Contemporary methods of assessment of cointegrating relations: an application to models of economic growth

Author: Mariya Neycheva

Abstract: The paper discusses a number of contemporary methods for assessment of cointegrating relations. Recent methods which allow for structural breaks in the data series are also considered. The method of Carrion-i-Silvestre and Sanso is applied in order to evaluate the impact of human capital on growth in the Bulgarian economy.

Keywords: cointegration, structural break, economic growth, human capital, Bulgaria

JEL: C3, E24, I25

1. INTRODUCTION

Cointegrating models are widely used to describe and estimate long-run economic relations such as the impact of production inputs on economic growth, fiscal sustainability, market demand function, consumption function, etc. The approach was introduced by Granger (1981; 1983) and was extensively developed in the paper of Engel and Granger (1987). The idea behind the cointegrated time series is to model the interaction between two variables (y, x) which are trending together. In this case the regression of one on the other is not spurious, but instead it tells something about the long-run relationship between them. Two nonstationary or random walk stochastic processes are said to be "cointegrated" if there exists a linear combination between them which is stationary i.e. I(0) despite the fact that each of them is integrated of order (1).

$$y(t) = a_0 + a_1 x(t) + u(t) \tag{1}$$

The regression of type (1) is known as cointegrating regression while the parameter a_1 is the cointegrating parameter. It represents the long-run, presumable stable, equilibrium relationship between the two stochastic processes denoted by *x(t)* and *y(t)*. The OLS estimation of eq. (1) produces a consistent estimate of the regression parameter a_1 which is actually "superconsistent" that is it tends to the true value faster than the usual OLS estimator[1]. The problem is that the t-statistics for a_1 does not necessarily have an approximate t-distribution.

[1] Prior to running the test for cointegration and solving the model a test for unit root should be applied to the regression variables.

The error term *u(t)* represents the so called "equilibrium error". It is used in linking the short-run dynamics of the dependent variable *y(t)* to its long-run value. The error-correction mechanism (ECM) corrects for disequilibrium. It regresses (see eq. 2) the change (first difference) of *y(t)* to the change of *x(t)* and the one-period lagged error term $\hat{u}(t-1)$ which resulted from estimation of eq. (1).

$$\Delta y(t) = b_0 + b_1 \Delta x(t) + b_2 \hat{u}(t-1) + \varepsilon_t \tag{2}$$

The term $\Delta\mathrm{x}(\mathrm{t})$ captures the short-run disturbances in the factor variable while the lagged error correction term u represents the adjustment toward the long-run equilibrium. The value of b_2 in eq. (2) which is expected to be negative displays how much of the discrepancy between the actual and the long- run value of y is eliminated or corrected over each time period (quarter, year, etc.). More explanations on the nature of cointegration could be found in the relevant literature (see, for example, Greene 2003; Gujarati 2004; Wooldridge 2013). The full mathematical description is beyond the scope of this paper.

The recent methods aim to overcome some of the difficulties surrounding the estimation of the cointegration models. The following lines present a summary of widely used assessment procedures in the studies on growth, namely the methods of Stock and Watson (1993), Gregory and Hansen (1996) as well as of Carrion-i-Silvestre and Sanso (2006). The latter two allow for a structural break in the data series.

2. CONTEMPORARY METHODS OF ASSESSMENT OF COINTEGRATING RELATIONS

Stock and Watson (1993) developed estimators for cointegration models which involve general I(d) processes with deterministic component. In the I(1) case with a single cointegrating vector, one should regress one of the variables on the levels of the remaining variables, leads and lags of their first differences ($\Delta\mathrm{x_i}(\mathrm{t}+\mathrm{j})$), and a constant using either ordinary or generalized least squares. Thus, their Dynamic Ordinary Least Square (DOLS) method (eq. 3) assures the exogeneity of the factor variable x(t) and the approximately normal t-statistics of a_1 which indicates its long-run impact on the dependent variable (y). The possible serial correlation in the error term u(t) in eq. 3 might be dealt with by computing a serial correlation-robust standard errors for a_1. The second contribution is that their model produces asymptotically efficient estimators in finite samples.

$$y(t) = a_0 + a_1 x(t) + \sum_{j=-k}^{k} f_i * \Delta x_i(t+j) + u(t) \tag{3}$$

Stock and Watson apply the DOLS estimation to the long-run demand for money (M1) by regressing the latter onto real net national product, the net national product price deflator both in a log form as well as on the commercial paper rate in percent.

Gregory and Hansen (1996a, 1996b) extend the contributions in the field of cointegration by proposing a test which allows for a regime shift either in the slope alone i.e. regression coefficient a_0 or in the entire coefficient vector (a_1). Moreover, they consider cases where the intercept or slope has a single break of unknown timing. That makes the process of finding the break point more formal and objective. In general, the method being illustrated by the long-run money demand function is useful for modeling processes which are following a certain pattern of relation over a fairly long period of time, and then shift to a “new” long-term relation. A dummy variable (φ) denotes the break at the (unknown) moment τ (eq. 4):

$$\varphi = 0,\ \text{if } t \le \tau;\ 1, \text{if } t > \tau \quad (4)$$

One might choose among the following forms of structural change in case of a single variable regression (Gregory and Hansen, 1996a):

1. Level shift (C):

$$y(t) = a_0 + a_1 x(t) + a_2 \varphi + u(t) \quad (5)$$

In this case (eq. 5) a_0 represents the intercept before the shift whereas a_2 stands for the change in the intercept at the time of the shift. The slope coefficient (a_1) is held constant thus implying that the equilibrium relation has shifted in a parallel fashion. The addition of a time trend results in a level shift with trend model (eq. 6).

2. Level shift with trend (C/T):

$$y(t) = a_0 + a_1 x(t) + a_2 \varphi + a_3\, t + u(t) \quad (6)$$

3. Regime shift (C/S)

Another option is a change of the regression coefficient (a_1). Eq. (7) shows that not only does the steady-state relation move (a_2) parallel but it rotates as well (a_3).

$$y(t) = a_0 + a_1 x(t) + a_2 \varphi + a_3\, \varphi\, x(t) + u(t) \quad (7)$$

4. Regime and trend shift (C/S/T)

In a later study, Gregory and Hansen (1996b) extend the previous model by allowing also for a shift in the trend. The new slope coefficient a_5 presents the shift in the trend after the break (eq. 8).

$$y(t) = a_0 + a_1 x(t) + a_2 \varphi + a_3\, \varphi\, x(t) + a_4\, t + a_5 \varphi t + u(t) \quad (8)$$

Figure 1 gives an idea how the relations might be illustrated graphically.

Fig.1. A simple graphical illustration of models with structural break

a) shift in the level only   b) shift in both level and slope

The null hypothesis being tested is the existence of no cointegration against the alternatives in models 1 through 4. From a practical point of view, one should compute the cointegration test statistic for each possible regime shift (τ ϵ T) and take the smallest value i.e. the larges negative one across all possible break points. The value of T should be small enough (for example 0.15 - 0.85) in order a wide range of options to be exhausted. The computational procedure requires that for each τ a respective model of a type 1-4 be solved by OLS thus yielding a residual denoted by $\hat{u}_\tau(t)$. From these residuals the first-order serial correlation coefficient is calculated. The tools for testing the null involve the Phillips (1987) and the Augmented Dickey-Fuller (ADF) statistics (Zα(τ), Zt(τ), ADF(τ)). Of these statistics the smallest values across all values of τ are used (Zα*, Zt*, ADF*) since those values constitute evidence against the null. For an application of the method see, for example, Golinell and Orsi (2000).

The paper of Carrion-i-Silvestre and Sanso (2006), CS henceforth, addresses one weakness in the relevant literature, specifically the definition of the null as “no cointegration”. Contrary to that, they generalize the contribution in the field by developing a test for the null of cointegration. The regression model resembles the abovementioned DOLS procedure of Stock and Watson (1993). The robustness of the regression output is ensured by an inclusion of additional variables which account for a structural break in the data series (g(t)), the potential endogeneity of the regressors as well as the serial correlation in the errors by the lags and leads of the first differences of the explanatory variables ($\Delta x_i(t+j)$). The test is powerful for small samples, as well. The options for both an a priori known and an unknown break point had been examined. The procedure requires an estimation of models of the following general form (eq. 9):

$$y(t) = a_0 + \sum a_i * x_i(t) + \sum_{j=-k}^{k} f_i * \Delta x_i\,(t+j) + g(t) + u(t) \tag{9}$$

The vector of slope coefficients [a_i] measures the long-run impact of the regression variables. The function denoted by g(t) represents the nature of the structural change which best describes the co-integrating relation. The authors consider six alternative specifications. It is allowed for a structural shift that changes not only the deterministic component but the cointegrating vector by adding an interaction between the independent variable and a break dummy. The main difference with the GH model commented above is the inclusion of lead and lags for the factor variables and the construction of the test statistics. Here the Lagrange-Multiplier (LM) statistics is applied to test whether a long-run relation exists. In case of unknown date of the break, the latter is estimated as the point of time which minimizes the sequence of the sum of squared residuals. The next section demonstrates how the method might be applied to estimate the long-run function of growth in Bulgaria.

3. APPLICATION OF CARRION-I-SILVESTRE AND SANSO METHOD TO A MODEL OF BULGARIA’S GROWTH

The empirical model which demonstrates the CS method is based on the neoclassical approach. In the augmented Solow model (Mankiw et al., 1992) human capital (H) is included as a third input alongside physical capital (K) and labor (L). Thus, the extended Cobb-Douglas aggregate production function with constant returns to scale takes the form:

$$Y_t = A_t * K_t^{\alpha} * H_t^{\beta} * L_t^{1-\alpha-\beta} \tag{10}$$

The parameters α and β measure the output elasticity with respect to physical and human capital, respectively; A indicates technological progress. I share the widely accepted view that the stock of human capital depends heavily on the educational status of the population. Moreover, two components might be distinguished according to highest educational level attained: active population having completed upper secondary education at most (variable sec) and active population with tertiary education (high). Their elasticities in the production function are β_1 and β_2, respectively.

Taking into account the above assumption, it could be proven that the functional relationship between the steady-state level of income per effective worker (y*) and the stock of human capital has the following form[2]:

$$\ln y^* = a_0 + \frac{\alpha}{1-\alpha} \ln s_k - \frac{\alpha}{1-\alpha} \ln(n + \delta + g) + \frac{\beta_1}{1-\alpha} * \ln sec^* + \frac{\beta_2}{1-\alpha} * \ln high^* \tag{11}$$

Table1. Unit root test of Zivot and Andrews for real GDP

	Model 1*		Model 2		Model 3	
	Test statistics	Break point	Test statistics	Break point	Test statistics	Break point
rgdp_BG	-2.3435	2002:2	-3.6192	2006:4	-3.5121	2006:3

*The null hypothesis is that the series is integrated without an exogenous structural break. Model 1 assumes a break in the level, model 2 – a break in the slope, while model 3 – both.

Source: Author's calculations

The empirical specifications in this paper are based on eq. (11). Seasonally adjusted quarterly data for the Bulgarian economy over the period 2000:1 – 2013:4 are used. The Zivot and Andrews (1992) unit root test with an endogenously determined break has been applied (see, table 1).

The growth regression (eq. 12) relates Real Gross Domestic Product (RGDP) per unit of active population (25 – 64 years of age) in logs to a set of variables (Xi). The vector of slope coefficients [di] measures the long-run impact on growth of the regression variables.

$$ln\ RGDP_t = a_0 + \sum d_i * X_{i,t} + \sum_{j=-k}^{k} f_i * \Delta X_{i,t+j} + g(t) + \varepsilon_t \tag{12}$$

Four specifications for g(t) have been considered for each break quarter, as follows:

[2]The mathematical description of the steady-state equation might be seen in Neycheva (2016). Also, in Neycheva and Joensen (2017).

a shift in the level: $g_{An}(t) = \alpha + \theta * DU_t$ (13)

a level shift with trend: $g_A(t) = \alpha + \theta * DU_t + \varphi * t$ (14)

a regime shift: $g_D(t) = \alpha + \theta * DU_t + \sum_i \beta_i * X_{i.t} * DU_t$ (15)

a regime shift with trend: $g_E(t) = \alpha + \theta * DU_t + \varphi * t + \sum_i \beta_i * X_{i.t} * DU_t$ (16)

The dummy variable (DU) is equal to 0 up to the corresponding break quarter for real GDP (TB) and 1 after that.

$$DU_t = \begin{cases} 1, t > TB \\ 0, t \leq TB \end{cases} \tag{17}$$

Taking into consideration the three break points denoted in table 1, the result is a total of 12 regressions. Table 2 displays only those of them for which the sum of squared residuals attains its minimum value. The respective break quarter is also shown in the last column.

Table 2. Carrion-i-Silvestre and Sanso test for cointegration with structural break

Model	Bulgaria	
	SC statistic	Break quarter
An	0.0248(1) (0.0874)	2006:3
A	0.0204(1) (0.0621)	2006:3
D	0.0208(1) (0.0514)	2006:4
E	0.0529 (0.0329)	2008:3

**Only the break points which minimize the sum of squared residuals have been selected.
(1)Significant at the 0.05 level.
Source: Author’s calculations

It must be noted that the main drawback of the method is that Carrion-i-Silvestre and Sanso have tabulated critical values for up to four factor variables only. Nonetheless, the present calculations are far below the critical values at the 5% level of significance presented in brackets, thus suggesting that the null of co-integration cannot be rejected in all cases except model E.

The regression output for eq. 12 is shown in table 3. Three modifications have been evaluated, specifically model An and model D (see, table 2) as well as the “no break” case. Model E has been rejected since the CS test could not prove the cointegrating link as it was explained above. Model A has also been dropped because the Doornik – Hansen test was unable to prove that residuals follow a normal distribution. Variable sec and high denote the percentage of the active population (25 -64 years of age) with upper secondary or tertiary education completed. The real business investments are presented by rinv whereas fdi stands for foreign direct investments. The value of export (exp) is also introduced as it is considered a key determinant of growth in our economy especially before the recent financial crisis (2008-2009).

Table 3. Regression output: impact of education on growth in Bulgaria

Dependent variable: rgdp*	1 break (06:Q3)	2 break (06:Q4)	3 no break case
sec	-0.375(1) (0.099)	-0.962(1) (0.078)	-0.662(1) (0.079)
high	-0.159(2) (0.067)	0.117 (0.238)	-0.135(2) (0.060)
rinv	-0.007 (0.008)	0.047 (0.048)	-0.008 (0.009)
fdi	0.125(1) (0.004)	0.097(1) (0.027)	0.125(1) (0.006)
exp	0.129(1) (0.014)	0.161(1) (0.027)	0.131(1) (0.017)
sec_dummy**	-	0.188 (0.368)	-
high_dummy	-	-0.584(3) (0.325)	-
rinv_dummy	-	0.082 (0.050)	-
fdi_dummy	-	0.047 (0.033)	-
exp_dummy	-	0.228(1) (0.041)	-
rgdp_ dummy	-0.018(1) (0.005)	-1.368 (2.215)	-
Normality of residual	3.542 (0.170)	1.260 (0.532)	2.009 (0.366)

*The dependent variable is log of real GDP per unit of active population. One lag and lead have been included. Standard errors are presented in parentheses.

(1), (2), (3) Significant at the 0.01, 0.05 and 0.10 level, respectively.

Source: Author's calculations based on Eurostat data

The general conclusion emerging from this econometric exercise is that an increase of the human capital stock in Bulgaria is not positively related to the increments of aggregate product. The effect of upper secondary education is always both negative and statistically significant. The regression coefficient for tertiary education is either below zero or positive but statistically not significant. This somewhat puzzling result is supported by the model containing no break (model 3). It must be mentioned that the signs and the level of significance of the slope coefficients are not vulnerable to the changes of either the number of lags and leads or the break point. The absolute values of the coefficients for secondary education are higher than that for tertiary education therefore it might be concluded that secondary education is more strongly linked to the long-run growth path.

4. CONCLUDING REMARKS

The paper presents a comparative review of a number of contemporary methods which are used to estimate cointegrating relations. Specifically, it compares the DOLS model of Stock and Watson (1993), Gregory and Hansen's method (1996a,b) allowing for a structural break as well as the method of Carrion-i-Silvestre and Sanso (2006). The latter could be considered as an extension of the former two since it combines the structural break approach with the dynamic OLS. Its application to a model linking education to growth in Bulgaria produces robust results. In light of the recent financial crisis, the inclusion of a structural break better fits the time series and should be taken into account.

References

1. Carrion-i-Silvestre J. and Sanso, A. (2011). Testing the Null of Cointegration with Structural Breaks. Oxford Bulletin of Economics and Statistics, vol. 68, pp. 623-646.

2. Dowrick S. (2003). Ideas and Education: level or growth effects? WP 9709, NBER.

3. Engle R. and Granger, C. (1987). Co-integration and error correction: representation, estimation and testing. Econometrica, vol. 55, pp. 251-276.

4. Golinelli, R. and Orsi, R. (2000). Testing for structural change in cointegrated relationships: analysis of price-wage models for Poland and Hungary. Economics of Planning, vol. 33, pp. 19-51.

5. Green, W. (2003). Econometric Analysis. 5th Ed. Upper Saddle River, New Jersey, US: Prentice Hall.

6. Gregory, A. and Hansen, B. (1996). Residual-based tests for cointegration in models with regime shifts. Journal of Econometrics, vol. 70, pp. 1-26.

7. Gregory A. and Hansen, B. (1996). Tests for Cointegration in Models with Regime and Trend Shifts. Oxford Bulletin of Economics and Statistics, vol. 58, pp. 555-560.

8. Gujarati, D. (2004). Basic econometrics. 4th Ed.,The McGraw-Hill Companies.

9. Kwiatkowski D., Phillips, P., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, vol. 54, pp.159-178.

10. Mankiw G., Romer, D. and Weil, D. (1992). A Contribution to the Empirics of Economic Growth. Quarterly Journal of Economics, vol. 107, pp. 407-437.

11. Neycheva, M. (2016). Secondary versus higher education for growth: the case of three countries with different human capital's structure and quality. Quality and Quantity, vol. 50, pp. 2367-2393.

12. Neycheva, M. and Joensen, K. (2017). Higher Educational Attainment for Growth: the MRW model for Iceland. Scandinavian Journal of Educational Research, DOI: 10.1080/00313831.2017.1357144

13. Phillips. P. (1987). Time series regression with a unit root. Econometrica, vol. 55, pp. 277-301.

14. Stock, J. and Watson, M. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. Econometrica, vol. 61, pp. 783-820.

15. Wooldridge, G. (2013). Introductory Econometrics. A modern Approach. 5th Ed., Mason, US: South-Western Cengage Learning.