Incorrectly specified lags in cointegrating regressions

by
Virmantas Kvedaras
Vilnius university, Department of Economics
(e-mail: virmantas.kvedaras@ef.vu.lt)

Abstract

The finite sample effects of incorrectly shifting the cointegrated variables in time is studied and the consequences for testing for cointegration using the Dickey-Fuller test is discussed. It is shown analytically that misspecifications creates autocorrelated errors in cointegrating equation that can be substantial even for quite small incorrect lag choices. The error dynamics due to lagged variables can be simply eliminated by shifting them in time, therefore the dynamics of the error term and asocciated biases in parameter estimates can be reduced. The simple test to detect incorectly lagged cointegrated variables is proposed. The finite sample effect on the parameter estimates and the power of the Dickey-Fuller test when it is applied to test for stationarity of the cointegrating regression error is explored by Monte Carlo simulations. The empirical example with the potential lag shift in the cointegrating regression is presented using the Lithuanian investment and export data.

1. Introduction

In principle the definition of cointegrated variables does not eliminate possibility of lags in cointegrating relation, however, presently, contemporaneously cointegrated variables are at the centre of theoretical investigations and practical applications. The probable reason of the dominating role of contemporaneous variables is that cointegration is associated with some equilibrium relationship in economic terms. In static equilibrium the time dimension is not present so that only contemporaneous relationship is important. However this does not eliminate possibility that variables are related with lags in general, these just become unimportant in equilibrium. Therefore it is not so evident that only contemporaneous relationship should be analyzed. Due to adjustment costs, inferences obscured in seasonal fluctuations, absence of instantaneous signal transmissions, capacity limits and other causes there are good economic and other reasons to expect some lags in cointegrating relationships and it is not evident, why this should not hold in the long-run as well.

The aggregation level has an impact for the discussed possible mismatching of timing in modelled cointegrating relationships. If data are sufficiently aggregated in time (for instance, yearly or even more aggregated), there might be no need to account for lagged relationships, because they are dominated by aggregated contemporaneous properties. However in some situations, for instance, modelling economies in transition, practically the only solution for econometric modelling is to use more frequent then yearly data, otherwise data sample reduces to around six or seven observations. Besides that currently more and more disaggregated data are available and with more frequent observations lagged relationships become relevant given that the increase in sample size does not counteract this negative effect.

Inspired by these motivations in this article we study the effect of incorectly specified lags in a single equation cointegration case when Dickey-Fuller (DF) tests is used for testing the stationarity of cointegrating regression errors. The following section presents some analytical result and questions, whether and in what cases lag misspecification is important. Then using the Monte Carlo simulations with the cointegrating regression with i.i.d. residuals and exogenous regressor the third section address the question how much it is important in terms of the precision of the parameter estimate and the power of the test. The fourth section presents an empirical example where time shifts in cointegration seems to be present and the fifth section concludes.

2. Lag shifts in cointegrating regression

2.1. Contemporaneous versus lagged cointegration

Let y_t and x_t be cointegrated as defined by following data generating process (DGP)

$$y_t = c + \beta x_{t-k} + u_t, u_t \sim i.i.d.(0, \sigma_u^2), \quad (1a)$$

$$\Delta x_t = \zeta_t, \zeta_t \sim i.i.d.(0, \sigma_\zeta^2), \quad (1b)$$

where $\beta, \sigma_u^2, \sigma_\zeta^2$ are the respective parameters, Δ denotes the first difference operator, and $k > 0$ denotes the order of lag and we assume that u_t and ζ_t are independent. Suppose that the conjectured testing equation is as follows

$$y_t = c + \beta x_t + v_t, \quad (2)$$

where v_t is assumed to be stationary. Note that the errors of the true relationship (1a) and the conjectured one in (2) are related by

$$v_t = u_t - \beta \sum_{i=0}^{k-1} \Delta x_{t-i}. \quad (3)$$

By definition u_t is stationary and stationarity of v_t therefore depends only on the last term in (3), which for finite lag k and infinite process $\{x_t\}_{t=1}^{\infty}$ is stationary as well. Therefore for infinite process there is no difference whether we use the definition of the process as in (1a) or in (2). In empirical applications however we deal with finite samples an increase in k relatively to T means that the sum of Δx_t is approaching the realization of integrated variable x_t. Consequently, with increasing shift parameter k, v_t becomes more alike a nonstationary variable. This transfer of dynamics then can obscure estimation of parameters of the static regression (2) and the procedure of inference about cointegration.

2.2. The impact on error dynamics

It is well known (see Banerjee et al.,1986) that omitting the error dynamics and estimating the static cointegrating regression (2) with ordinary least squares (OLS) result in significant biases in small samples. Nevertheless, the popular single equation Dickey Fuller or augmented Dickey Fuller (ADF) tests rely on the procedure first estimating the static regression (2) and then testing for stationarity of the residual term. It is important then to avoid the error term dynamics that can be simply eliminated by correctly specifying the timing of cointegrating variables. In this section we study the lag shifts impact on the error dynamics of the static equation (2) and draw some conclusions about the effect for the DF test.
Assume that the DGP for the process under investigation is as defined in (1), and suppose that (2) is the conjectured test equation. DF test then would test for stationarity of the residuals of the conjectured cointegrating relationship. Test procedure would differ, depending on whether β is assumed to be known or not, there are as well different other peculiarities of choosing the test form, however we abstract from these problems and following assume that β and the form of the relationship up to a lag shift are known (for DF and ADF statistic description and further analysis see Hamilton, 1994).
The cointegration test using DF statistic then reduces to testing for stationarity of v_t, where the hypotheses $H_0 : \rho = 1,\ H_1 : \rho < 1$, i.e. the null is that v_t is nonstationary, and the test is conducted using the test equation

$$v_t = \rho v_{t-1} + \varepsilon_t, \quad (4)$$

where ε_t is assumed to be i.i.d.
Because we assumed that variables are cointegrated as defined in (1) following we study the shift effects

under alternative hypothesis of stationarity of v_t. In this section we establish how the lag shifts affects the value of ρ and the variance of the residual term of the test equation under the alternative hypothesis and in the later sections using simulations we evaluate how these affect the precision of parameter estimates and the power of the test for certain DGP.

First note that when there is no shift and $k = 0$ the relationship in (1) under alternative of cointegrated variables implies that $v_t = u_t$ and $\rho = \frac{E(u_t u_{t-1})}{E(u_t^2)} = 0$, with $\varepsilon_t = u_t$ respectively. Now assume that there is a shift of one lag and $k = 1$, then from (3) and (1b) $v_t = u_t - \beta\zeta_t$ and

$$\rho = \frac{E(v_t v_{t-1})}{E\left(v_t^2\right)} = \frac{E[(u_t - \beta\zeta_t)(u_{t-1} - \beta\zeta_{t-1})]}{\sigma_u^2 + \beta^2\sigma_\zeta^2} = 0.$$

So that the value of $\rho = 0$ is not affected, however from (4) then we have $\varepsilon_t = u_t - \beta\zeta$ meaning that the error variance, as given in the denominator of the middle term in equation above, is increasing and inference about the parameter estimate is less efficient compared with the correctly specified case. Additionally note that although explanatory variable is exogenous as defined in (1), incorrect lag specification induces endogeneity problem as can be seen from (3).

For $k \geq 2$ from (3) and (1b) follows that $v_t = u_t - \beta\sum_{i=0}^{k-1}\zeta_{t-i}$ and

$$\rho = \frac{E(v_t v_{t-1})}{E\left(v_t^2\right)} = \frac{(k-1)\beta^2\sigma_\zeta^2}{\sigma_u^2 + k\beta^2\sigma_\zeta^2}, \quad (5)$$

that holds as well for the case $k = 1$, however for $k \geq 2$ the expression indicates that under alternative of cointegrated variables the autocorrelation coefficient $\rho \neq 0$. It increases with bigger k, and becomes closer and closer to 1 that is just the value of the correlation coefficient under the null hypothesis of the test for nonstationarity. For instance for the parameter values that are used in later simulations $\beta = \sigma_u^2 = \sigma_\zeta^2 = 1$ and with $k = 2$ and $k = 4$, say, $\rho = \frac{1}{3}$ and $\frac{3}{4}$ respectively. Besides that with the increasing k the dynamics of the error becomes more complicated with nonzero higher order autocorrelations. Consequently, the misspecifications of k can create substantial error dynamics and negatively affect the static least squares performance in finite samples as well as the power of the test, in the later case even when assuming that β is known instead of estimating it.

2.3. Testing for lag shifts

The relationship between the errors of correctly specified and shifted specifications (3) defines the effect of misspecification. It can be used in several ways, but the most straightforward is to put (3) into (2) that gives the cointegrating regression augmented with the contemporaneous and lagged values of explanatory variables. However to distinguish between dynamics induced by the misspecification of lagging and due to other reasons first obtaining residuals of (2) and then applying ARMAX where explanatory variables are the dynamic terms suggested by (3) could be preferred. In later simulations we deal with the first approach as we use i.i.d. errors u_t and it enables direct comparison of the parameter estimates of β. Therefore we estimate following augmented test equation

$$y_t = \beta x_t - \sum_{i=0}^{k-1}\beta_i \Delta x_{t-i} + u_t, \quad (6)$$

Note that besides the significance of the parameters of dynamic terms they should not be significantly different from the parameter of the level term with a negative sign. Therefore testing procedure could be split into two steps. First the significant dynamic terms in (6) are established and then it is tested whether they do not differ significantly among themselves and whether they are equal to the level parameter with a negative sign. If both criteria are satisfied then data do not contradict the model in (6) and respectively the hypothesis about the lags in cointegrating regression. The order of the shift in the original cointegrating regression then is straightforwardly indicated by the significant dynamic terms, i.e. the significant terms plus one. We present an example of the described test in section 4 after discussing the simulation results in section 3.

3. Small sample effects of shifts on parameter estimates and DF test power

In this section we study the effect of lag shifts in the cointegrating regression on the precision of parameter estimates and the power of the DF test. Because we showed that the shifts results in the dynamics of the error our simulations regarding the effects on the parameter estimates are closely related to Banerjee et al. (1986) results, however we study particularly the effect that arrises from incorrect timing in cointegrating regression rather than autocorrelated error in general. Additionally we report the effects on the power of the DF test when the two step Engle-Granger procedure is used to test for stationarity. Namely, first the test equation (2) with $c = 0$ is estimated by ordinary least squares and then a bit modified version of equation (4)

$$\Delta v_t = \delta v_{t-1} + \varepsilon_t,$$

is used to obtain OLS estimate of parameter ρ in order to test the hypothesis $H_0 : \delta = 0,\ H_1 : \delta < 0$. The DGP is as defined in (1) with fixed values of $c = 0, \sigma_u^2 = 1$, and $\sigma_\zeta^2 = 1$ and normally distributed u_t. We report the results for some different values of the cointegrating parameter β for several sample sizes T. The results are obtained by using 1000 repetitions with the EViews 4.1 package and therein built in random number generator. In Table 1 $\hat{\beta}_a$ denotes the average of the 1000 estimates of the parameter β, DF denotes the power calculated for the 5% significance level when β differs from zero and the respective size when $\beta = 0$. Because shifting lags creates certain dynamics we could expect that augmentations could solve the problem, therefore besides the DF test in the table we report as well the power of the ADF test, where augmentations where chosen based on Akaike information criteria.

Table 1The simulated lag shifts effects on static cointegrating regression parameter estimates and the power of DF and ADF tests

		$k = 0$			$k = 1$			$k = 2$		
	T	$\hat{\beta}_a$	DF	ADF	$\hat{\beta}_a$	DF	ADF	$\hat{\beta}_a$	DF	ADF
$\beta = 1$	30	0.996	0.999	0.958	0.824	0.99	0.943	0.677	0.801	0.749
	50	0.999	1	0.99	0.896	1	0.988	0.813	0.995	0.975
	100	1	1	0.999	0.946	1	1	0.897	1	0.997
	200	1	1	1	0.975	1	1	0.949	1	1
$\beta = 0.7$	30	0.699	0.995	0.954	0.586	0.994	0.942	0.345	0.954	0.906
	50	0.700	1	0.997	0.635	1	0.995	0.398	1	0.996
	100	0.700	1	1	0.662	1	1	0.450	1	1
	200	0.701	1	1	0.681	1	1	0.475	1	1
$\beta = 0$	30	0.039	0.076	0.104	0.022	0.071	0.095	−0.025	0.07	0.122
	50	0.032	0.069	0.084	−0.014	0.066	0.081	0.016	0.068	0.088
	100	−0.016	0.066	0.073	−0.007	0.058	0.058	0.018	0.058	0.059
	200	0.001	0.058	0.062	0.004	0.04	0.043	0.001	0.055	0.055

Table 1 cont.

		$k = 3$			$k = 4$			$k = 12$		
	T	$\hat{\beta}_a$	DF	ADF	$\hat{\beta}_a$	DF	ADF	$\hat{\beta}_a$	DF	ADF
$\beta = 1$	30	0.559	0.502	0.44	0.453	0.366	0.324	0.001	0.624	0.352
	50	0.713	0.882	0.852	0.647	0.68	0.612	0.197	0.462	0.404
	100	0.844	1	0.996	0.804	0.998	0.978	0.469	0.48	0.25
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

$\beta = 0.7$	30	0.282	0.862	0.782	0.216	0.772	0.676	0.017	0.778	0.43
	50	0.352	0.996	0.952	0.311	0.982	0.922	0.103	0.806	0.668
	100	0.427	1	1	0.399	1	0.992	0.229	0.978	0.54
	200	0.459	1	1	0.451	1	1	0.351	1	0.772
$\beta = 0$	30	−0.004	0.091	0.131	−0.019	0.065	0.125	−0.037	0.089	0.225
	50	−0.005	0.071	0.079	0.008	0.07	0.08	0.006	0.048	0.064
	100	−0.016	0.065	0.067	0.004	0.052	0.056	−0.001	0.059	0.065
	200	−0.032	0.056	0.066	0.038	0.054	0.057	−0.022	0.041	0.043

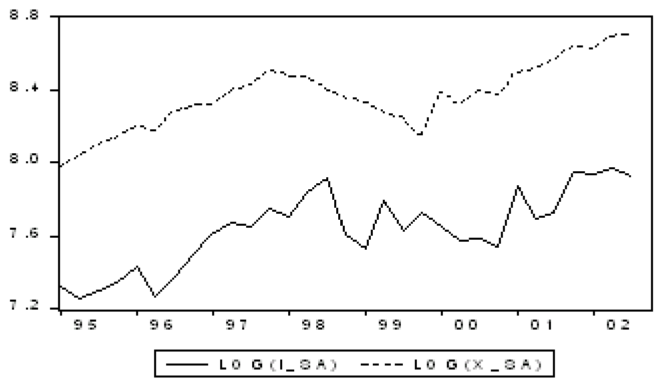
Several conclusions can be drawn from these simulations. First, as was predicted when lag shift $k = 1$ there is practically no loss in power, however the loss in efficiency of the parameter estimates for $\beta \neq 0$ is seeming and becomes quite substantial for small samples, i.e. up to around one fifth. Second, as can be expected, the most vulnerable are very small samples and for high lag misspecifications it becomes substential, e.g. for $T = 30$ and $k = 4$, which is quite possible yearly mismatching when dealing with quarterly data, the avarage of estimated parameter is more than halved and the power becomes only around 37 percents instead of almost perfect power performance. Recalling that here we deal with the simple case and the low power performance of the unit root tests under certain circumstances is well known this suggests that in small samples one must do as much as possible to avoid the error dynamics and particularly induced by possible lagged relationships. Third, the augmentation with lags does not solve the problem created by shifted variables.

As these simulations indicate there are good reasons to account for error dynamics in small samples and, if the problem of dynamics can be solved simply shifting the cointegrating variables, it seems that such procedure is significant.

4. Empirical application

In this section we provide an example where the established results seems to apply in practice. We analyze the relationship between Lithuanian real investment and export data. Analysis of the relationship between investment and export variables is inspired by the small open nature of the Lithuanian economy, where export led growth of economy seems to have firm grounds as was showen by earlier experience and particularly Russian crisis. Besides quite crucial role as a demand factor - about 70 percents of manufacturing production is exported - export additionally plays an important role in financing investments, because in Lithuania only ten percents of investments are financed from borrowed means. Therefore it is quite natural to expect that developments in exports will have an effect on the investments. It is not clear whether one should look however at the impact of export on investment directly or through the domestic production, due to high degree of openness however we assume that there is direct effect of exports on investments. This seems to be supported by the visual inspection of the seasonally adjusted - the Census X-12 multiplicative method was used - data plot (see Figure 1), as well the contemporaneous static least squares regression results are quite satisfactory (see Table 2 in appendix A), with marginal first order residual correlation (correlation coefficient value is 0.31 and its empirical significance level is 0.07), but with the joint hypothesis about the significance of the first six autocorelation lags rejected with the empirical significance level 0.27 (the Ljung-Box Q_6 was used for testing).

Figure 1 Dynamics of seasonally adjusted imports and exports (1995:1 - 2002:3 quarterly data)



The DF test of the residuals - AIC and SC criteria did not suggest inclusion of additional augmentations - rejects the null of nonstationarity for test formulation with a constant (the value of the test statistic is -3.85). So that there seems good grounds to estimate contemporaneous cointegrating vector. The inspection of CUSUM plot reveals however possible problems with stability of parameters (see Figure 2 in appendix B).

Despite quite satisfactory results of the preceding analysis we should remember that investment decision process usually takes time, therefore we could expect that there are some lags in export-investment relationship and not accounting for them could result in biased estimates of the parameters, e.g. due to error dynamics. Therefore we estimated the extended regression suggested by section 2.3. As can be seen from Table 3 in Appendix A there are two terms - we checked for other lags as well, but they were seemingly not significant - that closely satisfy the predictions made in section 2.3, when lag shifts are present. Namely, the parameter estimates of included dynamic terms are negative and, although here distribution of the test statistic due to cointegrating regression is nonstandard, the estimates of the parameters of the dynamic terms clearly does not differ significantly from the contemporaneous parameter estimate of the level term and each other as well. This seems to suggest that there might be a time shift in the formulated cointegrating relationship. Therefore we estimated the static cointegrating regression with two lags between variables as suggested by the significant dynamic terms. The estimation results (see Table 4 in Appendix A) are more satisfactory than earlier with contemporaneous variables. There are no standard problems, e.g. the first order autocorrelation empirical significance level now equal to 0.88 and Ljung-Box Q_6 to 0.49, and this specification is even preferable to that one with dynamic terms as suggested by the adjusted coefficient of determination, as well AIC and SC criteria. The unit root test statistic for the residuals nonstationarity increases as well to -4.71. More than that the CUSUM plot now (see Figure 3) does not show any significant departure from the assumption about stability of parameters.

It is interesting to note that assuming two lag shift, the first order autocorrelation induced by the misspecification of timing calculated using the equation in (5) and based on the estimates of the "correct" specification with lagged variables is equal to 0.26. This quite closely corresponds to that one of earlier mentioned 0.31 and estimated directly from potentially misspecified model. And although in this example the cointegration was found already initially with contemporaneous variables, as simulations in earlier section show there might be cases where disregarding time lags can result in incorrect conclusion about the cointegrating relationship, and, probably, better parameter estimates were obtained by regarding the lag structure in cointegrating relationship.

5. Conclusions

Currently the cointegration mainly is treated as contemporaneous phenomenon. However in the world of inevitable disequilibrium there are good economic and other reasons to expect some lagged relationships to hold even in the long run. In infinite processes cointegration with lags is not a reasonable concept, because it always can be transformed into contemporaneous relationship. In the real world however we deal with finite samples and capturing some lagged relationships might be sensible, because modelling contemporaneous cointegration instead of lagged one transfers the dynamics into error term. As the simulations show the influence of such transfer of dynamics in small samples can be quite substantial. And because there is straightforward possibility to test whether dynamics is due to shifted cointegrating relationship or due to other reasons it seems important to test for the origin of the dynamics and to diminish the biases in parameter estimates and to increase the efficiency as well as power of some cointegration tests.

This paper studied only the very limited case with i.i.d. errors, however the presented empirical example indicates that even such cases have applications in the real world. The extensions however of the analysis are needed for more general error processes, multivariate cointegration case, as well possibly more richer dynamic relationships than rather simple lag shift in time.

References

Banerjee, A., J.J. Dolado, D.F. Hendry, G.W. Smith (1986), Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence, Oxford Bulletin of Economics and Statistics, 48, 253-277.

Hamilton J.D. (1994), Time Series Analysis, Princeton University Press.

Appendix A

Hereafter I and X denote respectively investment and export variable. _SA signifies seasonally adjusted variables.

Table 2The estimated cointegrating regression with contemporaneous variables

Dependent Variable: LOG(I_SA)
Method: Least Squares
Sample(adjusted): 1995:1 2002:3
Included observations: 31 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.424433	0.976876	-0.434480	0.6672
LOG(X_SA)	0.963417	0.116761	8.251185	0.0000

R-squared	0.701284	Mean dependent var	7.634043
Adjusted R-squared	0.690983	S.D. dependent var	0.212846
S.E. of regression	0.118320	Akaike info criterion	-1.368510
Sum squared resid	0.405988	Schwarz criterion	-1.275995
Log likelihood	23.21191	F-statistic	68.08205
Durbin-Watson stat	1.365640	Prob(F-statistic)	0.000000

Table 3The test regression for shifted variables in cointegration relationship

Dependent Variable: LOG(I_SA)
Method: Least Squares
Sample(adjusted): 1995:3 2002:3
Included observations: 29 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.229954	1.028723	-1.195612	0.2431
LOG(X_SA)	1.063971	0.122883	8.658400	0.0000
DLOG(X_SA)	-0.835849	0.289928	-2.882950	0.0080
DLOG(X_SA(-1))	-0.762572	0.282802	-2.696491	0.0124

R-squared	0.757018	Mean dependent var	7.657810
Adjusted R-squared	0.727860	S.D. dependent var	0.198425
S.E. of regression	0.103513	Akaike info criterion	-1.570806
Sum squared resid	0.267871	Schwarz criterion	-1.382213
Log likelihood	26.77668	F-statistic	25.96276
Durbin-Watson stat	1.903021	Prob(F-statistic)	0.000000

Table 4The cointegrating regression after regarding time shifts

Dependent Variable: LOG(I_SA)
Method: Least Squares
Sample(adjusted): 1995:3 2002:3
Included observations: 29 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.850992	0.959501	-0.886911	0.3830
LOG(X_SA(-2))	1.020062	0.115005	8.869680	0.0000

R-squared	0.744490	Mean dependent var	7.657810
Adjusted R-squared	0.735027	S.D. dependent var	0.198425
S.E. of regression	0.102140	Akaike info criterion	-1.658465
Sum squared resid	0.281682	Schwarz criterion	-1.564168
Log likelihood	26.04774	F-statistic	78.67121
Durbin-Watson stat	1.941972	Prob(F-statistic)	0.000000

Appendix B

Figure 2The CUSUM test plot for the initial cointegrating regression with contemporaneous variables

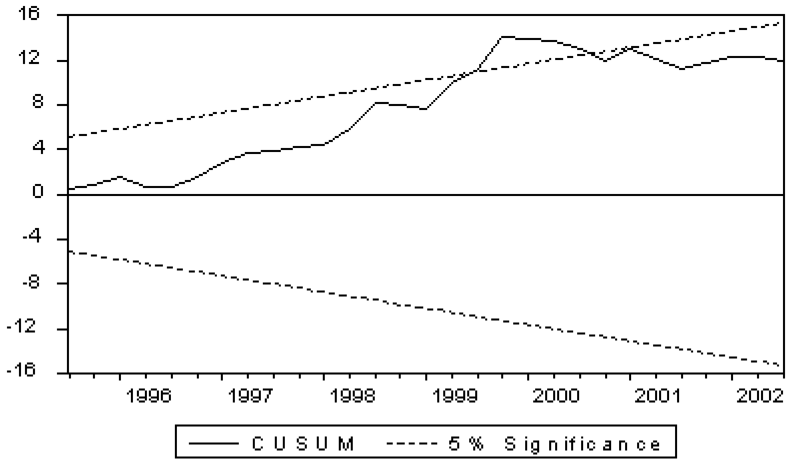
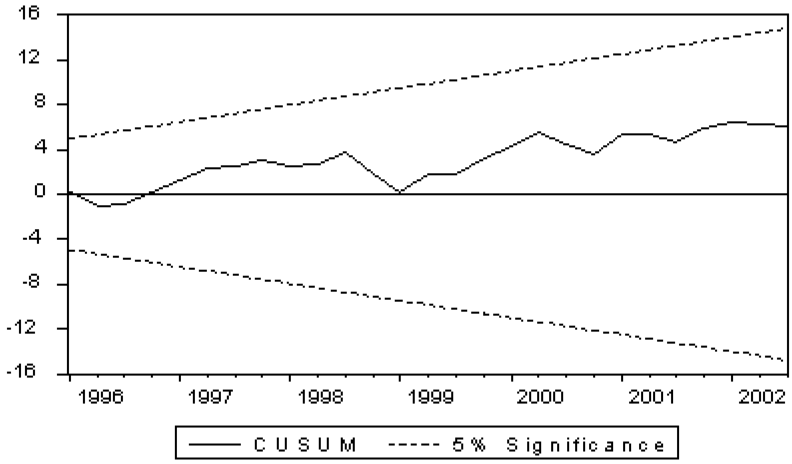


Figure 3The CUSUM test plot for cointegrating regression with lagged variables



This document created by Scientific WorkPlace 4.0.