3 SPURIOUS REGRESSION AND COINTEGRATION. NUMERICAL EXAMPLE: ROMANIA'S M2 MONEY DEMAND

Gheorghe RUXANDA*
Andreea BOTEZATU**

Abstract

Economic time series are, in their vast majority, integrated series so, their modelling procedure stumbles upon the problem of spurious regression. When existent, cointegration is the simplest way of eliminating the illogical correlation established between time series due to the presence of trends. The analysis of macroeconomic time series through cointegration is a common fact. Modelling the Romanian M2 money demand through cointegration and vector error correction led to somewhat significant results being a starting point for future, more complex research.

Key words: spurious regression, cointegration, money demand, error correction mechanism.
JEL classification: C22, C32, E41, C51, C52

1. Introduction

Starting with the 1980s, cointegration – one possible way of handling spurious regression – has widely been used in macroeconomic analysis. Long-run money demand is by far the most tackled upon issue when it comes to cointegration, together with UIP being the most common numerical example in econometric textbooks, due to strong economic evidence for their existence.

The paper is structured in two parts, the first one being an overview concerning spurious regression and cointegration while the second part concentrates on the attempt of identifying a long-run equation for M2 money demand.

The article has an exploratory nature, the purpose of the performed analyses being only to identify the possibility of Romanian money demand further and more complex studies.

---

* Ph.D Professor Academy of Economic Studies, Bucharest, email: ghrux@ase.ro
** Ph.D Canditate, Academy of Economic Studies, Bucharest, email: andreeabotezatu@gmail.com

2. Spurious regression and cointegration

When the analysed data series contain unit roots the regression equation by which they can be modelled is inadequate – *spurious* – as it shows illogical correlations between series. This type of relationship is due to the presence of trends in the data series, the processes not necessarily having the same causal phenomena. So, if the analysed series are not stationary, the regression seems to be statistically significant, even though the only thing present is period correlation and not causal relations between the data (Harris, 1995).

The statistical tests validate the regression coefficients, whenever the series contain trends. When time series contain unit roots, statistical tests overestimate the dependency between the variables, and the estimators are doubtful. Often, the null of no relation between the variables is rejected even when it is inexistent. Generating two independent random walk series and estimating a regression between them, Granger and Newbold (1974) demonstrated that in a substantially high number of cases the regression proved to be valid according to the t-test.

Following the example of Granger and Newbold, we repeatedly (40000 times) generated 2 random walk processes (each with 100 cases) and estimated the regression between them. Even though the series were independent, the estimated regressions' coefficients proved to be valid according to the t test in 76% of the cases. At the same time, R squared registered values higher than 20% in 46% of the developed regressions.

Starting as early as 1926, the presence of spurious regressions was detected in a study of Yule who showed the existence of significant correlation between the number of marriages and the mortality rate throughout 1866-1911. Another example of famous illogical correlation was the one demonstrated by Hendry in 1980 between the price level and the quantity of rain (Phillips, 1998).

The problem of spurious regression can be eliminated by differencing the data, but this implies the loss of long-run information content in the data.

Cointegration

Modelling time series so as to keep their long-run information content can be done through cointegration. This is a relatively new field of analysis, but extremely rapidly evolving. The term was firstly used in 1986 in the March edition of *Oxford Bulletin of Statistics and Econometrics*, even though references to this term were found dating back to 1964 in Sargan's papers concerning the error correction mechanism.

Engle-Granger cointegration procedure

Clive W. Granger introduced the term of cointegration and demonstrated its likeliness while attempting to prove the opposite – the impossibility of obtaining an I(0) series by regressing two I(1) processes.

Engle and Granger (1987) give the following definition to cointegration: "the components of the vector x, are said to be cointegrated of order d, b, denoted $x \sim CI(d,b)$, if:

- (i) all components of *x* are I (d);

- (ii) there exists a vector *a*(≠ 0) so that:

$$z = a'x \sim I(d-b),\ b > 0.$$

The vector *a* is called the cointegrating vector."

For a number of just two variables x_t and y_t both I(1), the previous description becomes:

$$z_t = x_t - a\, y_t \qquad (1)$$

When z_t is I(0), the constant *a* acts in the in the sense of cancelling the lung-run components of x_t and y_t.

Summarising, the cointegration equation shows the evolution and the long-run relationship between the variables, any shifts in the data due to various shocks are considered to be temporary and the data to be reverting to their long-run path.

Johansen cointegration technique

The Johansen test permits the identification of multiple cointegration relationships. In describing the Johansen technique, the starting point is a vector y_t which can be expresses as a VAR with k lags:

$$y_t = A_1 * y_{t-1} + A_2 * y_{t-2} \ldots\ldots. + A_k * y_{t-k} + \varepsilon t \qquad (2)$$

where: y_t is a vector (n x 1); A_i is the parameters matrix (n x n).

Transforming (2) in an error correction mechanism, the following equation is obtained:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \ldots. + \Gamma_1 \Delta y_{t-k+1} + \Pi\, y_{t-k} + \varepsilon_t \quad (3)$$

where:

$$\Gamma_i = -(I - A_1 - \ldots - A_i),\ i = \overline{1, k-1};\ \Pi = -(I - A_1 - \ldots - A_k),\ \Pi = \alpha\beta',$$

α represents the speed of adjustment and β the matrix of long-run coefficients.

The number of the cointegrating relationships is given by the rank of Π, and three possibilities arise: (1) the rank equals n, all variables in y are I(0) (stationary); (2) the rank is 0, there exists no cointegrating relationship between variables and (3) the rank is lower than n, when there exist a maximum of n-1 cointegrating relationships.

The estimation procedure of α and β' was developed by Søren Johansen through reduced rank regression (a detailed presentation can be found in Johansen (1991)). The tested null hypothesis is the existence of a maximum of r cointegration vectors in $\Pi = \alpha\beta'$.

Lütkepohl and Saikkonen (1999) formulate the hypothesis of the Johansen cointegration test as follows:

$$H_0\,(r_0):\ rk(\Pi) = r_0 \text{ with the alternative } H_1\,(r_0):\ rk(\Pi) > r_0,$$

$$\text{or, } H_0\,(r_0):\ rk(\Pi) = r_0 \text{ with the alternative } H_0\,(r_0):\ rk(\Pi) = r_0 + 1;$$

In order to establish the number of cointegrating relations the characteristic roots of matrix Π are determined ($\hat{\lambda}_1, \hat{\lambda}_2, \ldots \hat{\lambda}_{n-1}$). The two test statistics, for determining the number of significant eigenvalues, are:

$$LR_{trace}(r) = -T\sum_{i=r+1}^{n} \ln(1-\hat{\lambda}_i), \tag{4}$$

the trace statistic and:

$$LR_{\max}(r,r+1) = -T\ln(1-\hat{\lambda}_{r+1}) = LR_{trace}(r) - LR_{trace}(r+1), \tag{5}$$

the maximum eigenvalue statistic, $r = 0,1,...,n-1$.

The computed values are compared to the critical values, by this determining the exact number of cointegrating equations.

The results of the Johansen cointegration test are influenced by the considered lag length. For determining the lag lenth, the following criteria are used: LR (Likelihood Ratio Criterion), AIC (Akaike Information Criterion), SIC (Schwarz Information Criterion), FPE (Final Prediction Error), HQ (Hannan-Quinn Information Criterion).

As Philips (1998) shows, the cointegration equation only explains the trend of one variable by the means of the causal relation from other variables which can, in their turn, be endogenous, and it doesn't explain the trend itself.

3. Numerical Example: Romania's M2 Money Demand

Even though suffered a de-emphasis in the 1980s as shown by Duca and VanHoose (2004), the interest for money demand modelling revived and it remains important both for policy implementation and for theoretical reasons. As it is known, after Romania adopted the inflation targeting regime, monetary aggregates lost their role of policy anchors. Still, their importance is beyond doubt especially in the context of future euro-area membership, as it constitutes ECB's first pillar in the assessment of risk to price stability (its stated main objective). Money demand equation is extremely important for the ECB in order to establish its monetary policy, since the great number of studies on the behaviour of M3 in the euro- area (Coenen Gűnter, Vega Jean-Luis, 1999; Brand Claus, Cassola Nuno, 2000) and also in the new member states by the means of panel cointegration techniques (Dreger et al.). The interest for money demand analysis in Romania and the importance of monetary aggregates analysis is shown in the NBR's papers (Antohi et al., 2007).

A detailed analysis of how money demand was modelled throughout time can be found in Sriram Subramanian (2001). Budina et al. (2006) analysed Romanian money demand and its influence over inflation throughout the hyperinflation period 1996-2000, pointing out the existence of a stable and not affected by shocks long-run equation. Romania's money demand was the subject of analysis also in Antonescu et al., 2004 and Pelinescu et al, 2001. As part of a panel of data, Romanian money demand is studied in Fidrmuc, 2006. Other significant studies concerning M2 modelling are the ones for the Czech Republic (Arlt et al., 2001), Latvia (Tillers, 2004), Armenia (Poghosyan, 2003) and Nigeria (Enisan, 2006).

The general form of money demand is $(M_2/P)=f(Y, OC)$. The variables used in the regression are real M2[1] (ln_m2_r_sa), as proxy for the income the industrial production index is used due to its monthly frequency (ln_ipi_sa), as opportunmity costs for holding money the interst rate for outstanding deposits (r_dob_out) the ron/eur exchange rate (ln_ron_eur) and the consumer price index are used (ln_cpibl).

The considered data cover the period December 2004 – December 2007 and all series except for the interest rate enter the equation in logarithms. The series affected by seasonal factors were firstly seasonally adjusted with the Census X12 procedure implemented in Eviews.

Following the Box-Jenkins approach, the first stage is data pre-testing, consisting in unit root analysis by the means of Augmented-Dickey Fuller and Phillips-Perron tests. The results of these tests indicated that all series are integrated of order 1 – I(1). The second stage, estimation and re-specification, consists in the Johansen cointegration analysis.

The majority of lag length criteria suggest the use of a lag of 2 in the analysis:

Table 1

Lag length criteria

VAR Lag Order Selection Criteria
Endogenous variables: LN_M2_R_SA R_DOB_OUT LN_IPI_SA LN_CPIBL LN_RON_EUR

Exogenous variables: C

Sample: 2004M12 2007M12

Included observations: 34

Lag	LogL	LR	FPE	AIC	SC	HQ
0	266.3299	NA	1.45e-13	-15.37234	-15.14788	-15.29580
1	398.7991	218.1847	2.66e-16	-21.69407	-20.34728*	-21.23477
2	431.9799	**44.89158***	**1.82e-16***	-22.17529	-19.70617	**-21.33325***
3	457.6215	27.14998	2.32e-16	-22.21303*	-18.62159	-20.98825

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

At this lag length, both the trace statistic and the maximum eigenvalue suggest the existence of one cointegration equation.

1 M2= currency in circulation + overnight deposits (in lei, euros and other currencies)+ deposits redeemable at notice up to 3 months (in lei, euros and other currencies) +deposits with agreed maturity up to 2 years (in lei, euros and other currencies).

Table 2

Johansen cointegration test

Sample (adjusted): 2005M03 2007M12
Included observations: 34 after adjustments
Trend assumption: Linear deterministic trend
Series: LN_M2_R_SA R_DOB_OUT LN_IPI_SA LN_CPIBL LN_RON_EUR
Lags interval (in first differences): 1 to 2
Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.822076	101.3200	69.81889	0.0000
At most 1	0.494771	42.62242	47.85613	0.1420
At most 2	0.300317	19.40916	29.79707	0.4638
At most 3	0.186521	7.266828	15.49471	0.5468
At most 4	0.007269	0.248042	3.841466	0.6185

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.822076	58.69761	33.87687	0.0000
At most 1	0.494771	23.21326	27.58434	0.1646
At most 2	0.300317	12.14233	21.13162	0.5337
At most 3	0.186521	7.018786	14.26460	0.4869
At most 4	0.007269	0.248042	3.841466	0.6185

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

The resulted long-run equation, the cointegration equation, is adequate from the point of view of the coefficients' signs.

Table 3

Cointegration equation

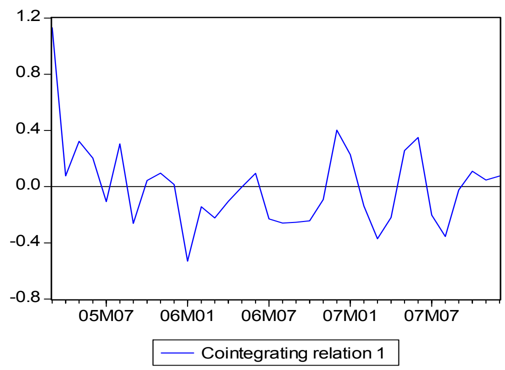
1 Cointegrating Equation(s):	Log likelihood	436.3103

Normalized cointegrating coefficients (standard error in parentheses)

LN_M2_R_SA	R_DOB_OUT	LN_IPI_SA	LN_CPIBL	LN_RON_EUR
1.000000	0.267062	-4.930242	57.54844	-3.377176
	(0.02914)	(0.63175)	(8.23569)	(0.95565)

Figure 1

Cointegration equation



From statistical point of view the coefficients are significant as the t-test shows.

Table 4

Cointegrating equation:

Cointegrating Eq:	CointEq1
LN_M2_R_SA(-1)	1.000000
R_DOB_OUT(-1)	0.267062
	(0.02914)
	[9.16357]
LN_IPI_SA(-1)	-4.930242
	(0.63175)
	[-7.80413]
LN_CPIBL(-1)	57.54844
	(8.23569)
	[6.98769]
LN_RON_EUR(-1)	-3.377176
	(0.95565)

	[-3.53391]
C	-253.3509

Cointegrating equation is:

LN_M2_R_SA(-1)=253.35 - 0.27* R_DOB_OUT(-1) + 4.93* LN_IPI_SA(-1) - 57.55* LN_CPIBL(-1) + 3.38* LN_RON_EUR(-1).

The obtained equation shows the direct relationship between income (approximated by the industrial production) and money, and the inverse relation with the interest rate, in accordance with economic theory. The income coefficient, higher than 1, suggests the monetization phenomenon which affected the Romanian economy throughout the analysed period. The positive sign of the exchange rate coefficient might suggest the wealth argument from the economic literature, as it shows that as the leu depreciates the demand for money increases. In fact, the economic literature isn't precise when it comes to the sign of the exchange rate in connection with the money demand. When a negative sign (most common in money demand studies) is obtained, the exchange rate, expressed as units of domestic currency per unit of foreign currency, behave as an opportunity cost variable, showing the substitution between currencies. On the other hand, when the sign is positive, the coefficient suggests that inflationary effect of depreciation and consequently a higher demand for money. Usually the substitution happens during periods of hyperinflation.

The connection between money demand, inflation and income revealed by the analysis confirm the monetarist view that inflation is everywhere a monetary phenomenon and by this the importance of monetary aggregates analysis for policy making is stressed.

Granger proved that cointegrated series can be modelled by ECM as well as the fact that variables entering an error correction mechanism are cointegrated. By building an ECM with the variables entering the cointegration equation, a relationship containing both the long and the short run information is obtained (lr in the ECM below represents the long run component):

Table 5

Error correction mechanism

Dependent Variable: D(LN_M2_R_SA)
Method: Least Squares
Date: 03/27/08 Time: 15:06
Sample (adjusted): 2005M03 2007M12
Included observations: 34 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.021569	0.002486	8.674644	0.0000
LR	0.021152	0.009449	2.238676	0.0330
D(LN_IPI_SA(-1))	0.266934	0.095172	2.804757	0.0089
D(LN_CPIBL(-1))	-1.209331	0.541493	-2.233326	0.0334
D(LN_IPI_SA(-2))	0.264789	0.083022	3.189373	0.0034

R-squared	0.314839	Mean dependent var	0.023447
Adjusted R-squared	0.220334	S.D. dependent var	0.015986
S.E. of regression	0.014116	Akaike info criterion	-5.548024
Sum squared resid	0.005778	Schwarz criterion	-5.323559
Log likelihood	99.31641	F-statistic	3.331458
Durbin-Watson stat	1.793778	Prob(F-statistic)	0.023146

Residual analysis

Residuals are not affected by autocorrelation as the Breusch-Godfrey Serial Correlation LM test shows, are homoskedastic (table 7) and follow the normal distribution (graph 2).

Table 6

Breusch-Godfrey Serial Correlation LM Test

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.880306	Probability	0.426222
Obs*R-squared	2.081346	Probability	0.353217

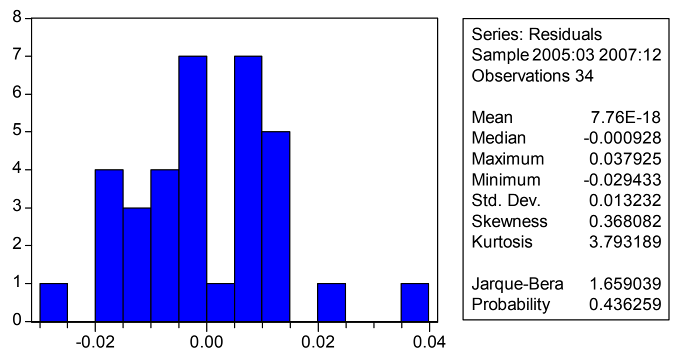
Table 7

White Heteroskedasticity Test

White Heteroskedasticity Test:			
F-statistic	1.313311	Probability	0.282177
Obs*R-squared	10.06071	Probability	0.260791

Figure 2

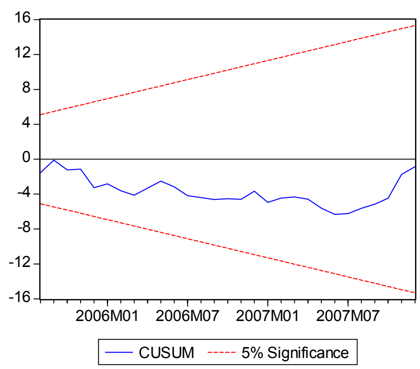
Residuals' histogram

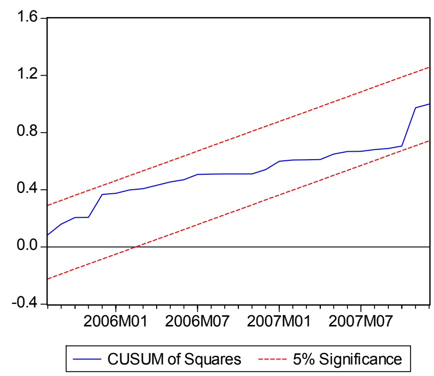


Stability testing

As mentioned by Brown, Durbin and Evans (1975), regressions aren't always constant over time especially when they involve economic data series. Hence, they proposed the CUSUM and CUSUM of squares methods based on recursive residuals in order to test for the model's long-run constancy.

Figure 3: CUSUM test **Figure 4: CUSUM of squares**





The CUSUM tests show a stable ECM equation. CUSUM test (Cumulative Sum of Recursive Errors) calculates the W statistic:

$$W_k = \sum_{j=p+1}^{k} \varpi_j / \sigma_{\varpi} ,$$

where ϖ_j is the recursive residual and σ_{ϖ} is the standard error of regression. Under the hypothesis of the parameters stability, the W statistic is situated inside the confidence interval.

Analysing the one and the N-step probability tests, some signs of instability can be detected.

Figure 5: One-step probability

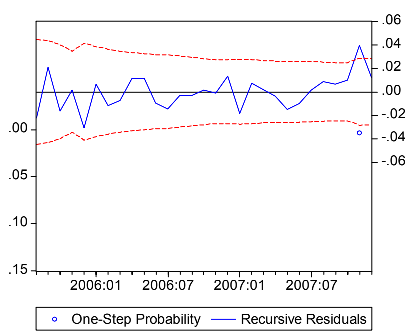
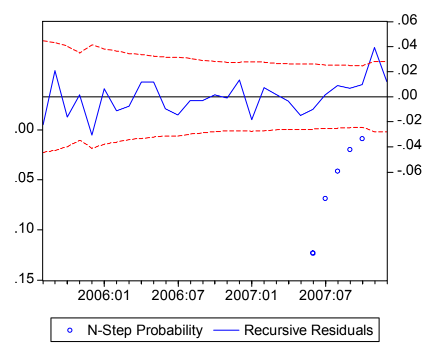


Figure 6: N-step probability



All in all, the error correction representation of the M2 isn't the best representation of the money demand – the R-squared coefficient being relatively small and some instability signs being present.

4. Conclusions

Applying the Johansen cointegration technique an adequate equation seemed to be obtained. But, it has to be regarded with due caution, and the resulted equation shouldn't be considered as a viable long-run money demand equation.

As Granger (1997) showed the results of the cointegration test are to a large extent influenced by the chosen lag length and in the present case, another lag length wouldn't have led to results similar to the ones presented above.

Also, the long-run component of data series needs time to accumulate in the data, the sample size being quite small in the considered case. Consequently, the coefficients of the above developed equations should be prudently analysed and attention not be paid to their value but more to their sign.

Even so, the analysis performed in this paper is an important starting point of future, more detailed research. Cointegration, extremely analysed and described in studies and scientific papers is one of the greatest discoveries of the 20th century, being also the solution – when existent among data – to spurious regression.

References

Akinlo Enisan, 2006, The Stability of Money Demand in Nigeria: An Autoregressive Distributed Lag Approach, “Journal of Policy Modeling”, (28): 445-452.

Antohi D., Stere T., Udrea I., Bistriceanu G., Botezatu A., 2007, Evoluții monetare în economia românească: determinanți şi implicații, “Caiete de studii”, no. 21.

Andronescu, Andreea, Hassan Mohammadi, James E. Payne, 2004, Long-run estimates of money demand in Romania, “Applied Economics Letters”, 11(14: 861 – 864.

Arlt J., Guba M., Radkovski S., Sojka M., Stiller V., 2001, Influence of Selected Factors on the Demand for Money, Working Paper, No. 30, Praha.

Brand Claus, Cassola Nuno, 2000, A Money Demand System for the Euro Area M3, ECB, Working Papers no. 39.

Brown R. L., Durbin J., Evans J. M, 1975, Techniques for testing the constancy of regression relations over time, “Journal of Royal Statistical Society”, (37): 149-192.

Budina Nina, Maliszewski Wojciech, Georges de Menil, Turlea Geomina, 2006, Money, Inflation and output in Romania, 1992-2000, “Journal of Internatiuonal Money and Finance”, (25): 330-347.

Coenen Gűnter, Vega Jean-Luis, 1999, The Demand for M3 in the Euro Area, ECB Working Papers, no. 6.

Dreger Christian, Hans-Eggert Reimers, Barbara Roffia, 2006. Long Run Money Demand in the New EU Member States with Exchange Rate Effects, ECB Working Paper no. 628.

Duca John V., Vanhoose David D., 2004, Recent Developments in Understanding the Demand for Money, “Journal of Economics and Business”,(56): 247-272.

Engle Robert, Granger Clive W J, 1987, Co-integration and Error Correction: Respresentation, Estimation and Testing, “Econometrica”, 55(2): 251-276.

Engle Robert, White Halbert (ed.), 1999, Cointegration, Causality, and Forecasting, Oxford University Press.

Fidrmuc, Jarko, 2006. Money Demand and Disinflation in Selected CEEC's during the Accession to the EU, Munich Economics Working Paper, Department of Economic, University of Munich.

Granger Clive W. J., 1997, On Modelling the Long Run in Applied Economics, "The Economic Journal", (107): 169-177.

Granger Clive W. J., 2004, Time Series Analysis, Cointegration, and Applications, "The American Economic Review", 94(3): 421-425.

Harris R. I. D., 1995, Using Cointegration Analysis in Econometric Modelling, Prentice Hall Publishing, Essex England.

Johansen Søren, 1991, Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, "Econometrica", 59(6): 1551-1580.

Millern Stephen, 1991, Monetary Dynamics: An Applied Cointegration and Error-Correction Modeling, "Journal of Money, Credit and Banking", 23(2): 139-154.

Miyao Ryuzo, 1996, Does a Cointegrating M2 Demand Relation Really Exists in the United States?, "Journal of Money, Credit, and Banking", 28(3): 335-380.

Pelinescu Elena şi Cornelia Scutaru, 2001. A Dynamic Model of the Money Demand in Romania, "Romanian Journal of Economic Forecasting", 1-2/2001.

Phillips Peter C. B., 1998, New Tools for Understanding Spurious Regression, "Econometrica", 66(6): 1299-1325.

Poghosyan Tigran, 2003, Demand for Money in Armenia: Partial Adjustment Approach, Working Papers no. 03/09, Central Bank of Armenia.

Soderlin Paul, Vredin Anders, 1996, Applied Cointegration Anlysis in the Mirror of Macroeconomic Theory, "Journal of Applied Econometrics", 11: 363-381.

Subramanian Sriram S., 1999, Survey on Demand for Money: Theoretical and Empirical Work with Special Reference to Error-Correction Models, IMF Working Paper no. 64.

Tillers Ivars, 2004, Money Demand in Latvia, Bank of Latvia, Working Paper, 3/2004, Riga.