Exact simulation-based inference for autoregressive processes based on induced tests*

Jean-Marie Dufour†
Université de Montréal

Malika Neifar ‡
Institut Supérieur de Gestion de Sousse

First version: June 2003
Revised: March 2004

ABSTRACT

In this paper, we propose exact inference methods for autoregressive models of given order p $(p \geq 1)$, which may be nonstationary and include a drift term. We study the problem of testing any hypothesis that sets the complete vector of autoregressive coefficients. This is done by first transforming the model to eliminate serial dependence under the null hypothesis, and then testing whether autocorrelation remains present in the transformed data. Tests for dependence at different lags are then combined using the methods proposed by Tippett (1931) and Fisher-Pearson [Fisher (1932), Pearson (1933)] for combining independent tests. In view of the dependence among the different tests, the size of the combined procedures is controlled by using Monte Carlo test techniques. The construction of valid confidence sets based on these tests is discussed. Numerical illustrations based on simulated data are also presented.

Keywords: autoregressive process; exact inference; induced test; test combination; Monte Carlo test.

*This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Institut de Finance mathématique de Montréal (IFM2), 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, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour.

‡ Département des méthodes quantitatives et technologie de l'information, Institut Supérieur de Gestion de Sousse. Rue Abdel Aziz Elbahi C.P.763, Sousse 4000, Tunisie. TEL (FAX): 011 216 73 367 128
Courriel: malika.neifar@isgs.rnu.tn and malika.neifar@topnet.tn.

1. Introduction

Statistical inference (tests and confidence regions) on autoregressive (AR) models constitute a basic problem in time series analysis, statistics and econometrics. Applied methods are generally based on unreliable asymptotic approximations even under strong parametric assumptions (such as, Gaussian innovations). But the actual level of asymptotically based tests can differ markedly from the posted level even with reasonably large samples, especially for processes of order greater than one [see Blough (1992), Dufour (2003) and Stock (1994)].

In this paper, we consider the problem of testing any hypothesis that sets the full vector of the autoregressive coefficients. We propose tests which are applicable on both stationary and nonstationary processes including a drift term, with possibly non-Gaussian errors. Further, we consider the problem of building exact confidence intervals and regions for the AR coefficients.

For that purpose, the data are first transformed under the null hypothesis so that the data become independent. Then the filtered data are tested for the presence of serial dependence at several lags. This raises the problem of combining tests against serial dependence at different lags. In Dufour and Neifar (2001), we considered a combination technique based on the Boole-Bonferroni inequality. This has the disadvantage of producing conservative tests, hence a power loss. Here, we consider induced tests based on two approaches originally suggested for independent test p-values, namely, the minimum criterion of Tippett (1931) and the Fisher-Pearson product criterion [Fisher (1932) and Pearson (1933)]; for further discussion of induced tests, see Dufour and Torrès (1998) and Dufour and Khalaf (2002). Since the tests are not independent, the overall size of the test is controlled by the technique of Monte Carlo tests [see Dufour and Khalaf (2001) and Dufour (2002)]. The technique of MC tests allows one to obtain provably exact randomized tests in finite samples using very small numbers of MC replications of the test statistics under null hypothesis.

This paper is organized as follows. In section 2, we describe the model and test problems studied. In section 3, the test statistics are derived. The implementation by the MC test technique is explained in section 4. Numerical illustrations based on simulated data are presented in section 5.

2. Framework

We consider a parametric autoregressive model of order p [AR(p)]:

$$y_t = \beta + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t \,, \tag{2.1}$$

$$u_t = \sigma \varepsilon_t \text{ where } \varepsilon_t \overset{i.i.d.}{\sim} D \,,\ t = 1, \ldots, T \,, \tag{2.2}$$

where D is a completely specified distribution, the errors $\boldsymbol{u} = (u_1, \ldots, u_T)'$ are independent of the initial values $y_0, y_{-1}, \ldots, y_{-p+1}$, the parameters $\varphi_1, \varphi_2, \ldots, \varphi_p$, β and σ^2 are unknown, $T \geq p+1$ and $p \geq 1$. Assumption (2.2) means that the errors are independent and identically distributed (i.i.d.) according to a distribution which is specified up to an unknown scale parameter. An important special case of (2.2) is the one where the errors are i.i.d. Gaussian:

$$u_t \overset{i.i.d.}{\sim} [0,\ \sigma^2],\ t = 1, \ldots, T \,. \tag{2.3}$$

But other distributions could be considered, such as more heavy-tailed distributions (e.g., a Cauchy distribution).

It will useful here to reparameterize this model as in Dickey (1976) and Beveridge and Nelson (1981). This yields the equivalent form:

$$\begin{aligned} y_t &= \beta + \big(\sum_{j=1}^{p} \varphi_j\big) y_{t-1} + \sum_{j=1}^{p-1} \big(-\sum_{i=j+1}^{p} \varphi_i\big)[y_{t-j} - y_{t-(j+1)}] + u_t \\ &= \beta + \theta_1 y_{t-1} + \sum_{j=2}^{p} \theta_j \triangle y_{t-(j+1)} + u_t\,, \quad t = 1, \ldots, T\,, \end{aligned} \tag{2.4}$$

where $\theta_1 = \varphi_1 + \varphi_2 + \cdots + \varphi_p$, $\theta_j = -\sum_{i=j}^{p} \varphi_i$, and $\triangle y_{t-j} = y_{t-j} - y_{t-(j+1)}$, $j \geq 2$. Now, unknown parameters are $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$, β and σ.

We wish to study the problem of testing any hypothesis which sets the complete vector of autoregressive coefficients $\boldsymbol{\theta}$ in (2.4) at any specified value $\boldsymbol{\theta}_0$ [in an admissible set S]:

$$H_0(\boldsymbol{\theta}_0) : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{against} \quad H_a(\boldsymbol{\theta}_0) : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0\,. \tag{2.5}$$

3. Test statistics

In order to test $H_0(\boldsymbol{\theta}_0) : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, we consider the following transformation of the data:

$$z_t(\boldsymbol{\theta}_0) = y_t - \theta_{10} y_{t-1} - \theta_{20} \triangle y_{t-1} - \cdots - \theta_{p0} \triangle y_{t-p+1} = \beta + u_t\,,\ t = 1, \ldots, T\,. \tag{3.1}$$

Thus, under $H_0(\boldsymbol{\theta}_0)$, the filtered variables $z_t(\boldsymbol{\theta}_0)$, $t = 1, \ldots, T$ are distributed like $u_t\,,\ t = 1, \ldots, T$, except possibly for the addition of an unknown constant β. Under the alternative hypothesis $H_a(\boldsymbol{\theta}_0)$, the same variables are autoccorrelated, following an ARMA(p, p) process. Consequently, we can test $H_0(\boldsymbol{\theta}_0)$ by testing whether the variables $z_t(\boldsymbol{\theta}_0)\,,\ t = 1, \ldots, T$, are mutually independent. Since the autocorrelation structure of an ARMA(p, p) process is completely specified by its first p autocorrelations, it will be sufficient to test the independence of $z_t(\boldsymbol{\theta}_0), t = 1, \ldots, T$, against autocorrelations at lags $1, 2, \ldots, p$ [see Neifar (1996) for more details].

To do so, we can use p statistics of the form:

$$D_j(\boldsymbol{\theta}_0) = \frac{\boldsymbol{z}(\boldsymbol{\theta}_0)'\, \bar{\boldsymbol{A}}_j\, \boldsymbol{z}(\boldsymbol{\theta}_0)}{\boldsymbol{z}(\boldsymbol{\theta}_0)'\, \bar{\boldsymbol{B}}_j\, \boldsymbol{z}(\boldsymbol{\theta}_0)}\,,\ j = 1, \ldots, p\,, \tag{3.2}$$

where $\bar{\boldsymbol{A}}_j = \boldsymbol{M}\boldsymbol{A}_j\boldsymbol{M}$, $\bar{\boldsymbol{B}}_j = \boldsymbol{M}\boldsymbol{B}_j\boldsymbol{M}$, $\boldsymbol{M} = \boldsymbol{I}_T - \frac{1}{T}\boldsymbol{\iota}'\boldsymbol{\iota}$, $\boldsymbol{\iota} = (1, 1, \ldots, 1)'$ and $\boldsymbol{B}_j$ is a positive definite matrix. $D_j(\boldsymbol{\theta}_0)$ may be interpreted as a test statistic designed to especially sensitive to serial dependence at lag j. Many commonly used autocorrelation tests – for example, tests based on Durbin-Watson statistics and sample autocorrelations – involve statistics of the form (3.2). But the null hypothesis we focus on is independence (which entails the absence of serial dependence at all lags) between $z_t(\boldsymbol{\theta}_0), t = 1, \ldots, T$.

Under the assumptions (2.1) - (2.2) and $H_0(\boldsymbol{\theta}_0)$, $D_j(\boldsymbol{\theta}_0)$ has a distribution which does not depend on nuisance parameters (β and σ) and can be easily simulated by MC techniques [or calculated for example by Imhof's (1961) algorithm in the Gaussian case]. Further, this distribution does not depend on $\boldsymbol{\theta}_0$. Note that an error normality assumption is not required for this invariance to hold.

In order to test the hypothesis of independence, we will combine tests based on $D_j(\boldsymbol{\theta}_0)$, $j = 1, \ldots, p$, in a way that will control the overall level of the procedure. Specifically, we consider two methods for combining the tests (or p-values) without using a conservative bound, such as the Bonferroni inequality: namely, the minimum criterion of Tippett (1931) and the Fisher-Pearson product criterion [Fisher (1932) and Pearson (1933)].

For that purpose, we start by considering the survival function (or p-value function) of $D_j(\boldsymbol{\theta}_0)$ under $H_0(\boldsymbol{\theta}_0)$:

$$G_j(x) = \mathsf{P}[D_j(\boldsymbol{\theta}_0) \geq x]\,.$$

On evaluating this function at $x = \hat{D}_j(\boldsymbol{\theta}_0)$, the observed value of $D_j(\boldsymbol{\theta}_0)$, we get the marginal significance level of the test of $H_0(\boldsymbol{\theta}_0)$ based on $D_j(\boldsymbol{\theta}_0)$:

$$p_j = G_j[\hat{D}_j(\boldsymbol{\theta}_0)]\,.$$

For one-sided tests, the critical region with level α_j $(0 < \alpha_j < 1)$ for each test $D_j(\boldsymbol{\theta}_0)$ has the form

$$p_j \leq \alpha_j,$$

$j = 1, \ldots, p$. Similarly, to get a two-sided test with level α_j based on $D_j(\boldsymbol{\theta}_0)$, we can take

$$\min\{p_j,\, 1 - p_j\} \leq \alpha_j/2\,. \tag{3.3}$$

If the exact distribution of the statistic $D_j(\boldsymbol{\theta}_0)$ is unknown, we can also use an asymptotic approximation or simulate it through a Monte Carlo experiment independent of the data. For example, we can consider $t_j = |\sqrt{T}\, r_j(\boldsymbol{\theta}_0)|$ which is asymptotically $\mathrm{N}[0,\, 1]$ under H_0, where $r_j(\boldsymbol{\theta}_0)$ is the lag j sample autocorrelation based on the transformed data. To the extent that the joint distribution of the statistics $D_j(\boldsymbol{\theta}_0)$, $j = 1, \ldots, p$, is free of nuisance parameters and can be simulated (in the present case, this is easy to do), the fact that individual p-values are only approximate can be dealt with automatically by the MC test technique described in the following section.

A difficulty we meet here consists in combining multiple tests based on different statistics of the form $D_j(\boldsymbol{\theta}_0)$, $j = 1, \ldots, p$. We will consider here two ways of combining multiple test p-values, the minimum criterion of Tippett (1931) and the Fisher-Pearson product criterion [Fisher (1932) and Pearson (1933)], which can be described as follows:

1. Tippett's minimum p-value criterion is based on

$$F_{\min} = \inf_{j=1,\ldots,p} \{p_j\},$$

or

$$\bar{F}_{\min} = 1 - \inf_{j=1,\ldots,p} \{p_j\}\,;$$

the null hypothesis is rejected when $F_{\min}$ is small or, equivalently, when $\bar{F}_{\min}$ is large;

2. the Fisher-Pearson product criterion is based on

$$F_{\times} = \prod_{j=1}^{p} p_j$$

or

$$\bar{F}_{\times} = 1 - \prod_{j=1}^{p} p_j \, ;$$

the null hypothesis is rejected when $F_{\times}$ is small or, equivalently, when $\bar{F}_{\times}$ is large.

Even though the distribution of the above induced test statistics may be quite difficult to establish analytically, they do not involve nuisance parameters and can be easily simulated. Consequently, we can apply to them the technique of MC tests [Dufour (2002)], which is described in section 4.

Confidence sets can be obtained by "inverting" the above proposed tests for hypotheses $H_0(\boldsymbol{\theta}_0)$, i.e., by finding the set I of values $\boldsymbol{\theta}_0$ which are not rejected at level α. This yields a joint confidence set I with level $1 - \alpha$ for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$:

$$P[\boldsymbol{\theta} \in I] \geq 1 - \alpha. \tag{3.4}$$

By projection, we can also build simultaneous confidence intervals for the individual coefficients θ_i, $i = 1, \ldots, p$; for more details, see Dufour and Kiviet (1998), Dufour and Khalaf (2001) and Dufour (2002).

4. Finite-sample Monte Carlo tests

Consider a statistics, say S, for testing $H_0(\boldsymbol{\theta}_0)$ and assume its distribution is continuous. Let S_0 be the test statistic computed from the observed data. In view of the fact that S has a distribution that does not involve nuisance parameters under the null hypothesis, we can use Monte Carlo methods to simulate N i.i.d. replications $S_1, \ldots, S_N$ of S under $H_0(\boldsymbol{\theta}_0)$, independently of S_0. Since $S_0, S_1, \ldots, S_N$ are i.i.d., all rankings of these $N+1$ variables are equally probable. It follows that the rank $R_N(S_0)$ of S_0 has a uniform distribution over the integers $1, \ldots, N+1$.

Let us define the p-value function

$$p_N(x) = \frac{NG_N(x) + 1}{N + 1} \tag{4.5}$$

where

$$G_N(x) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{[0,\,\infty)}(S_j - x), \quad \mathbf{1}_A(x) = \begin{cases} 1\,, & \text{if } x \in A\,, \\ 0\,, & \text{if } x \notin A\,. \end{cases}$$

If N is chosen so that $\alpha(N+1)$ is an integer (e.g., for $\alpha = 0.05$, we can take $N = 19, 99, 199,$

etc.), the above uniformity result entails that

$$\mathsf{P}[p_N(S_0) \leq \alpha] = \alpha \tag{4.6}$$

under the null hypothesis. Thus the critical region $p_N(S_0) \leq \alpha$ has level α. In other words, the randomized critical region $p_N(S_0) \leq \alpha$ control the level of the test and has the same level as the critical region $G(S_0) \leq \alpha$. For further details and references, see Dufour and Kiviet (1998), Dufour and Khalaf (2001) and Dufour (2002).

5. Numerical illustration

In order to illustrate the above methodology, we generated data using AR(2) models with the following coefficients:

Model	M_1	M_2	M_3	M_4
Parameter values	$\theta_1 = 0, \theta_2 = 1$	$\theta_1 = 0, \theta_2 = 0$	$\theta_1 = 1, \theta_2 = 0$	$\theta_1 = 1, \theta_2 = 1$

where $u_t \overset{i.i.d.}{\sim} \mathrm{N}[0,\ \sigma^2],\ t = 1, \ldots, T,\ T = 52,\ \beta = 0,\ (y_0,\ y_{-1}) = (0,\ 0)$. For level $\alpha = 0.05$ and using $N = 19$ replications, we applied the method described in the previous sections to test hypotheses on $\boldsymbol{\theta} = (\theta_1, \theta_2)'$, using the Fisher-Pearson combination criterion $(\bar{F}_\times)$, and build the corresponding confidence set for $\boldsymbol{\theta}$ (with level $1 - \alpha$). The p-values for individual tests were approximated form a preliminary simulation (independent of the rest) using 199 replications. The joint confidence sets for (θ_1, θ_2) appear (in black) in tables 1 to 4. The triangles represent the boundaries of the region inside which the process is stable (i.e., the roots of the AR polynomial are outside the unit circle). The corresponding projection-based confidence intervals are given in the following table:

Model	M_1	M_2	M_3	M_4
θ_1	$[-0.14,\ 0.08]$	$[-0.42,\ 0.07]$	$[0.81,\ 1.27]$	$[0.99,\ 1.01]$
θ_2	$[0.91,\ 1.17]$	$[-0.11,\ 0.30]$	$[-0.22,\ 0.68]$	$[0.84,\ 1.19]$

We see from these numerical results that the confidence sets are quite precise and (as expected) cover the true parameter values. Of course, more precise confidence sets can be obtained by using larger numbers of Monte Carlo replications.[1]

References

Beveridge, S. and Nelson, C. (1981), 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycles", *Journal of Monetary Economics* **7**, 151– 174.

[1] Due to space limitations and the computer intensive nature of the procedures proposed above, we cannot present here a full-fledged power simulation or a detailed empirical application.

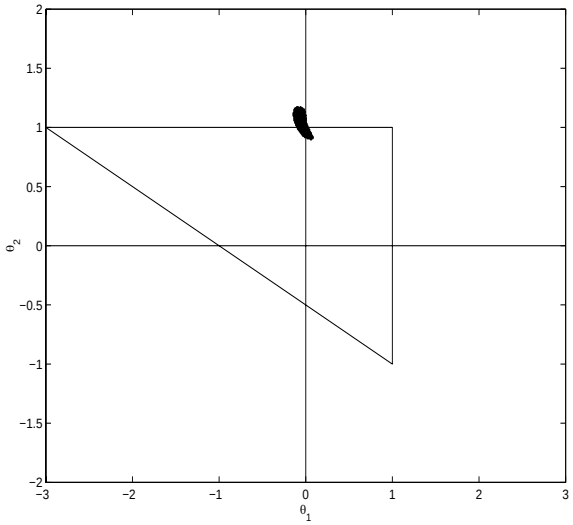

Table 1. Confidence set for model M1

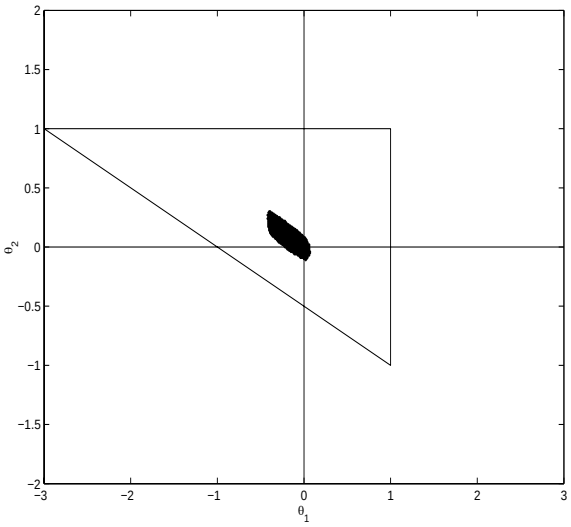

Table 2. Confidence set for model M2

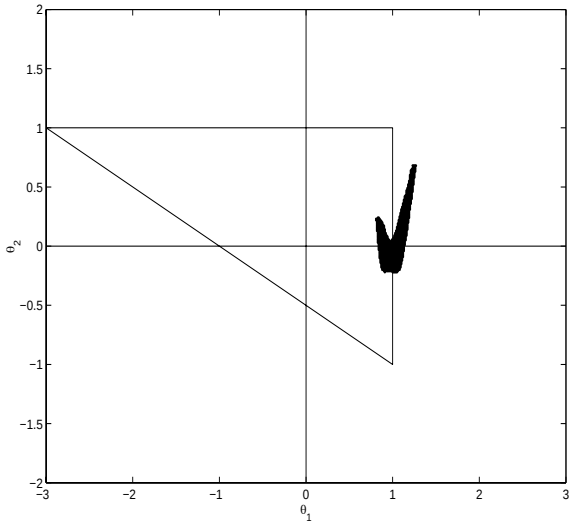

Table 3. Confidence set for model M3

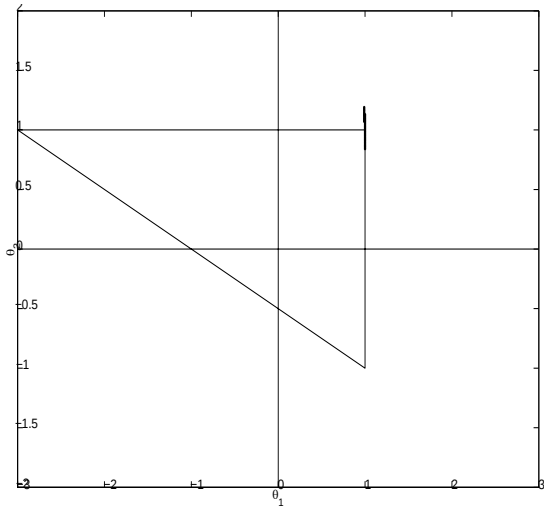

Table 4. Confidence set for model M4

Blough, S. R. (1992), 'The relationship between power and level for generic unit root tests in finite samples', *Journal of Applied Econometrics* **7**, 295–308.

Dickey, D. A. (1976), Estimation and Testing of Non Stationary Time Series, PhD thesis, University of Iowa.

Dufour, J.-M. (2002), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **forthcoming**.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. and Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, pp. 494–519.

Dufour, J.-M. and Khalaf, L. (2002), 'Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions', *Journal of Econometrics* **106**(1), 143–170.

Dufour, J.-M. and Kiviet, J. F. (1998), 'Exact inference methods for first-order autoregressive distributed lag models', *Econometrica* **66**, 79–104.

Dufour, J.-M. and Neifar, M. (2001), 'Méthodes d'inférence exactes pour des processus autorégressifs: une approche fondée sur des tests induits', *L'Actualité économique* **78**(1), 19–40.

Dufour, J.-M. and Torrès, O. (1998), Union-intersection and sample-split methods in econometrics with applications to SURE and MA models, *in* D. E. A. Giles and A. Ullah, eds, 'Handbook of Applied Economic Statistics', Marcel Dekker, New York, pp. 465–505.

Fisher, R. A. (1932), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

Imhof, P. J. (1961), 'Computing the distribution of quadratic forms in normal variables', *Biometrika* **48**, 419–426. Corrigendum 49 (1962), 284.

Neifar, M. (1996), Méthodes d'inférence exactes pour des modèles de régression avec erreurs autoregressives et applications macroéconomiques, PhD thesis, Université de Montréal.

Pearson, K. (1933), 'On a method of determining whether a sample of size n supposed to have been drawn from a parent population', *Biometrika* **25**, 379–410.

Stock, J. H. (1994), Unit root, structural breaks and trends, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics, Volume IV', North-Holland, Amsterdam, chapter 46, pp. 2740–2841.

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