Dynamic specification and cointegration

Boswijk, H.P.; Franses, P.H.

Published in:
Oxford Bulletin of Economics and Statistics

DOI:
10.1111/j.1468-0084.1992.tb00007.x

OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 54, 3 (1992)
0305-9049 $3.00

DYNAMIC SPECIFICATION AND COINTEGRATION*

Peter Boswijk and Philip Hans Franses

I. INTRODUCTION

Since the publication of Granger's (1981) seminal paper on cointegration, this topic has received considerable attention both in empirical and in theoretical research. Although initial results suggested that short-run dynamics may be neglected for estimating cointegrating relations, subsequent research has indicated that reliable inference necessitates a proper dynamic specification. The effect of dynamic model selection on the size and power of tests for cointegration and unit roots is investigated by e.g. Molinas (1986), Schwert (1989), Kunst (1989), and Franses (1990). In particular, Molinas showed that two independent autoregressive integrated moving average (ARIMA) processes with a strongly negative moving average (MA) parameter can appear to be cointegrated quite often if Engle and Granger's (1987) residual Dickey–Fuller test is used. Franses (1990) reports similar evidence for the Wald test (Boswijk, 1989).

In this paper we use some Monte Carlo experiments to investigate the effects of dynamic specification on the size and power of three cointegration tests. The first test, proposed by Engle and Granger (1987), is the residual augmented Dickey–Fuller unit root test (ADF). The second is a Wald test for the significance of the error correction mechanism in an autoregressive-distributed lag (ADL) model, suggested by Boswijk (1989) and further developed in Boswijk (1991). The third test is a likelihood ratio (LR) test in a vector autoregressive (VAR) model, proposed by Johansen (1988) and extended in Johansen and Juselius (1990). Two prototypical data generating processes (DGPs) for a bivariate time series $\{(y_t, z_t)\}$ are considered. In both cases z_t is exogenously generated by a random walk; in the first DGP y_t is determined by an ADL model, and in the second case by an autoregressive moving average model with explanatory variables (ARMAX). Hence for the first DGP standard regression equations can be expected to capture the dynamics completely, whereas in the second case they can merely provide an approximation. For both DGPs a cointegrated and a non-cointegrated

*Suggestions from Niels Haldrup, David Hendry and Jan Kiviet are gratefully acknowledged.

version is considered in order to evaluate both the size and the power of the tests. The same simulations are used to investigate whether model selection procedures based on the Lagrange-multiplier first-order serial correlation test (LMF) and the Schwarz criterion (SC) lead to a lag length selection which is optimal in terms of both size and power properties.

The plan of the paper is as follows. In Section II, we review the cointegration tests. In the next section, the Monte Carlo design is described, and some results are given. In the fourth section these results are interpreted, and their practical implications are discussed.

II. TESTING FOR COINTEGRATION

Consider the vector autoregressive model of order p, or VAR(p), for an $n \times 1$ vector time series $\{\boldsymbol{x}_t, t=1,\ldots, T\}$:

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Pi}_1 \boldsymbol{x}_{t-1} + \ldots + \boldsymbol{\Pi}_p \boldsymbol{x}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{2.1}$$

where $\boldsymbol{\mu}$ is an $n \times 1$ vector of intercepts, where $\boldsymbol{\Pi}_i$, $i=1,\ldots,p$, are $n \times n$ parameter matrices, and $\{\boldsymbol{\varepsilon}_t\}$ and white noise errors with covariance matrix Σ. Using the first-difference operator Δ, defined by $\Delta \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$, the model can be rewritten as

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{x}_{t-1} + \ldots \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{2.2}$$

where

$$\boldsymbol{\Pi} = -(\boldsymbol{I} - \boldsymbol{\Pi}_1 - \ldots - \boldsymbol{\Pi}_p),$$

$$\boldsymbol{\Gamma}_i = \boldsymbol{\Pi}_{i+1} + \ldots + \boldsymbol{\Pi}_p, \quad i=1,\ldots,p-1.$$

If the rank r of $\boldsymbol{\Pi}$ satisfies $0 < r < n$, this matrix can be decomposed as $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices. The Granger Representation Theorem, see Engle and Granger (1987), implies that in that case the components of $\boldsymbol{x}_t$ are non-stationary and integrated of order 1, whereas the linear combinations $\boldsymbol{\beta}'\boldsymbol{x}_t$ are stationary. The process is then said to be cointegrated, the columns of $\boldsymbol{\beta}$ are the cointegrating vectors, and $\boldsymbol{\alpha}$ contains the error correction coefficients.

In this paper we examine the behaviour of three procedures to test for cointegration. The methods differ mainly with respect to the choice of the model in which the cointegration properties are analysed, but also with respect to the additional assumptions that are required. Firstly, Engle and Granger (1987) suggest to estimate $\boldsymbol{\beta}$ in a static regression equation. Consider the case where $r=1$, so that $\boldsymbol{\beta}$ is an $n \times 1$ vector, normalized as $\boldsymbol{\beta}' = (1, -\boldsymbol{\theta}')$. Correspondingly, the vector $\boldsymbol{x}_t$ partitioned as $\boldsymbol{x}_t' = (y_t, z_t')$, where y_t is taken as the dependent variable, and the remaining variables in z_t are used as explanatory variables. This leads to the static regression equation:

$$y_t = a + \boldsymbol{\theta}' z_t + u_t. \tag{2.3}$$

The least-squares residuals from this cointegrating regression, $\hat{u}_t$, are checked for the presence of a unit root by the augmented Dickey–Fuller (ADF) test, which is the t-ratio of ρ in the auxiliary regression

$$\Delta\hat{u}_t = m + \rho\hat{u}_{t-1} + \gamma_1\Delta\hat{u}_{t-1} + \ldots + \gamma_{p-1}\Delta\hat{u}_{t-p+1} + e_t. \tag{2.4}$$

The asymptotic properties of the test are analysed in Phillips and Ouliaris (1990), where asymptotic critical values for the hypothesis $\rho = 0$ are tabulated (tables IIa–IIc).

The conditional dynamic regression procedure developed in Boswijk (1989) starts from the assumption that r is either 1 or 0, and that the vector z_t is weakly exogenous for the cointegration parameters. Under the assumption that $\{\boldsymbol{\varepsilon}_t\}$ are normally distributed, a conditional error correction model for y_t given z_t may be derived from (2.2) and reads as

$$\begin{aligned}\Delta y_t &= c + \boldsymbol{\delta}_0'\Delta z_t + \lambda(y_{t-1} - \boldsymbol{\theta}' z_{t-1}) + \sum_{j=1}^{p-1} (\varphi_j \Delta y_{t-j} + \boldsymbol{\delta}_j'\Delta z_{t-j}) + \eta_t \\ &= c + \boldsymbol{\delta}_0'\Delta z_t + \boldsymbol{\pi}' x_{t-1} + \sum_{j=1}^{p-1} (\varphi_j \Delta y_{t-j} + \boldsymbol{\delta}_j'\Delta z_{t-j}) + \eta_t,\end{aligned} \tag{2.5}$$

where $\boldsymbol{\pi}' = \lambda(1, -\boldsymbol{\theta}')$. If $\lambda \neq 0$, then (2.5) may be seen as a reparametrized ADL$(p; p, .., p)$ model with y_t and z_t cointegrated. If $\lambda = 0$, then the model is unstable, so that the error correction mechanism is absent, and there is no cointegration. Thus if $\boldsymbol{\theta}$ is known, the hypothesis of no cointegration (or instability) can be tested with a t-statistic, as discussed by Kremers *et al.* (1992). However, in general this will not be the case. From the definition of $\boldsymbol{\pi}$, it is easily seen that $\lambda = 0$ implies $\boldsymbol{\pi} = 0$. Hence the null hypothesis may be tested by a Wald-type test, eventually corrected for the number of estimated parameters:

$$\text{Wald} = \hat{\boldsymbol{\pi}}'(\hat{V}[\hat{\boldsymbol{\pi}}])^{-1}\hat{\boldsymbol{\pi}} = (T - n(p+1))\frac{RSS_r - RSS_u}{RSS_u}, \tag{2.6}$$

where $\hat{\boldsymbol{\pi}}$ is the ordinary least-squares (OLS) estimator of $\boldsymbol{\pi}$ in (2.5), $\hat{V}[\hat{\boldsymbol{\pi}}]$ is the corresponding OLS covariance matrix estimator, and RSS_u and RSS_r denote the unrestricted and restricted ($\boldsymbol{\pi} = 0$) residual sum of squares, respectively. When the null hypothesis is rejected one has found a stable cointegrating relationship. The parameter estimates of λ and $\boldsymbol{\theta}$ can then be found from the estimate of $\boldsymbol{\pi}$ after suitable transformation. In Boswijk (1991) it is shown that the asymptotic distribution of the Wald statistic under the null hypothesis is a generalization of the distribution of the squared Dickey–Fuller t-statistic. Asymptotic critical values are obtained via simulation and tabulated in Boswijk (1991, tables A1–A3). Furthermore, a generalization for the case where $r > 1$ is presented, based on a system of simultaneous error correction equations in structural form.

In contrast with the above procedures, the maximum likelihood approach developed in Johansen (1988) and extended in Johansen and Juselius (1990) starts off from the statistical model that generates cointegration, viz. the VAR model in error correction form

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{x}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1}\Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{\varepsilon}_t. \tag{2.7}$$

where the disturbances are assumed to be normally distributed. The procedure does not necessitate any exogeneity assumptions nor a restriction on the value of r. It can be shown that, for fixed r, the maximum likelihood estimator for $\boldsymbol{\beta}$ in model (2.7) defines the linear combinations of $\boldsymbol{x}_{t-1}$ that correspond to the r largest squared canonical correlations of $\boldsymbol{x}_{t-1}$ with $\Delta \boldsymbol{x}_t$, corrected for lagged differences and deterministic variables. The test procedure of $r-1$ versus r is given by testing the significance of the r-th canonical correlation with an LR test (the $\lambda_{\max}$ test, see Johansen and Juselius, 1990). Asymptotic critical values are presented in Johansen and Juselius (1990, tables A1–A3). In related simulation studies, see e.g. Reimers (1991), it emerges that the LR statistic should be corrected for the number of estimated parameters to obtain satisfactory size properties in finite samples. This is accomplished by multiplying the test statistic by a factor $(T-np)/T$.

III. A MONTE CARLO EXPERIMENT

3.1. *The Monte Carlo Design*

We consider a bivariate time series $\{\boldsymbol{x}_t = (y_t, z_t)'\}$, which may be cointegrated via a single cointegrating relationship, so $r = 1$ or 0. We assume that $\{z_t\}$ is exogenously generated by a random walk:

$$\Delta z_t = \varepsilon_t, \tag{3.1}$$

with $\{\varepsilon_t\}$ i.i.d. $N(0, 1)$. For y_t, conditional on z_t, we use two different data generating processes:

$$\begin{aligned} &(\mathrm{I})\ \ \Delta y_t = \lambda(y_{t-1} - z_{t-1}) + 0.5\Delta z_t + 0.6\Delta y_{t-1} + \eta_t, \\ &(\mathrm{II})\ \ \Delta y_t = \lambda(y_{t-1} - z_{t-1}) + \eta_t - 0.6\eta_{t-1}, \end{aligned} \tag{3.2}$$

with $\{\eta_t\}$ i.i.d. $N(0, 1)$, independently of $\{\varepsilon_t\}$, and $\lambda \in \{0, -0.2\}$. The dynamics of DGP I is of the autoregressive type, whereas DGP II is an ARMAX model. If $\lambda = 0$, then y_t and z_t are not cointegrated, and even independent in case II. If $\lambda = -0.2$, the series are cointegrated with cointegrating vector $(1, -1)'$. The values of the parameters in (3.2) are chosen in order to generate reasonably apparent dynamic properties. The signal-to-noise ratio is somewhat arbitrary, and corresponds to a range of $(0.25, 0.55)$ for the coefficient of determination in (3.2).

For I and II (and $\lambda = 0, -0.2$) we generate 1,000 samples of 58 observations on $\{y_t, z_t\}$; each sample is constructed from 59 observations on $\{\varepsilon_t, \eta_t\}$, using the initialization methods described in Kiviet (1986, pp. 258–59). The first 8 observations are used only to create the necessary lags, leaving a sample size of 50 observations for estimation and testing purposes. Drawings from the standard normal distribution are performed by the function RNDN of the GAUSS statistical package, with a seed of 1590507862.

For each sample we calculate the ADF t-test statistics in a residual AR(p) model, the Wald statistic in an ADL($p;p,..,p$) model, and the likelihood ratio statistic in a VAR(p) model, with p ranging from 1 to 8. A lag length of, e.g., 8 is not often considered in samples as small as 50 observations, but we use this range for p to illustrate the influence of dynamic specification more clearly. In all regressions a constant term is included to avoid starting value dependencies. With these results, rejection frequencies of the tests at nominal 5 percent and 10 percent significance levels are computed. These yield empirical size estimates for $\lambda = 0$. For $\lambda = -0.2$, they provide an indication of the power of the tests. In order to keep as close as possible to the empirical practice, we do not estimate the empirical power by the rejection rates at the actual size. Although one could use repeated sampling methods to obtain approximate critical values, this does not correspond to the current practice.

For the cases where $\lambda = -0.2$ we also compute the Lagrange-multipler F-test (LMF, see Kiviet, 1986) for first-order serial correlation and the Schwarz criterion (SC). In case of the VAR model a variable addition F-test for the vector of residuals lagged one period is used to test for first-order VAR or VMA (vector moving average) disturbances. For each sample we choose a value of p by the following three distinct criteria. Firstly, we choose the most parsimonious mdel that has an insignificant LMF statistic at the 5 percent level. Secondly, the model with the lowest SC is selected, and thirdly, we choose only from the models with insignificant LMF statistic the model with the lowest SC. With these results we compute the model selection frequencies for each lag. We only investigate the performance of these model selection procedures for $\lambda = -0.2$, because the cointegrated models are of primary concern in practice.

3.2. *Simulation Results*

The results are reported in Tables 1 and 2. To facilitate the evaluation and comparison of the test performances, we have represented the size and power results at the 5 percent nominal level and also the model selection results graphically in Figures 1 and 2 for DGP I and II, respectively.

First, we discuss the results for DGP I. From Figure 1(a) we see that the empirical size of the ADF test is smaller than five percent, and hence the test performs quite well in this respect. The rejection frequency under the alternative (henceforth RFA) is very low at one lag, reaches its maximum at two lags, and subsequently decreases with larger lag lengths. The results for

TABLE 1

Simulation Results for DGP I

(a) Rejection frequencies under H_0 (no cointegration) and H_a (cointegration) of the ADF test and AR lag length selection frequencies

	5% level		10% level		Selection frequency		
p	H_0	H_a	H_0	H_a	LMF	SC	LMF&SC
1	0.021	0.025	0.046	0.062	0.136	0.091	0.090
2	0.034	0.578	0.077	0.739	0.783	0.756	0.748
3	0.034	0.561	0.071	0.715	0.075	0.113	0.121
4	0.037	0.481	0.072	0.622	0.006	0.028	0.029
5	0.033	0.352	0.068	0.501	0.000	0.007	0.007
6	0.029	0.269	0.057	0.408	0.000	0.002	0.002
7	0.031	0.190	0.057	0.318	0.000	0.003	0.003
8	0.029	0.147	0.067	0.277	0.000	0.000	0.000

(b) Rejection frequencies under H_0 (no cointegration) and H_a (cointegration) of the Wald test and ADL lag length selection frequencies

	5% level		10% level		Selection frequency		
p	H_0	H_a	H_0	H_a	LMF	SC	LMF&SC
1	0.280	0.195	0.335	0.268	0.008	0.012	0.004
2	0.080	0.716	0.136	0.841	0.935	0.937	0.913
3	0.084	0.528	0.135	0.684	0.056	0.037	0.069
4	0.076	0.408	0.119	0.540	0.001	0.011	0.011
5	0.067	0.290	0.123	0.429	0.000	0.001	0.001
6	0.070	0.239	0.127	0.354	0.000	0.002	0.002
7	0.078	0.169	0.133	0.283	0.000	0.000	0.000
8	0.077	0.142	0.124	0.235	0.000	0.000	0.000

(c) Rejection frequencies under H_0 (no cointegration) and H_a (cointegration) of the LR test and VAR lag length selection frequencies

	5% level		10% level		Selection frequency		
p	H_0	H_a	H_0	H_a	LMF	SC	LMF&SC
1	0.259	0.120	0.328	0.177	0.017	0.067	0.016
2	0.069	0.523	0.126	0.667	0.915	0.920	0.914
3	0.067	0.324	0.119	0.458	0.057	0.011	0.059
4	0.062	0.212	0.115	0.336	0.006	0.002	0.006
5	0.053	0.139	0.097	0.250	0.002	0.000	0.002
6	0.046	0.112	0.091	0.212	0.001	0.000	0.001
7	0.055	0.085	0.105	0.152	0.002	0.000	0.002
8	0.050	0.071	0.102	0.129	0.000	0.000	0.000

Note: Critical values are from Phillips and Ouliaris (1990), table IIb, for the ADF test; from Boswijk (1991), table A2, for the Wald test; and from Johansen and Juselius (1990), table A2, for the LR test. For the LMF tests a 5 percent significance level is used. The table is based on 1,000 replications and a sample size of 50.

TABLE 2

Simulation Results for DGP II

(a) Rejection frequencies under H_0 (no cointegration) and H_a (cointegration) of the ADF test and AR lag length selection frequencies

	5% level		*10% level*		*Selection frequency*		
p	H_0	H_a	H_0	H_a	*LMF*	*SC*	*LMF&SC*
1	0.780	0.733	0.839	0.843	0.826	0.775	0.759
2	0.341	0.254	0.449	0.395	0.165	0.186	0.201
3	0.149	0.114	0.227	0.203	0.009	0.034	0.033
4	0.098	0.071	0.167	0.155	0.000	0.005	0.006
5	0.065	0.052	0.117	0.114	0.000	0.000	0.001
6	0.043	0.037	0.086	0.074	0.000	0.000	0.000
7	0.041	0.031	0.074	0.065	0.000	0.000	0.000
8	0.036	0.028	0.073	0.063	0.000	0.000	0.000

(b) Rejection frequencies under H_0 (no cointegration) and H_a (cointegration) of the Wald test and ADL lag length selection frequencies

	5% level		*10% level*		*Selection frequency*		
p	H_0	H_a	H_0	H_a	*LMF*	*SC*	*LMF&SC*
1	0.728	0.977	0.802	0.993	0.368	0.460	0.305
2	0.303	0.894	0.412	0.958	0.454	0.369	0.441
3	0.135	0.782	0.202	0.868	0.145	0.134	0.188
4	0.098	0.649	0.163	0.776	0.029	0.028	0.050
5	0.066	0.532	0.120	0.665	0.003	0.007	0.012
6	0.050	0.407	0.104	0.567	0.001	0.001	0.003
7	0.055	0.309	0.101	0.436	0.000	0.001	0.001
8	0.056	0.205	0.093	0.338	0.000	0.000	0.000

(c) Rejection frequencies under H_0 (no cointegration) and H_a (cointegration) of the LR test and VAR lag length selection frequencies

	5% level		*10% level*		*Selection frequency*		
p	H_0	H_a	H_0	H_a	*LMF*	*SC*	*LMF&SC*
1	0.616	0.884	0.714	0.949	0.545	0.812	0.538
2	0.190	0.694	0.304	0.843	0.361	0.158	0.354
3	0.081	0.550	0.137	0.693	0.079	0.029	0.092
4	0.054	0.420	0.109	0.578	0.015	0.001	0.016
5	0.041	0.310	0.084	0.457	0.000	0.000	0.000
6	0.033	0.205	0.075	0.338	0.000	0.000	0.000
7	0.033	0.140	0.070	0.251	0.000	0.000	0.000
8	0.029	0.106	0.072	0.186	0.000	0.000	0.000

Note: See Table 1.

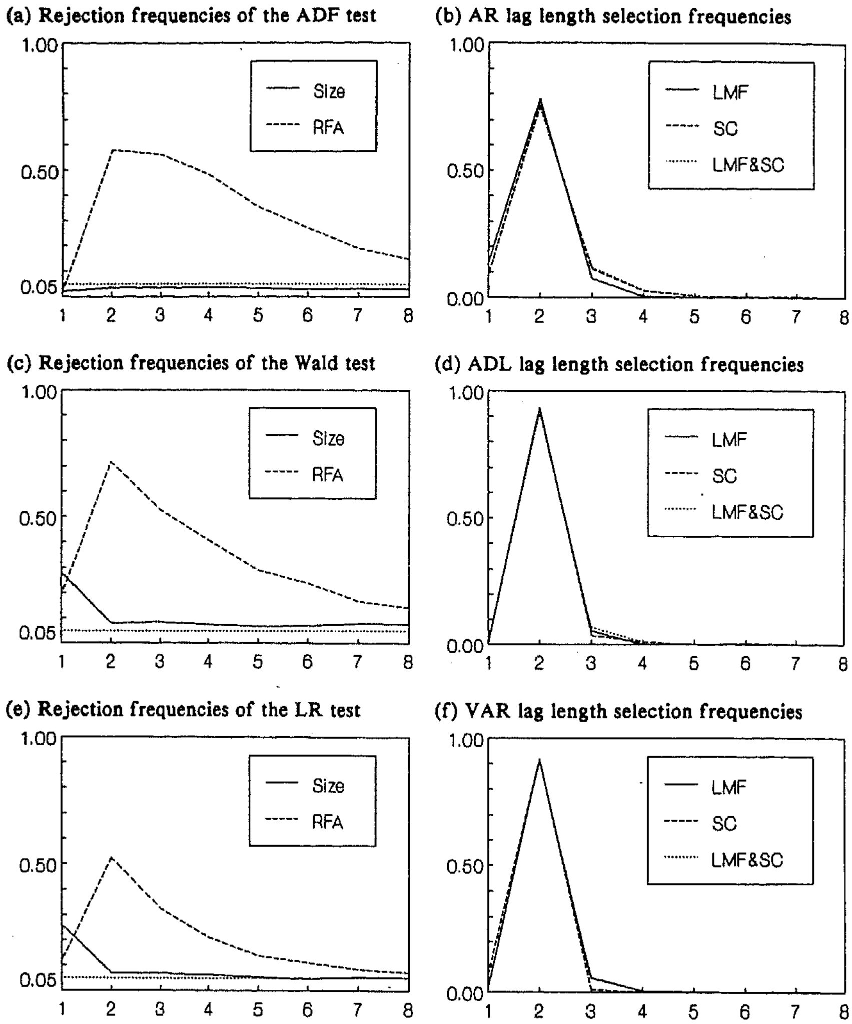

Fig. 1. Simulation results for DGP I (5% nominal significance level)

the 10 percent nominal significance level in Table 1(a) show the same pattern. Hence we conclude that a lag length of 2 is optimal for this procedure in this case. To a large extent the three model selection procedures lead to this optimal choice (see Figure 1(b)): in about 75 percent of the cases a model with two lags is selected.

The size and power characteristics for the Wald test are quite different from the ADF test. In the underparametrized ADL(1, 1) model, the empirical size is much too high (even higher than the power). Notice however that with our specification strategies this model is hardly ever selected. For all models

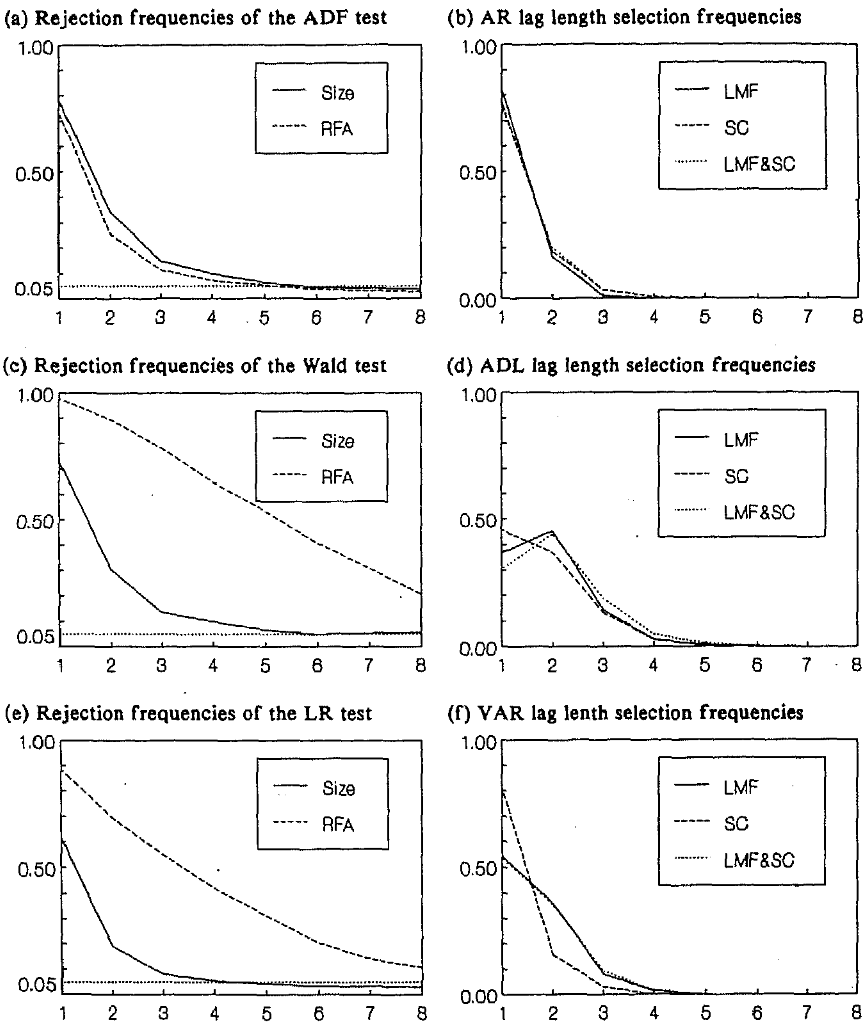

Fig. 2. Simulation results for DGP II (5% nominal significance level)

with $p > 1$, the empirical size is about 7.5 percent, which is still too high. Similarly, at a 10 percent nominal level the empirical size stabilizes at about 12.5 percent. However, with a sample size of only 50 and up to 18 explanatory variables we cannot expect much better results. Just as for the ADF test, the empirical power of the Wald test rises steeply if the lag length is raised from 1 to 2, and deteriorates with further lag increases. Again the optimal lag length is 2, which is quite adequately indicated by the model selection procedures; in over 90 percent of the cases this model is selected by all three models (see Table 1(b)). The size and power performance of Johansen's LR test is similar to that of the Wald test. As in the previous cases, all model selection procedures indicate the optimal lag length of 2 in most of the replications.

Summarizing the simulation results for DGP I, we note the following points. Firstly, a lag length of 2 is optimal in terms of size and power for all tests. For the Wald and LR tests this may be explained because the ADL(2, 2) and VAR(2) models encompass the true DGP; for the ADF test this is not the case, but a residual AR(2) model seems to mimic the (static residual) dynamics fairly adequately. A theoretical explanation for this phenomenon may be based on the analysis of Kremers *et al.* (1992). Secondly, the power of the tests is negatively influenced by both over- and underparametrization. Thirdly, we find that the three model selection strategies investigated all lead to the same optimal lag length of 2. We repeated the simulation experiments for DGP I for a sample size of 100. As expected, this yields better size and power properties for all tests, but other than that, the previous conclusions continue to hold.

We now turn to the simulation results for the second DGP. In this case, the MA effects imply that none of the models exactly encompass the true DGP. However, the dynamic properties of this process may well be approximated by higher order AR, ADL or VAR models.

From Figure 2(a) we see that up to 5 lags are required in the ADF auxiliary regression in order to obtain an empirical size reasonably close to the nominal level; for models with $p<5$, the size is much too large, with a maximum of 78 percent at one lag. The RFA of the ADF test is uniformly lower than the size for this DGP. Hence if the empirical significance levels are used, we may expect a complete loss of power. The same results emerge if a nominal significance level of 10 percent is used. All model selection criteria that we consider will prefer the (very inadequate) low-order model; the first-order model is selected in over 75 percent of the replications.

The size curve of the Wald test in Figure 2(c) closely resembles that of the ADF test: again, a low lag length leads to extreme overrejection under the null, and 5 lags are needed to avoid this size distortion. However, the RFA of the Wald test is substantially better: although it also decreases with the lag length, the RFA at five lags is still larger than 50 percent. Unfortunately, this optimal model is not indicated by the LMF serial correlation test or Schwarz criterion: for all three selection strategies, a lower dimensional model is selected in over 98 percent of the replications. Thus, although the ARMAX model may be approximated by a higher order ADL model, the selection of the appropriate model will be troublesome.

As with DGP I, the results for the LR test in a VAR model resemble those of the Wald test. Both the empirical size and the RFA decrease more rapidly, so that a lag length of 4 is now optimal. Although the shape of the model selection curves is somewhat different from the shapes for the other approaches, the same conclusion that they do not indicate enough lags applies here.

The results for the second DGP may be summarized as follows. For all test procedures, a lag length of about 5 is optimal, but this is rarely indicated by the model selection procedures that we have considered. Secondly, the properties of the ADF test disqualify it for practical use in this type of model.

IV. DISCUSSION

Our Monte Carlo experiment suggests the following conclusions about the effects on cointegration testing of dynamic (mis-) specification. An insufficient lag length can lead to substantial size distortions, and in particular to drastic overrejection of the null hypothesis. On the other hand, overparametrization of the dynamic structure will lead to a loss of power. This implies that for each DGP and test procedure there is exactly one lag length which is optimal. Standard model selection procedures do not indicate an appropriate lag length if the model provides only an approximation to the true dynamics. Because tests for cointegration are in fact tests for unit roots, these results can be expected to apply to univariate unit root testing as well, see e.g. Schwert (1989) and Kunst (1989).

The size distortion may be intuitively explained as follows. The cointegration approaches investigated consider the significance of correlations between error correction terms and first differenced variables after correction for deterministics and dynamics. If the correction is insufficient, there will be some serial correlation left in the corrected first difference. This may have a positive effect on the squared correlations, leading to overrejection of the null hypothesis. Another explanation of the overrejection in the second DGP is that our choice of the negative MA parameter causes a near-cancellation of the unit root factor if $\lambda = 0$, so that y_t may appear to be stationary (cf. Molinas, 1986).

The overrejection of the null hypothesis of no cointegration suggests the possibility of *spurious cointegration.* The effects of dynamic misspecification in our case in fact do resemble the well-known spurious regressions results of Granger and Newbold (1974). They found that the hypothesis of no relationship between two random walks, which are independent in reality, is rejected much too often. In a way, the notion of cointegration was a reaction to these findings, and the purpose of error correction models is to safeguard us against these spurious effects. We have seen that as yet this attempt has only been partly successful. Even if a proper misspecification strategy is employed, we still may reject the hypothesis of no long-run relationship too often in case of underparametrized models. Hence a significant empirical cointegrating relationship should be interpreted with some caution.

We would expect that the LMF test, which is supposed to test for serial correlation, should indicate the inappropriateness of models with too few lags. This is the case for DGP I, where the true dynamic structure is of the autoregressive type, but LMF clearly fails when the autoregressive models are used as an approximation of the true ARMAX model. A possible remedy for this problem is to start with a model which has both autoregressive and moving average terms, i.e., a (V)ARMA(X) model. However the Monte Carlo evidence of Schwert (1989) on unit root testing in ARMA models suggests that this method still suffers from size distortions and may be inferior to a high order autoregression. The instrumental variable methods of Hall (1989) may provide an alternative solution to these problems.

Comparing the three tests, it seems that the Wald test in an ADL regression is superior in this experiment. It has the highest power at two lags with DGP I and satisfactory size and power performance with DGP II at five lags. However, we should note that in both DGPs z_t is weakly exogenous; this property is exploited in the Wald procedure and not in the other two approaches. On the other hand, the exogenous process that we have chosen fits neatly in a VAR model. Because the Wald test does not require modelling of the exogenous process, it can be expected to be less sensitive to departures from this assumption, such as general ARIMA processes with arbitrary error distributions. For the residual ADF test we note that its results for the second DGP seem to limit its practical usefulness.

Department of Actuarial Science and Econometrics, University of Amsterdam
Econometric Institute, Erasmus University Rotterdam

REFERENCES

Boswijk, H. P. (1989). 'Estimation and Testing for Cointegration with Trended Variables: A Comparison of a Static and a Dynamic Regression Procedure', Report AE 12/89, University of Amsterdam.

Boswijk, H. P. (1991). 'Testing for Cointegration in Structural Models', Report AE 7/91, University of Amsterdam.

Engle, R. F. and Granger, C. W. J. (1987). 'Cointegration and Error Correction: Representation, Estimation, and Testing', *Econometrica*, Vol. 55, pp. 251–76.

Franses, P. H. (1990). 'A Note on Spurious Cointegration', *Tinbergen Institute Research Bulletin*, Vol. 2, pp. 249–52.

Granger, C. W. J. (1981). 'Some Properties of Time Series Data and their Use in Econometric Model Specification', *Journal of Econometrics*, Vol. 16, pp. 121–30.

Granger, C. W. J. and Newbold, P. (1974). 'Spurious Regressions in Econometrics, *Journal of Econometrics*, Vol. 2, pp. 111–20.

Hall, A. (1989). 'Testing for a Unit Root in the Presence of Moving Average Errors', *Biometrika*, Vol. 76, pp. 49–56.

Johansen, S. (1988). 'Statistical Analysis of Cointegration Vectors', *Journal of Economic Dynamics and Control*, Vol. 12, pp. 231–54.

Johansen, S. and Juselius, K. (1990). 'Maximum Likelihood Estimation and Inference on Cointegration — with Applications to the Demand for Money', *BULLETIN*, Vol. 52, pp. 169–210.

Kiviet, J. F. (1986). 'On the Rigour of Some Misspecification Tests for Modelling Dynamic Relationships', *Review of Economic Studies*, Vol. 53, pp. 241–61.

Kremers, J. J. M., Dolado, J. J. and Ericsson, N. R. (1992). 'The Power of Cointegration Tests', *BULLETIN*, this issue.

Kunst, R. M. (1989). 'On the Influence of Order Selection on Correlation Analysis in Autoregressive Processes', Manuscript, Institute for Advanced Studies, Vienna.

Molinas, C. (1986). 'A Note on Spurious Regressions with Integrated Moving Average Errors', *BULLETIN*, Vol. 48, pp. 279–82.

Phillips, P. C. B. and Ouliaris, S. (1990). 'Asymptotic Properties of Residual Based Tests for Cointegration', *Econometrica*, Vol. 58, pp. 165–93.

Reimers, H.-E. (1991). 'Comparisons of Tests for Multivariate Cointegration', forthcoming in *Statistical Papers*.

Schwert, G. W. (1989). 'Tests for Unit Roots: A Monte Carlo Investigation', *Journal of Business & Economic Statistics*, Vol. 7, pp. 147–59.