$ is equivalent to a test of $R^*\boldsymbol{\delta} = \mathbf{0}$. Again, this is a hypothesis on the parameters of (5.5) which can be tested with the usual F procedure.

5.2. Exact tests on λ

The components of $\boldsymbol{\delta}$ in (5.5) must satisfy $\theta_{2k} = -\theta_{1k}\lambda, k = 0, 1, \ldots, K$ and $\beta = \lambda/\big(1 + \lambda^2\big)$. From these relations, we see that a test of $\lambda = \lambda_0$ can be performed by testing the joint hypothesis: $\theta_{2k} + \lambda_0\theta_{1k} = 0,\ k = 0, 1, \ldots, K,$ and $\beta = \lambda_0/(1 + {\lambda_0}^2)$. Using matrix notation we can easily write this set of restrictions as a linear hypothesis on the parameters of (5.5), *i.e.*, $\tilde{R}\boldsymbol{\delta} = \mathbf{r}_0$ with

$$\tilde{R} \equiv \begin{pmatrix} \lambda_0 I_{K+1} & I_{K+1} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & 1 \end{pmatrix}, \quad \mathbf{r}_0 \equiv \begin{pmatrix} \mathbf{0} \\ \lambda_0/(1 + {\lambda_0}^2) \end{pmatrix}.$$

Unlike for the pure AR(1) process of Section 4, we are now able to obtain a test for any *a priori* value λ_0 of the autocorrelation parameter λ.

5.3. Exact confidence sets for λ

In Section 4.3 we showed how to build an exact confidence region for λ at level $1 - \alpha$. This confidence region, denoted $C_{K+1}(y, \alpha)$, satisfies $\mathsf{P}\big[\{y : C_{K+1}(y, \alpha_1) \ni \lambda\}\big] = 1 - \alpha_1$ or $\mathsf{P}\big[A_{K+1}(\alpha_1)\big] = 1 - \alpha_1$, where $A_{K+1}(\alpha_1) \equiv \{y : C_{K+1}(y, \alpha_1) \ni \lambda\},\ \forall \alpha_1 \in (0, 1)$.

Similarly, we can also use the relation $\theta_{2k} + \lambda\theta_{1k} = 0, k = 0, 1, \ldots, K$, to derive an exact test of $H_0 : \lambda = \lambda_0$. This hypothesis entails $H_{0,k}(\lambda_0) : \mathbf{a}_k(\lambda_0)'\boldsymbol{\delta} = 0$, where $\mathbf{a}_k(x) \equiv \Big(x\boldsymbol{\iota}'_{k+1},\ \boldsymbol{\iota}'_{k+1},\ 0\Big)$, $\boldsymbol{\iota}_l$ being the l-th vector of the canonical basis of $\mathbb{R}^{K+1}$, $x \in \mathbb{R}$. The set $C_k(y, \alpha_1)$ of all values λ_0 of λ such that $H_{0,k}(\lambda_0)$ is not rejected at level α_1 is a $1 - \alpha_1$ confidence region for λ. Therefore $\mathsf{P}\big[A_k(\alpha_1)\big] = 1 - \alpha_1$, where $A_k(\alpha_1) \equiv \{y : C_k(y, \alpha_1) \ni \lambda\}$. Since this condition holds for any $k = 0, 1, \ldots K$, we can combine these regions to form a single confidence region for λ which has level $1 - \alpha$. Clearly, we have

$$\mathsf{P}\left[\bigcap_{k=0}^{K+1} A_k(\alpha_1)\right] = 1 - \mathsf{P}\left[\bigcup_{k=0}^{K+1} \overline{A_k(\alpha_1)}\right]$$

where $\overline{A_k(\alpha_1)}$ denotes the set of all y which are not in $A_k(\alpha_1)$, and

$$\mathsf{P}\left[\bigcup_{k=0}^{K+1} \overline{A_k(\alpha_1)}\right] \le \sum_{k=0}^{K+1} \mathsf{P}\left[\overline{A_k(\alpha_1)}\right] = (K + 2)\alpha_1\,,$$

hence

$$\mathsf{P}\left[\bigcap_{k=0}^{K+1} A_k(\alpha_1)\right] \geq 1-(K+2)\alpha_1$$

and choosing α_1 such that $\alpha_1 = \alpha/(K+2)$, we get

$$\mathsf{P}\left[\bigcap_{k=0}^{K+1} A_k(\alpha_1)\right] \geq 1-\alpha\,.$$

But $\bigcap_{k=0}^{K+1} A_k(\alpha_1) = \left\{y : \bigcap_{k=0}^{K+1} C_k(y,\alpha_1) \ni \lambda\right\}$. This shows that $C(y,\alpha) \equiv \bigcap_{k=0}^{K+1} C_k\left(\frac{\alpha}{K+2}\right)$ is a $1-\alpha$ confidence region for λ.

5.4. Exact tests of joint hypotheses

It is also possible to use (5.5) to derive an exact test of a linear hypothesis on the vector $\left(\lambda, \mathbf{b}^{(m)\prime}\right)'$, where $\mathbf{b}^{(m)}$ is an $m \times 1$ subvector of $\mathbf{b}$. Consider the null hypothesis

$$H_0 : \lambda = \lambda_0 \text{ and } R\mathbf{b}^{(m)} - \mathbf{r} = \mathbf{0}$$

where R is a known $q \times m$ matrix with rank q, $\mathbf{r}$ is a known $q \times 1$ vector and $\mathbf{b}^{(m)} = (b_{k_1}, b_{k_2}, \ldots\, b_{k_m})'$. The following equivalences hold

$$\left.\begin{array}{l} \lambda = \lambda_0 \\ R\mathbf{b}^{(m)} - \mathbf{r} = \mathbf{0} \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} \theta_{2k} + \lambda_0\theta_{1k} = 0\,,\ k \in \mathcal{K}_m \\ R\mathbf{b}^{(m)}\beta - \mathbf{r}\beta = \mathbf{0} \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} I_m\boldsymbol{\theta}_2^{(m)} + \lambda_0 I_m \boldsymbol{\theta}_1^{(m)} = \mathbf{0} \\ R\boldsymbol{\theta}_2^{(m)} + \mathbf{r}\beta = \mathbf{0} \end{array}\right.$$

where $\mathcal{K}_m \equiv \{k_1, k_2, \ldots\, k_m\}$, $\boldsymbol{\theta}_i^{(m)} \equiv (\theta_{ik_1}, \theta_{ik_2}, \ldots\, \theta_{ik_m})'$, $i = 1, 2$. Defining

$$Q \equiv \begin{pmatrix} I_m & \lambda_0 I_m & \mathbf{0} \\ 0 & R & \mathbf{r} \end{pmatrix}, \quad \boldsymbol{\delta}^{(m)} \equiv \begin{pmatrix} \boldsymbol{\theta}_1^{(m)} \\ \boldsymbol{\theta}_2^{(m)} \\ \beta \end{pmatrix},$$

we see that H_0 is equivalent to $Q\boldsymbol{\delta}^{(m)} = \mathbf{0}$. Finally H_0 appears as a linear hypothesis on the parameters of (5.5): $H_0 : \hat{R}\boldsymbol{\delta}^* = \mathbf{0}$ with $\hat{R} \equiv (Q\,,\,0)\,,\ \boldsymbol{\delta}^* \equiv \left(\boldsymbol{\delta}^{(m)\prime},\ \boldsymbol{\delta}^{)m(\prime}\right)',\ \boldsymbol{\delta}^{)m(} \equiv (\theta_{1,k}\ \theta_{2k}\,,\ k \notin \mathcal{K}_m)'$. Once again, the standard Fisher procedure solves the problem.

5.5. Linear regression models with AR(1) errors

We now show that model (5.1) with $p = 1$ includes as an important special case the linear regression model with AR(1) errors. This model is given by

$$Y_t = m_t + u_t\,, \quad u_t = \phi u_{t-1} + \varepsilon_t\,, \quad t = 1, 2, \ldots, T,$$

with $\varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$ and u_0 given. An alternative form of this model is

$$Y_t = m_t + \phi u_{t-1} + \varepsilon_t\,, \quad t = 1, 2, \ldots, T.$$

Since $u_t = Y_t - m_t\,, \quad t = 1, 2, \ldots, T$, we have

$$Y_t = m_t^* + \phi Y_{t-1} + \varepsilon_t\,, \quad t = 2, 3, \ldots, T, \tag{5.6}$$

where $m_t^* \equiv m_t - \phi m_{t-1}$. It is now clear that this model is a special case of (5.1). The procedures developed in the previous sections therefore apply to (5.6). In particular, exact inference in integrated AR(1) models is available.

5.6. A test on the order of an autoregression

We now turn to another kind of inference problem. We are no longer interested in inference on the components of the mean vector or autocovariance matrix, but rather on the order of the autoregression in AR(p) models. There is a situation in which Theorem **3.6** and its corollary are of special interest. Consider $\{X_t : t \in \mathbb{T}\}$, a stochastic process for which we know that one of the following representations is true:

$$\Phi(B)X_t = \varepsilon_t\,, \quad \text{where } \Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_{p_1} z^{p_1}\,,$$

$$\Psi(B)X_t = \nu_t\,, \quad \text{where } \Psi(z) = 1 - \psi_1 z - \psi_2 z^2 - \cdots - \psi_{p_2} z^{p_2}\,,$$

where ε_t and ν_t are both Gaussian white noises and $p_1 \neq p_2$ (we set $p_1 < p_2$). Suppose we wish to test $H_0 : \{X_t : t \in \mathbb{T}\} \sim$ AR(p_1) against $H_1 : \{X_t : t \in \mathbb{T}\} \sim$ AR(p_2). If H_0 is true, then $\{X_t : t \in \mathbb{T}\}$ is Markovian of order p_1, and we know from Corollary **3.7** that the coefficient of X_τ in the affine regression of X_t on p_2 leads and p_2 lags will be zero for any τ such that $|\tau - t| = p_1 + 1, \ldots, p_2$. Since the affine regression is a classical linear regression model, standard inference procedures apply. From the exposition of the procedures, it is clear that splitting the sample entails an information loss. We may then suspect the tests to lack power. We investigate this issue in the next section.

6. Combination of tests

One of the purposes of this paper is to improve the Ogawara-Hannan testing procedure. In the previous sections, we showed that Ogawara's results can be extended to a much wider class of processes than those considered in Ogawara (1951) and Hannan (1955a, 1955b, 1956). We also showed one can use these results to obtain finite sample inference procedures for a wide variety of econometric models. However, when we apply those, we are led to leave one half of the sample apart, at least. In this section, we discuss methods that allow one to make use of the full sample. We also present simulation results which show our method performs better than that of Ogawara and Hannan.

6.1. Theoretical results

Consider a statistical model characterized by a family of probability laws, parameterized by θ : $\mathcal{P} = \{\mathsf{P}_\theta, \theta \in \Theta\}$. Suppose we wish to test $H_0 : \mathsf{P} \in \mathcal{P}_0$ against $H_1 : \mathsf{P} \in \mathcal{P} \setminus \mathcal{P}_0$. If the model is identified, which will be assumed, this amounts to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\theta \in \Theta_0 \iff \mathsf{P}_\theta \in \mathcal{P}_0$. Assume we have m statistics $T_i, i \in J \equiv \{1, 2, \ldots, m\}$, that can be used for testing H_0. Further assume that under H_0, $\mathsf{P}_\theta\left[\{y : T_i(y) > t\}\right]$ is known, for all $t \in \mathbb{R}, i \in J$. The relation between these statistics is typically unknown or difficult to establish. We wish to combine the information provided by each of those m statistics on the true probability distribution of the model.

A natural way of doing this is to proceed as follows. Using the m statistics T_i, we build m critical regions $W_i(\alpha_i) \equiv T_i^{-1}\big((t_i(\alpha_i), \infty)\big)$, where the $t_i(\alpha_i)$'s are chosen so that $\mathsf{P}_\theta\left[W_i(\alpha_i)\right] = \alpha_i$. We reject H_0 with a test based an the i-th statistic if y is in $W_i(\alpha_i)$, or equivalently if the observed value t_i of T_i is in $(t_i(\alpha_i), \infty)$. Consider the decision rule which consists in rejecting H_0 when it has been rejected by at least one of the tests based on a T_i statistic. The rejection region corresponding to this decision rule is $\bigcup_{i \in J} W_i(\alpha_i)$. This test is called an *induced* test of H_0 [see Savin (1984)]. Its size is impossible or difficult to determine since the distribution of the vector $(T_1, T_2, \ldots, T_m)'$ is generally unknown or intractable. It is however possible to choose the α_i's so that the induced test has level α. We have

$$\mathsf{P}_\theta\Big[\bigcup_{i \in J} W_i(\alpha_i)\Big] \leq \sum_{i \in J} \mathsf{P}_\theta\left[W_i(\alpha_i)\right] \leq \sum_{i \in J} \alpha_i$$

and so we only need to choose the α_i's so that they sum to α. To our knowledge, there is no criterion for choosing the α_i's in a way that could be optimal in some sense. Without such a rule, we will set $\alpha_i = \alpha_0 = \alpha/m$ for all $i \in J$.

It is difficult to compare the power of an α level test based on a single statistic T_i with that of a α level induced test. The latter uses the information provided by the whole sample, but is obtained by combining m tests of level α/m only, whereas the former has level $\alpha > \alpha/m$, but only exploits a subsample. In other words, with respect to power, what can be gained from the larger sample size on which is based the induced test could be lost because the levels of the individual tests combined are lower (*e.g.*, α/m instead of α). We now present simulations that reveal the power increase associated with combining tests.

6.2. Power simulations for AR(1) processes

Let $\{Y_t : t \in \overline{T}\}$, where $\overline{T} = \{1, 2, \ldots, T\}$, a random process admitting an AR(1) representation

$$Y_t = \lambda Y_{t-1} + \varepsilon_t\,, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(\mathbf{0}, I_T)\,, \quad t \in \overline{T}\,, \tag{6.1}$$

with Y_0 given. For the sake of simplicity, we assume that T is even with $T = 2n$. Since $\{Y_t : t \in \overline{T}\}$ is a Markov process of order 1, the results of Section 2 apply and we know that: (1) $Y_{2t}\,,\ t = 1, 2, \ldots, n-1$, are mutually independent, conditionally to $(Y_1, Y_3, \ldots, Y_{2n-1})$; (2) $Y_{2t+1}\,,\ t = 1, 2, \ldots, n-1$, are mutually independent, conditionally to $(Y_2, Y_4, \ldots, Y_{2n})$. If we define two subsets of $\mathbf{T}$, $J_1 = \{2, 4, \ldots, 2n-2\}$ and $J_2 = \{3, 5, \ldots, 2n-1\}$, we obtain two transformed models of type (4.2):

$$Y_t = \frac{\lambda}{1+\lambda^2}(Y_{t+1} + Y_{t-1}) + \eta_{it}\,, \quad t \in J_i,\,, \quad \boldsymbol{\eta}_i \sim N(\mathbf{0}, \sigma_\eta^2 I_{n_i}) \tag{6.2}$$

where $\boldsymbol{\eta}_i \equiv (\eta_{it}, t \in J_i)'$, $i = 1, 2$, and $n_1 = n-1,\ n_2 = n$. In each of these two models it is possible to test $H_0 : \lambda = \lambda_0$ at level $\alpha/2$, as shown in Section 4. We combine these two tests according to the procedure described in 6.1.

In our simulations, we proceed as follows. We consider $\lambda_0 = 0,\ 0.5,\ 1$ and $T = 100$. For a set $V(\lambda_0)$ of S values of λ in a neighborhood of λ_0, we simulate a sample of size T from the AR(1) process (6.1). Then we form the two subsamples $(y_t : t \in J_i),\ i = 1, 2$, from which we test $H_0(\beta_0) : \beta = \beta_0$ in the transformed model (6.2), with $\beta_0 = \lambda_0/\big(1+\lambda_0^2\big)$. For purposes of comparison, these tests are performed at levels 5% and 2.5%. The two 2.5% level tests are combined to give a 5% level induced test. These computations are repeated 1000 times, for each value of λ in $V(\lambda_0)$. The number of rejections of $H_0(\beta_0)$ gives an estimation of the performance of the test. Results are shown in Figures 1 to 6 where the solid line (—) represents the 5% induced test and the dashed lines $(--)$ and $(-\cdot-)$ represent the 5% subsample-based tests.

Figures 1 to 3 display the estimated power function for $\lambda = 0,\ 0.5,\ 1$ respectively, whereas the last three (Figures 4 to 6) show the differences of rejection frequencies for $\lambda = 0,\ 0.5,\ 1$ respectively. More precisely these differences are computed as: *Number of rejections of* $H_0(\beta_0)$ *with the induced test – Number of rejections of* H_0 *with the test based on subsample* $(y_t : t \in J_i),:$ $i = 1, 2$.

Apart from the case where $\lambda_0 = 0$, the combination method leads to a power increase, relative to a 5% level test based on a subsample. When $\lambda_0 = 0$, the power loss from combining is about 8% at most, which appears small. For $\lambda_0 \neq 0$, it is important to note that the values λ and λ^{-1} yield the same value of β in (6.2). For example $\lambda = 0.5$ and $\lambda = 2.0$ both yield $\beta = 0.4$. In other words, unless we impose restrictions such as $|\lambda| \leq 1$ or $|\lambda| \geq 1$, the value of β does not completely identify λ. This explains the presence of the mirror peak at $\lambda = 2$ [Figure 2].

7. Conclusion

In this paper we proposed a method allowing one to make finite-sample inference on the parameters of autoregressive models. This was made possible by special properties of Markov processes. The conditions under which such results hold are very mild since their demonstrations only require the existence of density functions. In particular, they are general enough to be applied to multivariate and possibly non stationary and/or non-Gaussian processes. However, with the addition of conditional stationarity and normality assumptions, we were able to use these properties to derive exact tests and confidence regions on the parameters of AR(1) models. In order to apply our procedure, it is necessary to split the sample as two subsets of observations. Our simulations in the case of a pure AR(1) model showed that a combination of separate inference results based on these subsamples generally leads to an improvement in the performance of the procedure.

Our method displays several attractive features. First, since it is exact, it controls the probability

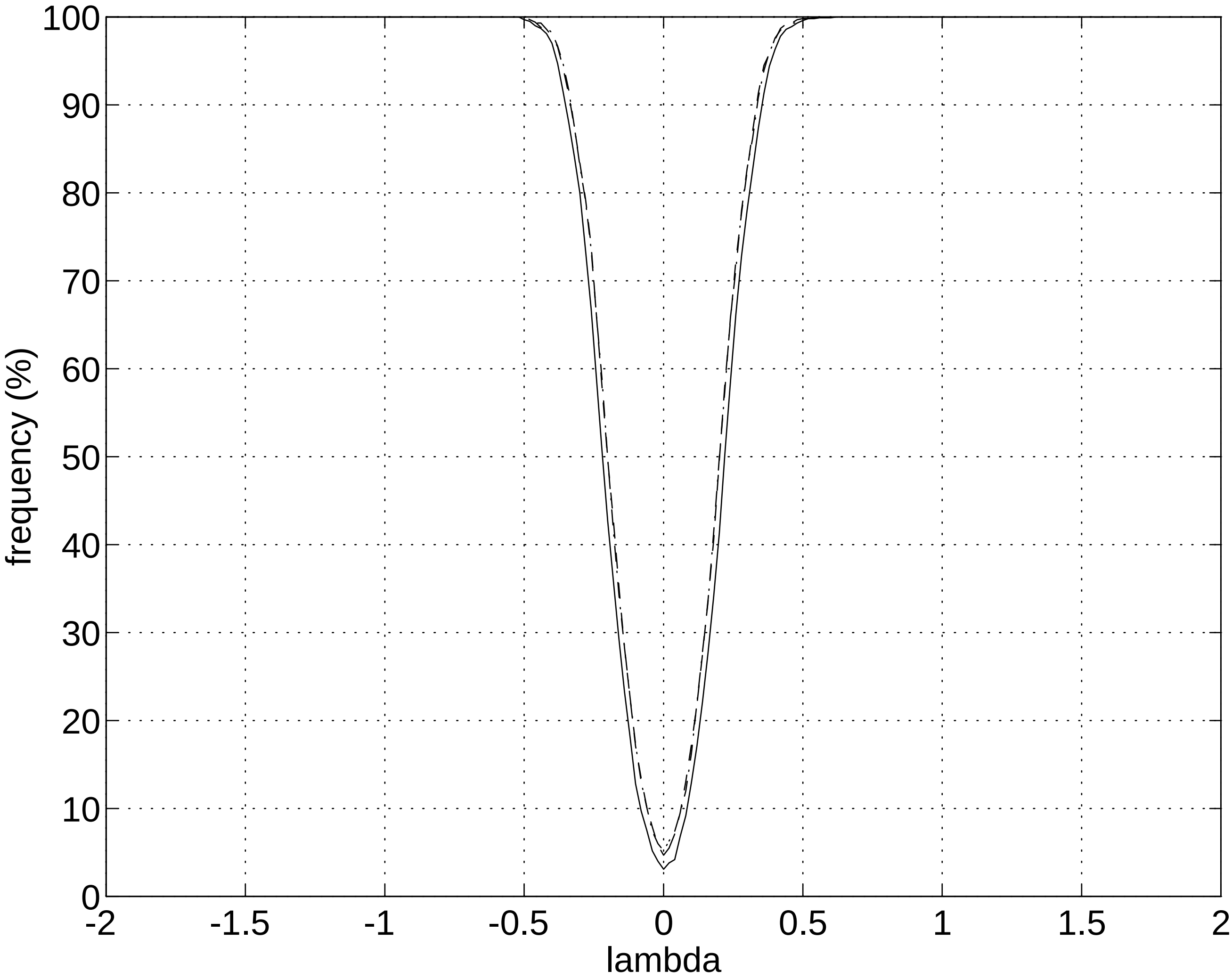

Figure 1. Rejection frequencies of $H_0 : \lambda = 0$

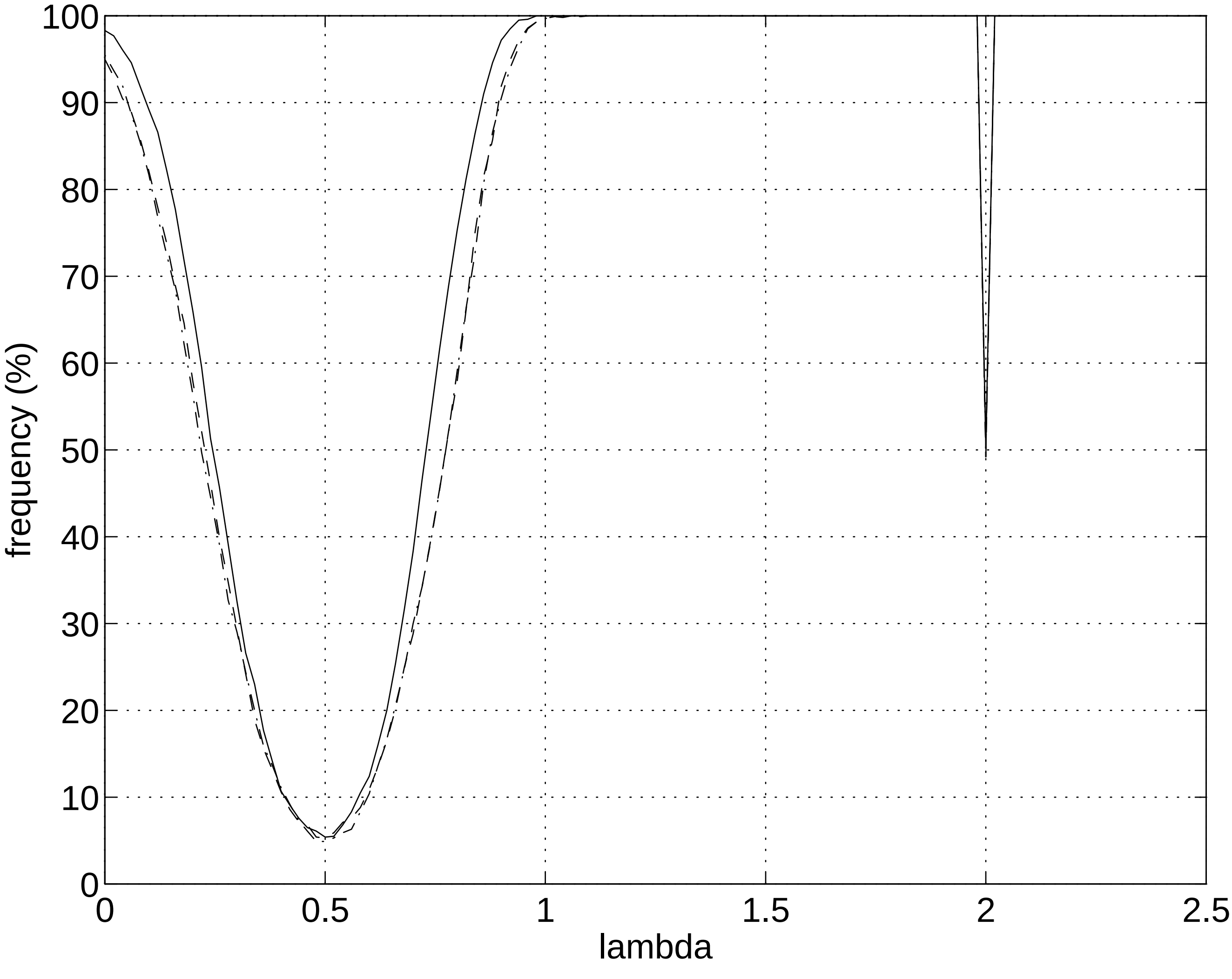

Figure 2. Rejection frequencies of $H_0 : \lambda = 0.5$

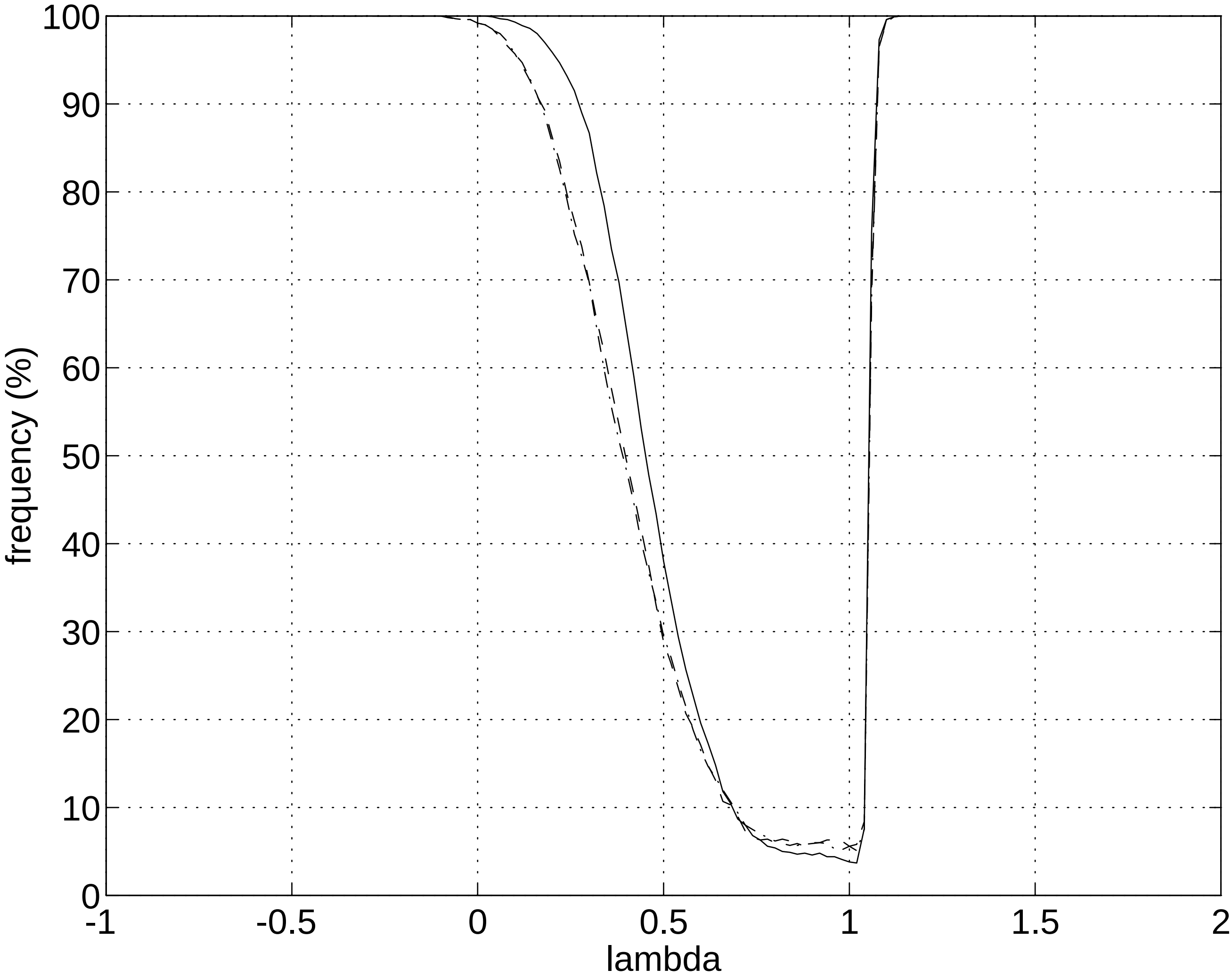

Figure 3. Rejection frequencies of $H_0 : \lambda = 1$

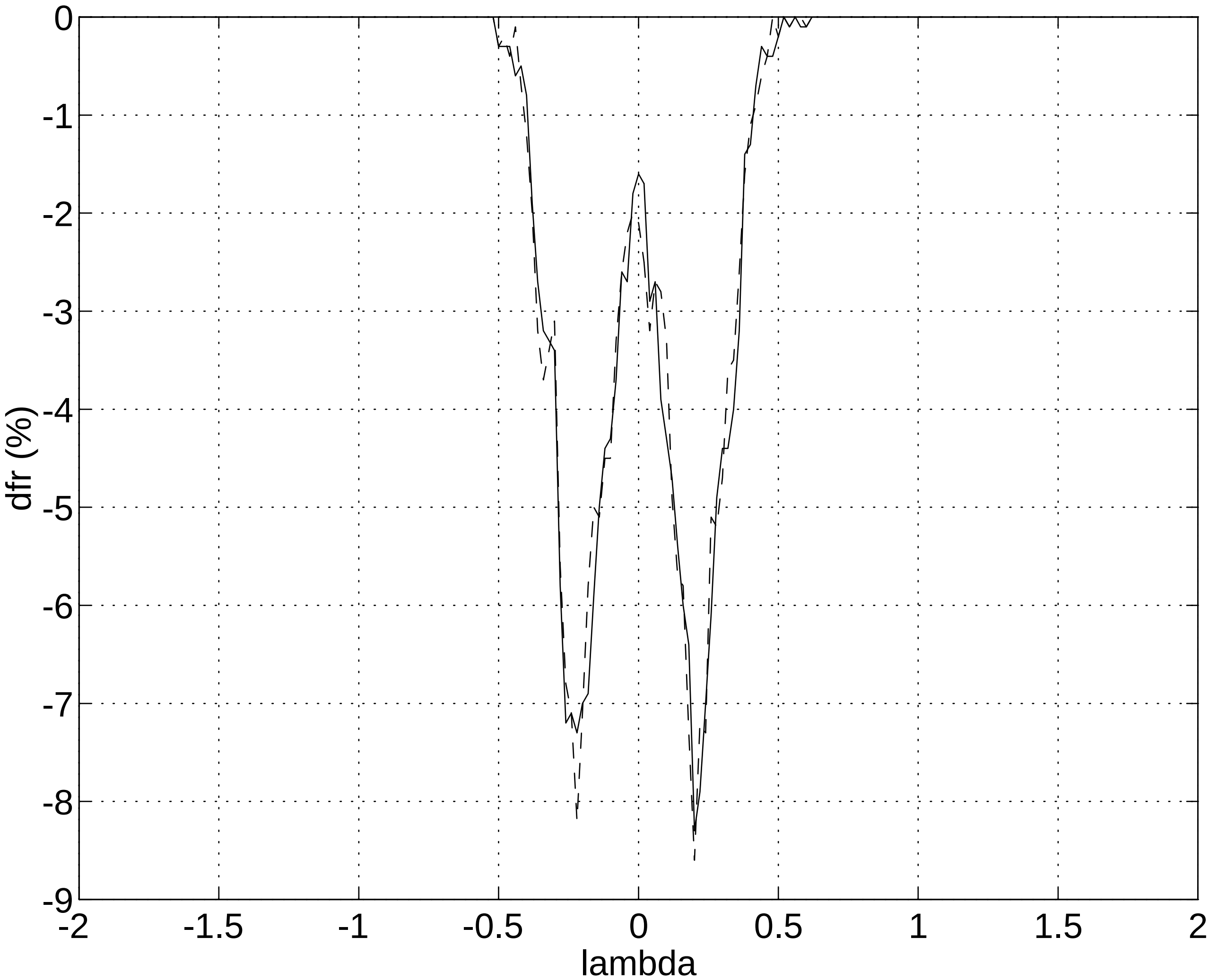

Figure 4. Differences of rejection frequencies for $H_0 : \lambda = 0$

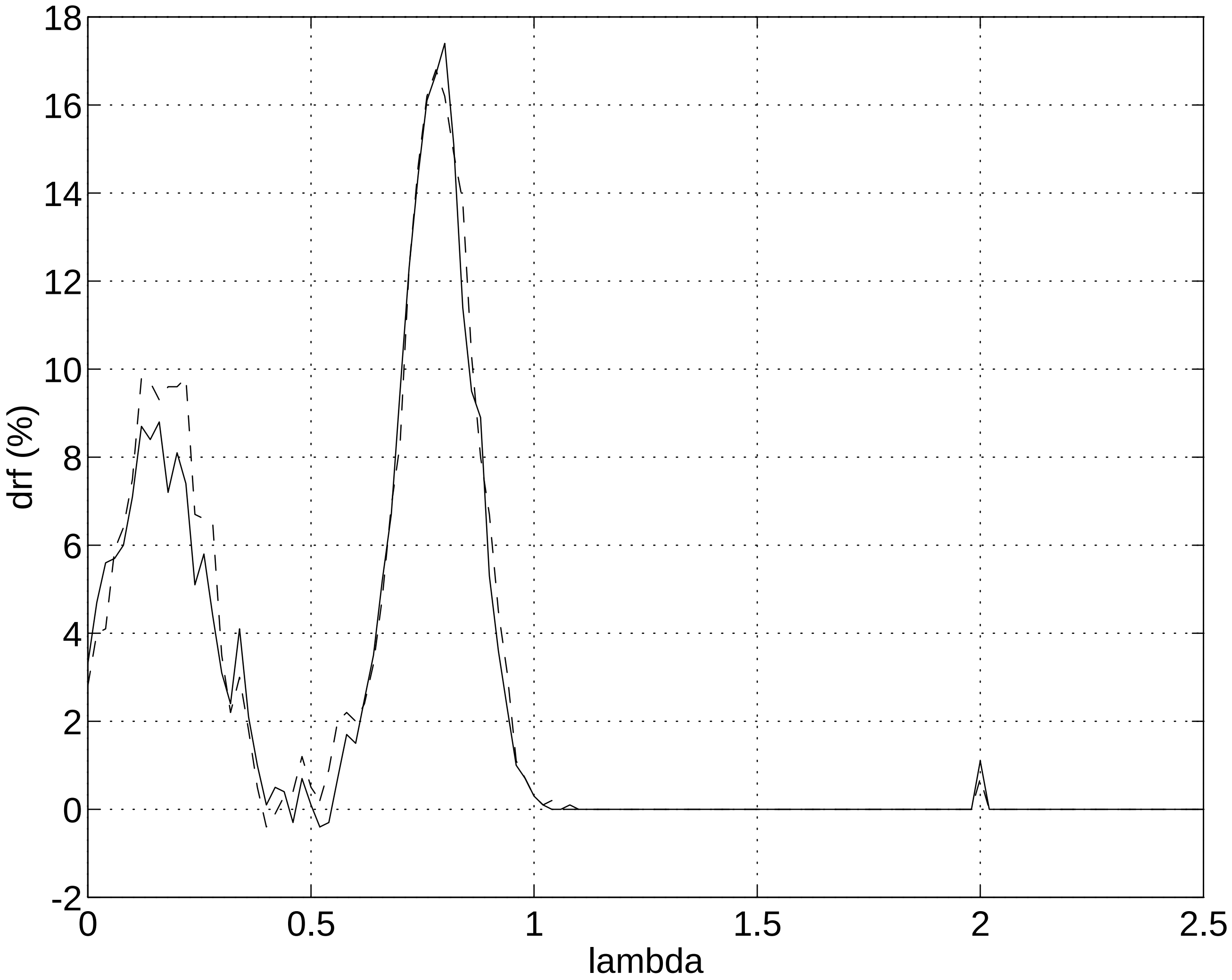

Figure 5. Differences of rejection frequencies for $H_0 : \lambda = 0.5$

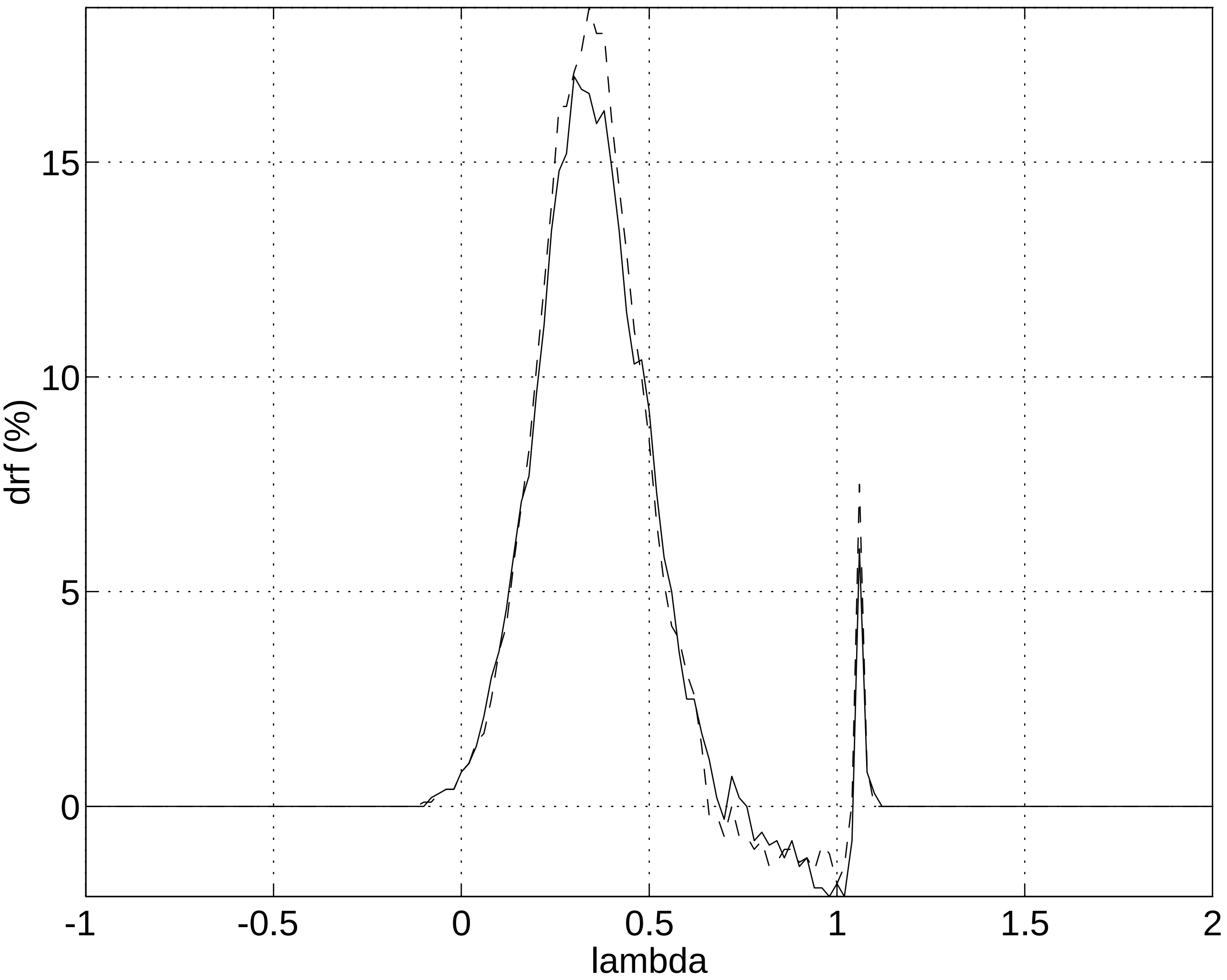

Figure 6. Differences of rejection frequencies for $H_0 : \lambda = 1$

of making a type I error. Second, it is readily applicable to a wide range of econometric specifications of AR(1) models. In particular, it can be used to deal with random walk models, models with a deterministic mean expressed as a linear combination of exogenous variables, including polynomial deterministic trends, *etc*. Third, the critical regions are built from standard distributions which, unlike most asymptotic procedures, do not change with the sample size and/or model specification. Finally, Monte Carlo experiments show that it has good power properties. For those reasons, we think that our procedure should be considered as a good alternative to asymptotic inference methods.

In Section 6, we argued that simulations of power functions were necessary because we could not say *a priori* whether the combination method yields more power. Indeed, on the one side we make use of the whole sample when combining, but on the other side we must lower the bound on the probability of making a type I error (the level) in each of the tests we combine. The former should increase the performance of the procedure whereas the latter should decrease it. The method is easily transposable to higher order autoregressive models and it appears quite plausible the same effect will take place in more general processes. It would certainly be of interest to study this issue further.

Of course, the finite-sample validity of the t and F-type tests described in sections 4 and 5 remain limited to models with Gaussian errors. As usual, these procedures will however be asymptotically valid under weaker distributional assumptions. Further, it is of interest to remember that the general theorems on Markovian processes given in Section 3 hold without parametric distributional assumptions. In particular the conditional independence and truncation properties do not at all require the Gaussian distributional assumption, hence opening the way to distribution-free-procedures. Similarly the test combination technique described in Section 6, which is based on the Boole-Bonferroni inequality, is by no way restricted to parametric models. For example, the latter might be applied to combine distribution-free tests or bootstrap tests [see Nankervis and Savin (1996)] which accommodate more easily non-Gaussian distributions. Such extensions go however beyond the scope of the present paper.

A. Appendix: Proofs

A.1 Proof of Theorem 3.1 We must show that

$$f_{X_{(p+1)}\,,\cdots_{(p+1)}\cdots,X_{n(p+1)}|A_{1,p}} = \prod_{t=1}^{n} f_{X_{t(p+1)}|A_{1,p}}\,.$$

The following equality is always true

$$f_{X_{(p+1)}\,,\cdots_{(p+1)}\cdots,X_{n(p+1)}|A_{1,p}} = f_{X_{(p+1)}|A_{1,p}} \prod_{t=2}^{n} f_{X_{t(p+1)}|A_{1,p},X_{(p+1)},\cdots_{(p+1)}\cdots,X_{(t-1)(p+1)}}\,. \tag{A.1}$$

Consider the t-th term of the product in (A.1) for $t \geq 2$:

$$\begin{aligned} f_{X_{t(p+1)}|A_{1,p},X_{(p+1)},\cdots_{(p+1)}\cdots,X_{(t-1)(p+1)}} &= f_{X_{t(p+1)}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1},A_{t+1,p}} \\ &= \frac{f_{X_{t(p+1)},A_{t+1,p}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}}}{f_{A_{t+1,p}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}}}\,. \end{aligned} \tag{A.2}$$

The numerator in (A.2) can be written

$$\begin{aligned} f_{X_{t(p+1)},A_{t+1,p}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}} &= \int\cdots\int f_{X_{t(p+1)},\cdots_{(1)}\cdots,X_{n(p+1)+p}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}} \\ &\qquad d(x_{(t+1)(p+1)},\cdots_{(p+1)}\cdots,x_{n(p+1)}) \end{aligned}$$

$$\begin{aligned} &= \int\cdots\int \prod_{s=t(p+1)}^{n(p+1)+p} f_{X_s|X_1,\cdots_{(1)}\cdots,X_{s-1}} d(x_{(t+1)(p+1)},\cdots_{(p+1)}\cdots,x_{n(p+1)}) \\ &= \int\cdots\int \prod_{s=t(p+1)}^{n(p+1)+p} f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}} d(x_{(t+1)(p+1)},\cdots_{(p+1)}\cdots,x_{n(p+1)}) \end{aligned}$$

where the last identity follows from the Markovian property $M(p)$. Set

$$g_1(a_{t+1,p},x_{t(p+1)}) \equiv f_{X_{t(p+1)},A_{t+1,p}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}}(a_{t+1,p},x_{t(p+1)})\,.$$

Similarly, we can write the denominator of (A.2) as

$$f_{A_{t+1,p}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}} = \int\cdots\int \prod_{s=t(p+1)}^{n(p+1)+p} f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}} d(x_{t(p+1)},\cdots_{(p+1)}\cdots,x_{n(p+1)})$$

and we denote $g_2(a_{t+1,p}) \equiv f_{A_{t+1,p}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}}(a_{t+1,p})$. Clearly, neither $g_1(a_{t+1,p},x_{t(p+1)})$ nor $g_2(a_{t+1,p})$ depends on $(X_{(p+1)},\ \cdots_{(p+1)}\cdots,\ X_{(t-1)(p+1)})$. Therefore these variables do not enter the ratio (A.2) and we may write the t-th term of the product (A.1) for $t \geq 2$ as

$$f_{X_{t(p+1)}|A_{1,p},X_{(p+1)},\cdots_{(p+1)}\cdots,X_{(t-1)(p+1)}} = f_{X_{t(p+1)}|A_{1,p}} .$$

Since this is true for any $t = 1, 2, \ldots, n$, we can factor the conditional density as

$$f_{X_{(p+1)}\cdots_{(p+1)}\cdots,X_{n(p+1)}|A_{1,p}} = \prod_{t=1}^{n} f_{X_{t(p+1)}|A_{1,p}}$$

which yields the result to be proved. $Q.E.D.$

A.2 Proof of Theorem 3.2 From Theorem **3.1**, $X_{p+1}, X_{2(p+1)}, \ldots, X_{n(p+1)}$ are mutually independent conditionally on $A_{1,p}$, hence

$$\begin{aligned}
f_{X_{t(p+1)}|A_{1,p}} &= f_{X_{t(p+1)}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1},X_{t(p+1)+1},\cdots_{(1)}\cdots,X_{(n+1)(p+1)}} \\
&= \frac{f_{X_{t(p+1)},\cdots_{(1)}\cdots,X_{(n+1)(p+1)}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}}}{f_{X_{t(p+1)+1},\cdots_{(1)}\cdots,X_{(n+1)(p+1)}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}}} \\
&= \frac{f_{X_{t(p+1)},\cdots_{(1)}\cdots,X_{(n+1)(p+1)}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}}}{\int f_{X_{t(p+1)},\cdots_{(1)}\cdots,X_{(n+1)(p+1)}|X_1,\cdots_{(1)}\cdots,X_{t(p+1)-1}} dx_{t(p+1)}} \\
&= \frac{\prod_{s=t(p+1)}^{(n+1)(p+1)} f_{X_s|X_1,\cdots_{(1)}\cdots,X_{s-1}}}{\int \prod_{s=t(p+1)}^{(n+1)(p+1)} f_{X_s|X_1,\cdots_{(1)}\cdots,X_{s-1}} dx_{t(p+1)}} \\
&= \frac{\prod_{s=t(p+1)}^{(n+1)(p+1)} f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}}}{\int \prod_{s=t(p+1)}^{(n+1)(p+1)} f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}} dx_{t(p+1)}} \qquad \text{(A.3)}
\end{aligned}$$

where the last equality is derived using the Markovian property $M(p)$. The product of conditional densities in the numerator of (A.3) can be splitted as $\prod_{s=t(p+1)}^{(n+1)(p+1)} f_{X_s|B_{s,p}} = G_1 \times G_2$, where

$$G_1 \equiv \prod_{s=t(p+1)}^{(t+1)(p+1)-1} f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}} \; , \; G_2 \equiv \prod_{s=(t+1)(p+1)}^{(n+1)(p+1)} f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}} \; .$$

Clearly, G_2 does not depend on $X_{t(p+1)}$. Therefore, the ratio (A.3) simplifies as

$$f_{X_{t(p+1)}|A_{1,p}} = \frac{G_1}{\int G_1 dx_{t(p+1)}} \; . \qquad \text{(A.4)}$$

Now, due to the Markovian property $M(p)$, any of the conditional densities in the product G_1 can be written as $f_{X_s|X_{s-p},\cdots_{(1)}\cdots,X_{s-1}} = f_{X_s|X_{t(p+1)-p},\cdots_{(1)}\cdots,X_{s-1}}$, $s = t(p+1), t(p+1)+1, \ldots, (t+$

$1)(p+1)$. Therefore it is easy to see that

$$G_1 = \prod_{s=t(p+1)}^{(t+1)(p+1)-1} f_{X_s|X_{t(p+1)-p},\cdots_{(1)}\cdots,X_{s-1}}$$

$$= f_{X_{t(p+1)},\cdots_{(1)}\cdots,X_{(t+1)(p+1)-1}|X_{t(p+1)-p},\cdots_{(1)}\cdots,X_{t(p+1)-1}}.$$

Hence

$$\int G_1 dx_{t(p+1)} = f_{X_{t(p+1)+1},\cdots_{(1)}\cdots,X_{(t+1)(p+1)-1}|X_{t(p+1)-p},\cdots_{(1)}\cdots,X_{t(p+1)-1}}$$

and

$$f_{X_{t(p+1)}|A_{1,p}} = \frac{G_1}{\int G_1 dx_{t(p+1)}}$$

$$= f_{X_{t(p+1)}|X_{t(p+1)-p},\cdots_{(1)}\cdots,X_{t(p+1)-1},X_{t(p+1)+1},\cdots_{(1)}\cdots,X_{(t+1)(p+1)-1}}.$$

Since $X_{t(p+1)+1} = X_{(t+1)(p+1)-p}$, we can use the notation of Section 3.1 to write $f_{X_{t(p+1)}|A_{1,p}} = f_{X_{t(p+1)}|B_{t(p+1),p},B_{(t+1)(p+1)}}$ which is the desired result. $Q.E.D.$

A.3 Proof of Theorem 3.6 We need to show that $f_{X_t|B_{t,q},B_{t+q+1,q}}$ does not depend on $X_{t-\tau}$ and $X_{t+\tau}$, for $\tau = p+1, p+2, \ldots, q$. We have:

$$f_{X_t|B_{t+q+1,q},B_{t,q}} = \frac{f_{X_t,B_{t+q+1,q}|B_{t,q}}}{f_{B_{t+q+1,q}|B_{t,q}}} = \frac{f_{X_t,B_{t+q+1,q}|B_{t,q}}}{\int f_{X_t,B_{t+q+1,q}|B_{t,q}} dx_t}.$$

Now, using the fact that $\{X_t : t \in \mathbb{T}\}$ is Markovian of order p, the numerator of this last term can be written $f_{X_t,B_{t+q+1,q}|B_{t,q}} = f_{X_t,\cdots_{(1)}\cdots,X_{t+q}|B_{t,q}} = \prod_{s=t}^{t+q} f_{X_s|B_{s,p}}$ so that

$$f_{X_t|B_{t+q+1,q},B_{t,q}} = \frac{\prod_{s=t}^{t+q} f_{X_s|B_{s,p}}}{\int \prod_{s=t}^{t+q} f_{X_s|B_{s,p}} dx_t} = \frac{\prod_{s=t}^{t+q} f_{X_s|B_{s,p}}}{\left(\prod_{s=t+p+1}^{t+q} f_{X_s|B_{s,p}}\right)\left(\int \prod_{\tau=t}^{t+p} f_{X_\tau|B_{\tau,p}} dx_t\right)}$$

$$= \frac{\prod_{s=t}^{t+p} f_{X_s|B_{s,p}}}{\int \prod_{s=t}^{t+p} f_{X_s|B_{s,p}} dx_t}.$$

It is easy to see that the variables X_s with $t+q \geq s \geq t+p+1$ and $t-p-1 \geq s \geq t-q$ do not appear in the latter expression. $Q.E.D.$

A.4 Proof of Theorem 3.8 Let $\{Y_t : t \in \mathbb{T}\}$ be a Gaussian process having the same first and second order moments as $\{X_t : t \in \mathbb{T}\}$. Then $\{Y_t : t \in \mathbb{T}\}$ must also satisfy the condition in the theorem

$$Y_t \perp (Y_1, \cdots_{(1)} \cdots, Y_{t-p-1}) | Y_{t-p}, \cdots_{(1)} \cdots, Y_{t-1}, \quad \forall t \geq p+1,$$

which is equivalent to the Markovian condition $f_{Y_t|Y_1,\cdots_{(1)}\cdots,Y_{t-1}} = f_{Y_t|Y_{t-p},\cdots_{(1)}\cdots,Y_{t-1}}$, $\forall t \geq p+1$, since $\{Y_t : t \in \mathbb{T}\}$ is Gaussian. From Theorem **3.1**, $Y_{p+1}, \cdots_{(p+1)} \cdots, Y_{n(p+1)}$ are mutually independent, conditional on $A^Y_{1,p}$, where $A^Y_{1,p}$ is defined like $A_{1,p}$ in Section 3.1 with X replaced by Y. Using the normality of $\{Y_t : t \in \mathbb{T}\}$, this is equivalent to

$$Y_{t(p+1)} \perp Y_{s(p+1)} | A^Y_{1,p}\,, \quad \forall t, s \quad \text{such that } 1 \leq t, s \leq n,\ t \neq s\,.$$

This is a condition on the first and second order moments of $\{Y_t : t \in \mathbb{T}\}$, which must also be satisfied by the first and second order moments of $\{X_t : t \in \mathbb{T}\}$. Hence, if $A_{1,p}$ denotes the set of X variables as defined in Section 3.1,

$$X_{t(p+1)} \perp X_{s(p+1)} | A_{1,p}\,, \quad \forall t, s \quad \text{such that } 1 \leq t, s \leq n,\ t \neq s\,.$$

$Q.E.D.$

A.5 Proof of Theorem 3.9 Let $\{Y_t : t \in \mathbb{T}\}$ be a Gaussian process having the same first and second order moments as $\{X_t : t \in \mathbb{T}\}$. From the proof of Theorem **3.8**, we know that $\{Y_t : t \in \mathbb{T}\}$ must also satisfy

$$f_{Y_t|Y_1,\cdots_{(1)}\cdots,Y_{t-1}} = f_{Y_t|Y_{t-p},,\cdots_{(1)}\cdots,Y_{t-1}}\,, \quad \forall t \geq p+1\,.$$

Then, from Theorem **3.2**, we have

$$f_{Y_{t(p+1)}|A^Y_{1,p}} = f_{Y_{t(p+1)}|B^Y_{(t+1)(p+1),p},B^Y_{t(p+1),p}}\,, \quad \forall t \text{ such that } 1 \leq t \leq n\,,$$

where for any s, $B^Y_{s,p} \equiv (Y_{s-p}, \cdots_{(1)} \cdots, Y_{s-1})$. Since $\{Y_t : t \in \mathbb{T}\}$ is Gaussian, this condition is equivalent to

$$Y_{t(p+1)} \perp B^Y_{s(p+1),p} | \left(B^Y_{(t+1)(p+1),p}, B^Y_{t(p+1),p} \right)$$

for all $t \geq 1$ and $s \geq 1$ such that $s \neq t$ and $s \neq t+1$. Since this is condition on the first and second order moments of $\{Y_t : t \in \mathbb{T}\}$, it must also be satisfied by those of $\{X_t : t \in \mathbb{T}\}$. $Q.E.D.$

A.6 Proof of Theorem 4.1 $\mathsf{E_L}\left[X_t | \big(B_{t+p+1,p}, B_{t,p}\big)\right] = \mathsf{E_L}\left[X_t | \big(B_{t+p+1,p}, B^*_{t,p}\big)\right]$ is the affine regression of X_t on $\big(B_{t+p+1,p}, B_{t,p}\big)$, where $B^*_{\nu,p} \equiv (X_{\nu-p}, X_{\nu-p+1}, \ldots, X_{\nu-1})$. The matrix of the coefficients of this regression is given by $\Psi_{12}\Psi_{22}^{-1}$, where

$$\Psi_{12} \equiv \mathsf{cov}\left[X_t, \big(B_{t+p+1,p}, B^*_{t,p}\big)\right]\,, \quad \Psi_{22} \equiv \mathsf{V}\left[\big(B_{t+p+1,p}, B^*_{t,p}\big)\right]\,.$$

Ψ_{22} is non-singular by the regularity assumption. We partition these matrices in the following way:

$$\Psi_{12} \equiv \begin{pmatrix} C_1 & C_2 \end{pmatrix}, \quad \Psi_{22} \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where

$$A_{11} \equiv \mathsf{V}\big(B_{t+p+1,p}\big)\,, \quad A_{22} \equiv \mathsf{V}\big(B^*_{t,p}\big)\,, \quad A'_{21} = A_{12} \equiv \mathsf{cov}\big(B_{t+p+1,p}, B^*_{t,p}\big)\,,$$

$$C_1 \equiv \mathsf{cov}\big(X_t, B_{t+p+1,p}\big), \quad C_2 \equiv \mathsf{cov}\big(X_t, B^*_{t,p}\big).$$

Since $\{X_t : t \in \mathbb{T}\}$ is weakly stationary, $C_1 = C_2 \equiv C$ and $A_{11} = A_{22} \equiv A_1$. We next show that $A_{12} = A_{21}$, *i.e.*, A_{12} is symmetric. The (i, j)-th element of this matrix is

$$\mathsf{cov}(X_{t+p+1-i}, X_{t-p+j-1}) = \gamma_{|t+p+1-i-t+p-j+1|} = \gamma_{|2(p+1)-(i+j)|}$$

where $\gamma_{|s-t|} \equiv \mathsf{cov}(X_s, X_t)$, and its (j, i)-th element is

$$\mathsf{cov}(X_{t+p+1-j}, X_{t-p+i-1}) = \gamma_{|t+p+1-j-t+p-i+1|} = \gamma_{|2(p+1)-(j+i)|}\,.$$

These two terms are identical and consequently $A_{12} = A'_{12} = A_{21} \equiv A_2$. The vector Π whose components are the coefficients of X_{t+k} and X_{t-k}, $1 \le k \le p$, in the affine regression of X_t on $\big(B_{t+p+1,p}, B^*_{t,p}\big)$ is given by

$$\Pi = \begin{pmatrix} C & C \end{pmatrix} \begin{pmatrix} A_1 & A_2 \\ A_2 & A_1 \end{pmatrix}^{-1}.$$

Define Π_1 and Π_2, the two $(1 \times p)$ subvectors of Π whose elements are the coefficients of the variables in $B_{t+p+1,p}$ and in $B^*_{t,p}$, respectively. Then

$$\left.\begin{aligned} C &= \Pi_1 A_1 + \Pi_2 A_2 \\ \\ C &= \Pi_1 A_2 + \Pi_2 A_1 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} A_1(\Pi_1 - \Pi_2) + A_2(\Pi_2 - \Pi_1) = 0 \\ \\ A_2(\Pi_1 - \Pi_2) + A_1(\Pi_2 - \Pi_1) = 0 \end{aligned}\right.$$

which is equivalent to

$$\Psi_{22} \begin{pmatrix} \Pi_1 - \Pi_2 \\ \Pi_2 - \Pi_1 \end{pmatrix} = 0\,.$$

Since the variance-covariance matrix Ψ_{22} is non singular, we must have $\Pi_1 = \Pi_2$. $Q.E.D.$

B. Appendix: Coefficients of two-sided autoregressions for AR(1) processes

The model is

$$Y_t = \phi Y_{t-1} + u_t\,, \quad t = 1, 2, \ldots, n\,,$$

$$u = (u_1, \ldots, u_n)' \sim N(0, \sigma_u^2 I_n)\,,$$

with Y_0 given. Rewriting $Y_t = \phi^t Y_0 + \sum_{i=0}^{t-1} \phi^i u_{t-i}$ and taking expectations, we get $\mathsf{E}(Y_t) = \phi^t Y_0$. The mean deviation process $\{X_t \equiv Y_t - \mathsf{E}(Y_t) : t = 1, 2, \ldots, n\}$ satisfies the autoregression $X_t = \phi X_{t-1} + u_t$.

B.1. Computation of first order moments

Define $\Psi_{12} \equiv \mathsf{cov}\Big[Y_{2t}, \big(Y_{2t+1}\,, Y_{2t-1}\big)'\Big]$ and $\Psi_{22} \equiv \mathsf{V}\big[(Y_{2t+1}\,, Y_{2t-1})'\big]$. From the definition of $\{X_t\,:\,t = 1, 2, \dots, n\}$, we have $X_t = \sum_{i=0}^{t-1} \phi^i u_{t-i}$ and $\mathsf{E}(X_t) = 0$, $\mathsf{E}(X_t^2) = \sigma_u^2 \sum_{i=0}^{t-1} \phi^{2i}$. Furthermore the autocovariances are

$$\mathsf{cov}(Y_{2t+1}, Y_{2t}) = \mathsf{E}(X_{2t+1}X_{2t}) = \sigma_u^2\phi \textstyle\sum_{i=0}^{2t-1} \phi^{2i}\,,$$

$$\mathsf{cov}(Y_{2t}, Y_{2t-1}) = \mathsf{E}(X_{2t}X_{2t-1}) = \sigma_u^2\phi \textstyle\sum_{i=0}^{2t-2} \phi^{2i}\,,$$

$$\mathsf{cov}(Y_{2t+1}, Y_{2t-1}) = \mathsf{E}(X_{2t+1}X_{2t-1}) = \sigma_u^2\phi^2 \textstyle\sum_{i=0}^{2t-2} \phi^{2i}\,,$$

hence

$$\Psi_{12} = \phi\sigma_u^2 \left(\sum_{i=0}^{2t-1} \phi^{2i}\,,\ \sum_{i=0}^{2t-2} \phi^{2i}\right), \quad \Psi_{22} = \sigma_u^2 \begin{pmatrix} \sum_{i=0}^{2t} \phi^{2i} & \phi^2 \sum_{i=0}^{2t-2} \phi^{2i} \\ \phi^2 \sum_{i=0}^{2t-2} \phi^{2i} & \sum_{i=0}^{2t-2} \phi^{2i} \end{pmatrix}.$$

B.2. The affine regression of Y_{2t} on $\big(Y_{2t+1}\,, Y_{2t-1}\big)'$ when $|\phi| \neq 1$

In general we have:

$$\mathsf{E_L}\left[Y_{2t}|\big(Y_{2t+1}\ Y_{2t-1}\big)\right] = \mathsf{E}(Y_{2t}) + \Psi_{12}\Psi_{22}^{-1} \begin{pmatrix} Y_{2t+1} - \mathsf{E}(Y_{2t+1}) \\ Y_{2t-1} - \mathsf{E}(Y_{2t-1}) \end{pmatrix}.$$

Using the fact that, for $|\phi| \neq 1$,

$$\sum_{i=0}^{k} \phi^{2i} = \frac{1 - \phi^{2(k+1)}}{1 - \phi^2}\,,$$

we obtain the following expressions:

$$\Psi_{12} = \frac{\phi\sigma_u^2}{1 - \phi^2}\Big(1 - \phi^{4t}\,,\ 1 - \phi^{4t-2}\Big), \quad \Psi_{22} = \frac{\sigma_u^2}{1 - \phi^2} \begin{pmatrix} 1 - \phi^{4t+2} & \phi^2(1 - \phi^{4t-2}) \\ \phi^2(1 - \phi^{4t-2}) & 1 - \phi^{4t-2} \end{pmatrix},$$

hence

$$\begin{aligned} \mathsf{E_L}\left[Y_{2t}|\big(Y_{2t+1}\,,\ Y_{2t-1}\big)\right] &= \mathsf{E}(Y_{2t}) + \tfrac{\phi}{1+\phi^2}\,\big(1\ \ 1\big) \begin{pmatrix} Y_{2t+1} - \mathsf{E}(Y_{2t+1}) \\ Y_{2t-1} - \mathsf{E}(Y_{2t-1}) \end{pmatrix} \\ &= a + \beta(Y_{2t+1} + Y_{2t-1}), \end{aligned}$$

where $a = \mathsf{E}(Y_{2t}) - \beta[\mathsf{E}(Y_{2t+1}) + \mathsf{E}(Y_{2t-1})]$ and $\beta = \phi/(1 + \phi^2)$. Since for all $t \geq 0$, $\mathsf{E}(Y_t) = \phi^k \mathsf{E}(Y_{t-k})\,,\ k = 0, 1, \dots, t$, $a = 0$.

B.3. The affine regression of Y_{2t} on $\left(Y_{2t+1}\, , Y_{2t-1}\right)'$ when $|\phi| = 1$

When $|\phi| = 1$, we have

$$\Psi_{12} = \phi\sigma_u^2 \big(2t\, ,\ 2t-1\big)\, , \quad \Psi_{22} = \sigma_u^2 \begin{pmatrix} 2t+1 & 2t-1 \\ 2t-1 & 2t-1 \end{pmatrix},$$

hence

$$\mathsf{E}_\mathsf{L}\left[Y_{2t} \middle| \big(Y_{2t+1}\, ,\ Y_{2t-1}\big)\right] = \frac{\phi}{2}(Y_{2t+1} + Y_{2t-1}) = \frac{\phi}{1+\phi^2}(Y_{2t+1} + Y_{2t-1}).$$

Note that from the derivations in the case where $|\phi| \neq 1$, we have $a = 0$ irrespective to the value of ϕ. In any case, the residual variance is

$$\mathsf{V}\Big[Y_{2t} - \mathsf{E}_\mathsf{L}\left[Y_{2t} \middle| \big(Y_{2t+1}\, ,\ Y_{2t-1}\big)\right]\Big] = \mathsf{V}(Y_{2t}) - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}' = \frac{\sigma_u^2}{1+\phi^2}\ ,\ \phi \in (-\infty, \infty).$$

References

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

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.

BURNSIDE, C., AND M. S. EICHENBAUM (1996): "Small-Sample Properties of GMM-Based Wald Tests," *Journal of Business and Economic Statistics*, 14, 294–308.

CAMPBELL, B., AND J.-M. DUFOUR (1997): "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter," *International Economic Review*, 38, 151–173.

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. (1989): "Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions," *Econometrica*, 57, 335–355.

——— (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., M. HALLIN, AND I. MIZERA (1998): "Generalized Runs Tests for Heteroskedastic Time Series," *Journal of Nonparametric Statistics*, 9, 39–86.

DUFOUR, J.-M., AND J. F. KIVIET (1996): "Exact Tests for Structural Change in First-Order Dynamic Models," *Journal of Econometrics*, 70, 39–68.

——— (1998): "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models," *Econometrica*, 66, 79–104.

DUFOUR, J.-M., AND O. TORRÈS (1998): "Union-Intersection and Sample-Split Methods in Econometrics with Applications to SURE and MA Models," in *Handbook of Applied Economic Statistics*, ed. by D. E. A. Giles, and A. Ullah, pp. 465–505. Marcel Dekker, New York.

GOURIÉROUX, C., AND A. MONFORT (1989): *Statistique et modèles économétriques, Volumes 1 et 2*. Economica, Paris.

HANNAN, E. J. (1955a): "An Exact Test for Correlation Between Time Series," *Biometrika*, 42, 316–326.

——— (1955b): "Exact Tests for Serial Correlation," *Biometrika*, 42, 316–326.

——— (1956): "Exact Tests for Serial Correlation in Vector Processes," *Proceedings of the Cambridge Philosophy Society*, 52, 482–487.

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.

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.

KIVIET, J. F., AND J.-M. DUFOUR (1997): "Exact Tests in Single Equation Autoregressive Distributed Lag Models," *Journal of Econometrics*, 80, 325–353.

KRISHNAIAH, P. R., AND V. K. MURTHY (1966): "Simultaneous Tests for Trend and Serial Correlation for Gaussian Markov Residuals," *Econometrica*, 34, 472–480.

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

LINNIK, U. V. (1949): "K Teorii Neodnorodnī'kh Tsepeĭ Markova," *Izvestiya Akademii Nauk SSSR, Seriya Matematicheskaya*, 13, 65–94, In Russian.

MAASOUMI, E. (1992): "Fellow's Opinion: Rules of Thumb and Pseudo-Science," *Journal of Econometrics*, 53, 1–4.

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 (1996): "The Level and Power of the Bootstrap t Test in the AR(1) Model with Trend," *Journal of Business and Economic Statistics*, 14, 161–168.

OGAWARA, M. (1951): "A Note on the Test of Serial Correlation Coefficients," *Annals of Mathematical Statistics*, 22, 115–118.

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

SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches, and M. D. Intrilligator, 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.

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