

Testing Cointegration for Czech Stock Market[#]

Tran Van Quang^{*}

Introduction

Unit root and cointegration analyses are an important achievement in econometric theory that has become one of the central interests of economists and financiers in the last decade. On the financial market, while the price of an individual stock may fluctuate randomly, some combinations of several stocks show that they may share a long-term equilibrium. In this case such behaviour can be used to realize a statistical arbitrage profit. Although the application of this technique is widespread around the world for various aims, we have not seen it to be used for Czech stock market so far. To fill this gap, the main objective of this paper therefore is to determine the linkages between prices of the most liquid stocks at the Czech stock market. Such linkages would enable a trading strategy which could lead to the mentioned arbitrage profit at the Czech stocks market.

This paper is divided into three sections. In section one, a summary of cointegration analysis is given. In section two, a cointegration technique is applied to test the linkage between prices of the most traded stocks on the Czech stock market, namely they are stocks of Telefonica O₂ (previously named Český Telecom) (abbreviated SPTT), Komerční banka (KB), České energetické závody (CEZ) and Unipetrol (UNPE). The results of the unit root ADF test of chosen stock prices time series and the cointegration test of a successful combination are shown in this section. In the last section, brief comments on the results of our analysis and some conclusions are presented.

1. Cointegration Analysis Summary

The definition in the simple case of 2 time series x_t and y_t , that both are integrated of order one ($I(1)$) and means that the process contains a unit root is the following: x_t and y_t are said to be cointegrated if there exists a parameter α such that $u_t = y_t - \alpha x_t$ is a stationary process.

This turns out to be a breakthrough invention of looking at time series because many financial time series behave that way. The first thing to notice is that financial series behave like $I(1)$ processes, e. i. they seem to drift all over the place; but the

[#] The support from the Czech Science Foundation under the Grant 402/03/H057 is gratefully acknowledged.

^{*} Ing. Tran Van Quang – research fellow; Center for Research in Economic Dynamics and Econometrics, Faculty of Finance and Accounting, University of Economics, Prague, nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic; <tran@vse.cz>.

second more important thing to notice here is that they seem to drift in such a way that they do not drift away from each other. The reason unit roots and cointegration is so important is the following. Let's consider the regression:

$$y_t = \alpha_0 + \alpha_1 x_t + u_t. \quad (1)$$

Firstly, if x_t is a random walk and that y_t is an independent random walk (so that x_t is independent of y_s for all s), then the true value of α_1 is of course 0, but the limiting distribution of $\hat{\alpha}_1$ is such that $\hat{\alpha}_1$ converges to a function of Brownian motions. This is called a spurious regression. It is well-known that the t-statistics of the regression diverges at rate \sqrt{T} , so it is impossible to detect the problem from the t-statistics. The problem does repeat itself in typical OLS regression output though when a very high R^2 and a very low Durbin – Watson value are found. Of course, if unit roots are found, then it is common to get standard consistent estimators by running the regression:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + e_t, \quad (2)$$

since it is now a regression of a stationary variable on a stationary variable, the difficulties mentioned before are overcome and classical statistical theory can be applied. Secondly, let's assume that x_t is a random walk and the y_t is another random walk, such that (1) holds true for non-zero α_1 , but with the error term u_t following a unit root distribution. In this case we can still get inconsistent estimates and we need to estimate the relation (2). This may also be called a spurious regression, even though there actually is a relationship between x_t and y_t . Finally, now let's consider that (1) holds with a stationary error term. This is exactly the case where x and y are cointegrated. In this case, $\hat{\alpha}_1$ is not only consistent, but converges to the true value at rate T and the OLS estimator is superconsistent. In the case where x_t is a simple random walk and u_t is serially uncorrelated we find that $T(\hat{\alpha}_1 - \alpha_1)$ is asymptotically distributed as a $\int_0^1 B_2 dB_1 / \int_0^1 B_2^2 dt$, where B_1 and B_2 are independent Brownian motions. This limiting distribution has a mean of zero, but more importantly the standard t-test is asymptotically distributed. In the situation where there is serial correlation or x_t may be correlated with u_s for some s we may not get a symmetric asymptotic distribution of $T(\hat{\alpha}_1 - \alpha_1)$, and even if the estimate converges at the very fast rate T , this may translate into non-negligible bias in typical macro data samples. In this situation the t-statistic is no longer asymptotically normal. The best way to test is this therefore to use the multivariate Johansen estimator, although one can also use Philips' fully modified estimator (the idea is to estimate a correction term, similarly to what is done in the Philips – Perron unit root tests (1988)), or one can allow for more dynamics in the relationship (1). It is obvious that in the situation where x_t and y_t are cointegrated the regression (2) is still consistent. So the differenced regression is always consistent and one could argue that it would be safe to always estimate this relation. The loss of

efficiency is, however, very significant and sometimes, time series are long enough that the gain in efficiency from running the cointegrating regression can be critical for getting decent results.

A stochastic process is said to be integrated of order p , denoted as $I(p)$, if it need to be differenced p times in order to achieve stationarity. More generally x_t and y_t are said to be cointegrated of order $CI(d, p)$ if x_t and y_t are both integrated of order p ; but there exist an α such that $y_t - \alpha x_t$ is cointegrated of order $d - p$. From now on we will only treat the $CI(1, 1)$ case, which will be referred to simply as cointegration. Most applications of cointegration methods treat that case, and it will allow for a much simpler presentation to limit the development to the $CI(1, 1)$ case.

For many purposes the above definition of cointegration is too limited. It is true that financial and economic series tend to move together, but in order to obtain a linear combination of the series, that is stationary, one may have to include more variables. The general definition of cointegration for the $I(1)$ case is therefore the following:

A vector of $I(1)$ variables y_t is said to be cointegrated if there exist a vector β_i such that $\beta_i' y_t$ is trend stationary. If there exist r such linearly independent vectors β_i , $i = 1, \dots, r$, then y_t is said to be cointegrated with cointegrating rank r . The matrix $\beta = (\beta_1, \dots, \beta_r)$ is called the cointegrating matrix.

One may note that $\beta_i' y_t$ is an r -dimensional vector of trend-stationary variables. This definition is symmetric in the variables, i.e. there is no designated left-hand side variable. This is usually an advantage for the statistical testing, but of course, it makes it harder for economic intuition.

The β_i vectors are individually identified only up to scale since $\beta_i' y_t$ stationary implies that $c\beta_i' y_t$ is stationary. This also implies that one can normalize one of the coefficients to one – but only in the case where one is willing to impose the a priori restriction that this coefficient is not zero. As far as the identification of the matrix β is concerned, it is also clear that if β is a cointegrating matrix, then $\beta F'$ is also a cointegrating matrix for any non-singular matrix F .

1.1 Cointegration in the Autoregressive Representation

The general VAR(k) model can be written as:

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta y_{t-j} + e_t, \quad (3)$$

as considered earlier.

If Π is equal to zero, this means that there is no cointegration. This is the model that is implicit in the Box – Jenkins method. The variables may be $I(1)$; but it can be

easily cured by taking differences (in order to achieve the usual asymptotic distribution theory).

If Π has full rank then all y_t must be stationary since the left hand side and the other right hand side variables are stationary (since we limit ourselves to variables that are either $I(0)$ or $I(1)$).

The most interesting case is when Π has less than full rank but is not equal to zero. This is the case of cointegration. In this case Π can be written as $\Pi = \alpha\beta'$ where α and β are $n \times r$ matrix. α and β are only identified up to non-singular transformations since $\Pi = \alpha F^{-1}(\beta F')' = \alpha\beta'$ for any non-singular F . This lack of identification can sometimes render results from multivariate cointegration analysis impossible to interpret and finding a proper way of normalizing β (and thereby α) is often the hardest part of the work. α can be interpreted as a speed of adjustment towards equilibrium.

1.2 Cointegration in the Moving Average Representation

The multivariate Wold representation states that the stationary series Δy_t can be written as:

$$(1 - L)y_t = \Psi(L)e_t, \quad (4)$$

which, according to the Beveridge – Nelson decomposition, can be written as

$$y_t = \Psi(1)S_t + \Psi^*(L)e_t, \quad (5)$$

where S_t is the n -dimensional random walk $S_t = \sum_{s=1}^t e_s$ and

$$\Psi^*(L) = (1-L)^{-1}(\Psi(L) - \Psi(1)). \quad (6)$$

Now βy_t is stationary in the case of cointegration, so that

$$\beta y_t = \beta\Psi(1)S_t + \beta\Psi^*(L)e_t, \quad (7)$$

is stationary, which implies that $\beta\Psi(1)S_t$ is equal to 0. This gives another characterization of cointegration that is useful for testing.

One can show that the representation (1) can be reformulated in the case of cointegration as

$$y_t = \tilde{\Psi}S_t^* + \Psi^*(L)e_t, \quad (8)$$

where S_t^* is the $(n-r)$ -dimensional random walk. This is called a common trend representation in Stock and Watson (1988), and this representation can be also used as the basis for cointegration tests.

1.3 Testing for Cointegration

1.3.1 The Engle – Granger test

The most well known test, suggested by Engle and Granger (1987), is to run a static regression (after first having verified that x_t and y_t are both $I(1)$)

$$y_t = \theta'x_t + e_t, \quad (9)$$

where x_t is one or higher dimensional. The asymptotic distribution of θ is not standard, but the test suggested by Engle and Granger was to estimate $\hat{\theta}$ by OLS and the test for unit roots in

$$\hat{e}_t = y_t - \hat{\theta}'x_t. \quad (10)$$

Since the unit root tests test the null hypothesis of a unit root, most cointegration tests test the null of no cointegration. Unfortunately, the limiting distribution of, for example, t-test, does not have the limiting distribution tabulated by Dickey and Fuller (1979, 1981). The limiting distribution does, however, resemble the Dickey – Fuller distribution even though we need a separate table for each dimension of the regressor. Typically, we will allow for dynamics in the residual and perform the equivalent of the ADF test (using the slightly different critical values in this case). Such a procedure is usually called a cointegration ADF test, or CADF test for short. Engle and Granger compared different tests and recommended the CADF test and supplied critical values based on Monte Carlo simulations in the case of just one regressor. They also extend those tables to the case of more than one regressor, and the most complete tables can be found in McKinnon (1991).

If the x_t series contain a trend or may contain a trend, then one should be careful to include a trend in the cointegration regression, otherwise the asymptotic critical values will be different. In the case of a one dimensional x_t , that include a deterministic trend, a regression of y_t on x_t that does not include that the trend will give you an asymptotically normal coefficient (this is not too surprising since deterministic trend always will dominate a unit root trend).

1.3.2 Estimation of the Parameters in Case of Cointegration

The issue of efficient estimation of parameters in cointegrating relationships is quite a different issue from the issue of testing for cointegration. Engle and Granger suggested the following simple two-step estimator. First, we estimate the static cointegrating relationship $y_t = \theta'x_t + e_t$, then we define $z_t = y_t - \hat{\theta}'x_t$, and finally estimate the error correction model

$$\Delta y_t = A_1(L)\Delta y_{t-1} + A_2(L)\Delta x_{t-1} + \gamma z_t + e_t. \quad (11)$$

The fact that $\hat{\theta}$ is superconsistent implies that the parameters of the lag polynomials have the same (asymptotically normal) distribution as they would have if θ had been known.

1.4 The Johansen ML Estimator

The best way of testing for unit roots is by using the system maximum likelihood estimator of Johansen (1988, 1990) is a test for cointegration restrictions in a VAR representation. This estimator also gives asymptotically efficient estimates of the cointegrating vectors (the betas) and of the adjustment parameters (the alphas). Johansen's method is the maximum likelihood estimator of the so-called reduced rank model. We will start with the AR (k) model:

$$\Delta y_t = \mu + \sum_{s=1}^{k-1} \Delta y_{t-s} + \Pi y_{t-k} + e_t, \quad (12)$$

which under the assumption of cointegration of order k can be written as

$$\Delta y_t = \mu + \sum_{s=1}^{k-1} \Delta y_{t-s} + \alpha \beta' y_{t-k} + e_t, \quad (13)$$

where α and β both have dimension $p \times k$. The number of parameters in the unrestricted model is $p + kp^2 + p(p+1)/2$. Let $Z_{0t} = \Delta y_t$, $Z_{1t} = (\Delta y_{t-1}, \dots, \Delta y_{t-k+1}, 1)'$ and $Z_{kt} = y_{t-k}$. Let's define the moment matrix as

$$M_{ij} = T^{-1} \sum_{t=1}^T Z_{it} Z_{jt}', \text{ where } i, j = 0, 1, k. \quad (14)$$

We first regress Z_{it} , $i = 0, k$ on Z_{1t} and get residuals R_{it} , $i = 0, k$. It appears as we are purging Δy_t and y_{t-k} of the short run parameters which can be considered noise in the cointegrating relation. We are also purging those variables of their mean, and if we want to include exogenous variables (e.g. dummy variables), they should also be included in the Z_{1t} vector. Let's denote the residual sum of squares from regressing Z_0 and Z_k on Z_1 as S_{ij} , $i, j = 0, k$ in other words:

$$S_{ij} = T^{-1} \sum_{t=1}^T R_{it} R_{jt}'. \quad (15)$$

The maximum likelihood estimator of α and β is a function of these residuals. Johansen shows that $\hat{\beta}$ can be found from choosing the eigenvectors $(\hat{v}_1, \dots, \hat{v}_r)$, where $\hat{V} = (\hat{v}_1, \dots, \hat{v}_p)$ are the eigenvectors of the equation:

$$|\lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}| = 0. \quad (16)$$

Normalized such that $\hat{V}' S_{kk} \hat{V} = I$, and ordered in the order of the corresponding eigenvalues such that $\hat{\lambda}_1 > \dots > \hat{\lambda}_p > 0$. Cointegration of order r implies that $\lambda_1 \neq 0, \dots, \lambda_r \neq 0$, while $\lambda_{r+1} = \dots = \lambda_p = 0$. Since the estimated eigenvalues are continuous random variables they are different from zero and each from other with a probability of 1. And it can be intuitively clear now that we want the eigenvectors

corresponding to the non-zero eigenvalues to be the estimators of the cointegrating vectors.

In order to find those eigenvalues left- and right-multiply the equation above by $S_{kk}^{-1/2}$ and we get the equivalent problem:

$$\left| \lambda - S_{kk}^{-1/2} S_{k0} S_{00}^{-1} S_{0k} S_{kk}'^{-1/2} \right| = 0. \quad (17)$$

The eigenvalues will be the ones that we are looking for. After being normalized, the eigenvectors (say u_i) such that $u_i' u_i = 1$, so we get:

$$(\hat{v}_1, \dots, \hat{v}_r) = (S_{kk}^{-1/2} u_1, \dots, S_{kk}^{-1/2} u_r). \quad (18)$$

In order to give some interpretation of this equation remember that the least squares $\hat{\Pi}$ can be obtained by regressing R_{0t} on R_{kt} . So the least squares estimate of Π is :

$$\hat{\Pi} = S_{kk}^{-1} S_{k0}. \quad (19)$$

Now we note that :

$$S_{kk}^{-1/2} S_{k0} S_{00}^{-1} S_{0k} S_{kk}'^{-1/2} = S_{kk}^{-1/2} S_{kk}^{-1/2} S_{k0} S_{00}^{-1/2} S_{00}^{-1/2} S_{0k} S_{kk}'^{-1/2} S_{kk}'^{1/2} = (S_{kk}^{-1/2} \hat{\Pi} S_{00}^{-1/2}) (S_{kk}^{-1/2} \hat{\Pi} S_{00}^{-1/2})' \quad (20)$$

The intuitively natural approach would be to consider the eigenvalues of $\hat{\Pi} \hat{\Pi}'$ and we can see that this is actually what the maximum likelihood algorithm does apart from the fact that $\hat{\Pi}$ has been normalized by left-multiplying by $S_{kk}^{-1/2}$ and right-multiplying by $S_{00}^{-1/2}$.

$$L_{\max}^{-2/T}(r) = |S_{00}| \prod_{i=1}^r (1 - \hat{\lambda}_i). \quad (21)$$

We can see that this is a function of the estimated eigenvalues where all the eigenvalues except the largest r eigenvector are set equal to zero. So for example, the test for one cointegrating vector against no cointegrating vector consists of testing whether the largest eigenvalue is significantly different from 0. Johansen further finds

$$\begin{aligned} \hat{\alpha} &= S_{0k} \hat{\beta}, \\ \{\hat{\Gamma}_1, \dots, \hat{\Gamma}_{k-1}, \hat{\mu}\} &= (M_{01} - \hat{\alpha} \hat{\beta}' M_{k1}) M_{11}^{-1}, \text{ and} \\ \hat{\Lambda} &= S_{00} - \hat{\alpha} \hat{\alpha}'. \end{aligned}$$

The likelihood ratio test statistic H for the hypothesis that $\Pi = \alpha \beta'$ is of rank r against the unrestricted model where Π has full rank p is

$$H = -2 \ln(Q) = -T \sum_{i=r+1}^p \ln(1 - \hat{\lambda}_i). \quad (22)$$

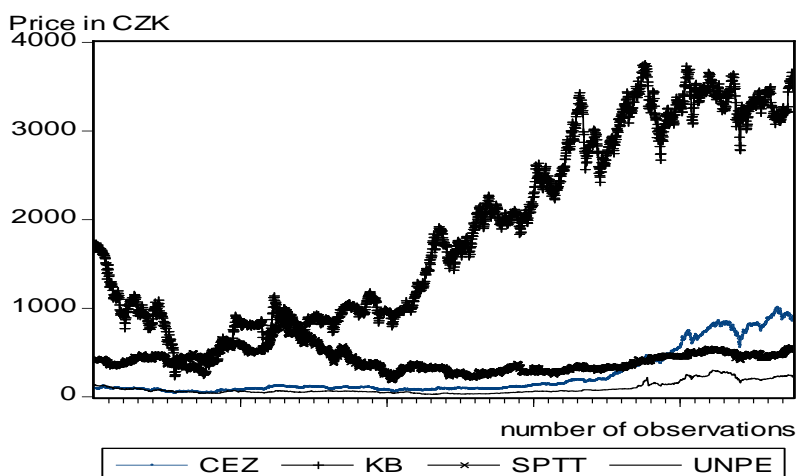
Note that the null hypothesis here is that there are $(p-r)$ unit roots. This corresponds to the simple residual based test previously, where we have $p = 2$ (if the X variable is one dimensional), and we test for 1 cointegrating relation, the null is then that there are 2 unit roots. This test statistic is often referred to as the trace statistic. Note that this statistic is expected to be close to 0 if there is at most r (linearly independent) cointegrating vectors. Another test that is often used is the λ – max test which looks at $-T \ln(1 - \hat{\lambda}_{r+1})$ – the idea being that if the $(r + 1)$ th eigenvalue can be accepted to be zero, then all the smaller eigenvalues can also. This test is a test of $r+1$ cointegrating vectors against r cointegrating vectors.

The asymptotic distribution of the likelihood ratio test is a functional of multivariate Brownian motion and is tabulated for values of p . Of course, sometimes we do not really want to test whether there is, say 3 cointegrating vectors against no integrating vectors, rather we want to make a decision as to what is the number of cointegrating vectors. In the situation where we directly want to test $r+1$ cointegrating vectors against r cointegrating vectors, we could use the λ – max test, but this test will not give us a consistent way of deciding the cointegration rank. A consistent way to do this, using the trace test, is to start by testing for zero cointegrating vector. If we reject zero cointegrating vectors, then we test for at most 1 cointegrating vectors. If this is not rejected, we stop and decide that $r = 1$ – if we reject this, we move on until we can no longer reject and stop there.

Even though there is a constant in the error correction representation (13), this may not translate into a deterministic trend in y_t . Johansen (1991) invented the likelihood ratio test for reduced rank in the case where there is a constant in the Error Correction Model but no trend in y_t . Johansen also discusses how to obtain a consistent test for the number of stochastic trends and for trend in y_t at the same time: in this case, all we have to do is to move the vector of ones in to Z_{kt} and delete it from Z_{lt} .

2. Application of Cointegration Technique on Czech Stocks

In this section, the cointegration of Czech stocks is tested on the most liquid stocks. They are stocks of Telefónica O₂ Česká republika (SPTT), České energetické závody (CEZ), Komerční banka (KB) and Unipetrol (UNPE). The daily data set of SPTT, CEZ, KB and UNPE from September 1st 1997 to February 28th 2007 is obtained from Reuters Česká republika. Another reason for choosing them is the existence of long time series needed for the analysis. The whole data set of these stocks includes 2384 daily observations of closing prices adjusted for dividends and splits. The courses of stocks prices in time are shown in Fig. 1 and their basic descriptive statistics are in Tab. 1.

Figure 1: The course of stock prices over time**Table 1: Basic descriptive statistics of investigated stock prices**

	CEZ	KB	SPTT	UNPE
Mean	240.07	1862.15	423.79	94.24
Median	105.50	1680.50	419.35	64.89
Maximum	1010.00	3760.00	967.30	299.50
Minimum	38.55	229.17	184.80	28.20
Standard deviation	261.76	1075.63	125.06	67.20
Skewness	1.54	0.27	0.90	1.44
Kurtosis	3.86	1.55	4.16	3.86
Jarque-Bera	1020.84	238.18	455.72	900.69
Probability	0	0	0	0
Sum	572324.10	4439376.00	1010321.00	224658.30
Sum Sq. Dev.	1.63E+08	2.76E+09	37269468	10762314

Although the ADF test and the cointegration test can be carried out straightaway on prices time series of the stocks of our interest, they were actually completed on their log price series in the statistical package Eviews 5 as recommended by Oomen (2006) for the purpose of stabilizing behaviour of these series. Logarithmization only reduces the scales of our original time series, but it does not change their character. The null hypothesis of the ADF test always is: a time series has a unit root. The results of the ADF test are shown in Tab. 2 to Tab. 5.

Table 2: Summary of ADF unit roots test for CEZ

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CEZ(-1)	0.000350	0.000527	0.664155	0.5067
D(CEZ(-1))	0.058698	0.020477	2.866540	0.0042
C	-0.000888	0.002692	-0.329907	0.7415
R-squared	0.003700	Mean dependent var		0.000930
Adjusted R-squared	0.002862	S.D. dependent var		0.022492
S.E. of regression	0.022459	Akaike info criterion		-4.752955
Sum squared resid	1.200028	Schwarz criterion		-4.745681
Log likelihood	5663.770	F-statistic		4.416983
Durbin-Watson stat	2.001748	Prob(F-statistic)		0.012170
Augmented Dickey-Fuller test statistic			t-Statistic	Prob.*
			0.664155	0.9914

*MacKinnon (1996) one-sided p-values.

Table 3: Summary of ADF unit roots test for KB

Variable	Coefficient	Std. Error	t-Statistic	Prob.
KB(-1)	-0.000636	0.000802	-0.792563	0.4281
D(KB(-1))	0.146874	0.020287	7.239683	0.0000
C	0.004918	0.005901	0.833395	0.4047
R-squared	0.021697	Mean dependent var		0.000308
Adjusted R-squared	0.020875	S.D. dependent var		0.027016
S.E. of regression	0.026733	Akaike info criterion		-4.404600
Sum squared resid	1.700121	Schwarz criterion		-4.397326
Log likelihood	5248.879	F-statistic		26.38124
Durbin-Watson stat	2.015503	Prob(F-statistic)		0.000000
Augmented Dickey-Fuller test statistic			t-Statistic	Prob.*
			-0.792563	0.8205

*MacKinnon (1996) one-sided p-values.

Table 4: Summary of ADF unit roots test for SPTT

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SPTT(-1)	-0.002893	0.001621	-1.784781	0.0744
C	0.017503	0.009748	1.795629	0.0727
R-squared	0.001336	Mean dependent var		0.000126
Adjusted R-squared	0.000917	S.D. dependent var		0.022691
S.E. of regression	0.022681	Akaike info criterion		-4.733769
Sum squared resid	1.224816	Schwarz criterion		-4.728921
Log likelihood	5642.286	F-statistic		3.185443
Durbin-Watson stat	2.022817	Prob(F-statistic)		0.074424
Augmented Dickey-Fuller test statistic			t-Statistic	Prob.*
			-1.784781	0.3885

*MacKinnon (1996) one-sided p-values.

The results of ADF test show us that we can not reject the null hypothesis that our time series of interest have a unit root at a level of significance of 5% (p-values always are higher). Another thing worth noticing is their differences do not have a unit root according to p-value in these tables. All stock prices time series of our interest have unit root, therefore they are all $I(1)$ processes and they might be cointegrated. Their cointegration of all combination from 2 to 4 of these stocks (there are 11 combinations of 2 to 4 stocks which can be made from 4 stocks of our interest) is tested by Johansen's procedure and only one combination of stocks CEZ, KB and UNPE is found to be cointegrated. The results of the cointegration test of this combination are shown below (Tab. 6 to Tab. 10).

Table 5: Summary of ADF unit roots test for UNPE

Variable	Coefficient	Std. Error	t-Statistic	Prob.
UNPE(-1)	-0.000517	0.000879	-0.588592	0.5562
D(UNPE(-1))	0.059091	0.020468	2.886976	0.0039
C	0.002447	0.003853	0.635081	0.5254
R-squared	0.003593	Mean dependent var		0.000214
Adjusted R-squared	0.002755	S.D. dependent var		0.025770
S.E. of regression	0.025734	Akaike info criterion		-4.480739
Sum squared resid	1.575481	Schwarz criterion		-4.473465
Log likelihood	5339.560	F-statistic		4.288828
Durbin-Watson stat	1.997543	Prob(F-statistic)		0.013827
Augmented Dickey-Fuller test statistic			t-Statistic	Prob.*
			-0.588592	0.8707

*MacKinnon (1996) one-sided p-values.

Table 6: Unrestricted cointegration rank test (trace test)

Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None*	0.010291	32.50034	29.79707	0.0238
At most 1	0.003308	7.890408	15.49471	0.4772
At most 2	3.42E-06	0.008146	3.841466	0.9276
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

Table 7: Unrestricted cointegration rank test (maximum eigenvalue test)

Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None*	0.010291	24.60993	21.13162	0.0155
At most 1	0.003308	7.882262	14.26460	0.3908
At most 2	3.42E-06	0.008146	3.841466	0.9276
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

Table 8: Unrestricted cointegrating coefficients (normalized by $b'S_{11}b=I$):

CEZ	KB	UNPE
-3.722893	1.780315	4.155830
-0.309171	1.902549	-0.353068
1.422079	-1.282161	0.518916

Table 9: Unrestricted adjustment coefficients (alpha):

D(CEZ)	-0.000389	0.000682	-3.42E-05
D(KB)	-0.002338	-0.000212	-2.36E-05
D(UNPE)	-0.001659	0.001119	5.82E-06
1 Cointegrating equation(s):		Log likelihood	16874.84

Both the trace tests and the maximum eigenvalue tests of Johansen's procedure have detected the presence of cointegration in price time series of stocks CEZ, KB and UNPE at level of significance 5% and the null hypothesis that there is no cointegration between them that can be rejected. A combination consisted of -3.722893 of stock of CEZ, 1.780315 of stock of KB and 4.155830 of stock of UNPE can be made and its behaviour in time will be mean-reverting and can be used for arbitrage purposes.

Table 10: Normalized cointegrating coefficients (standard error in parentheses)

CEZ	KB	UNPE
1.000000	0.000000	-1.306568
		(0.19029)
0.000000	1.000000	-0.397898
		(0.34102)

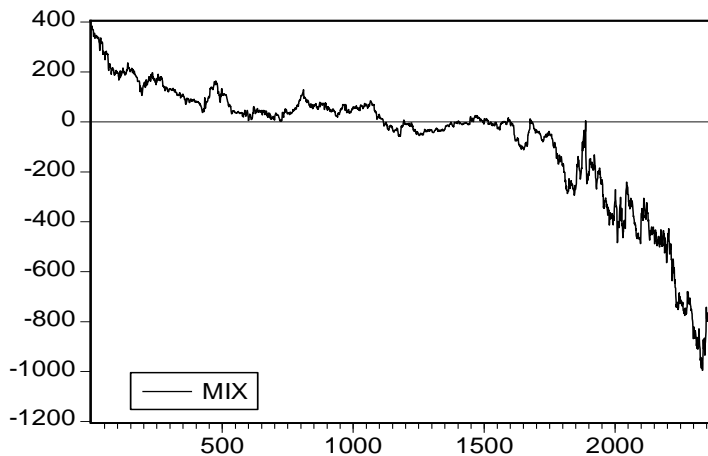
Table 11: Adjustment coefficients (standard error in parentheses)

D(CEZ)	0.001237 (0.00171)	0.000606 (0.00119)
D(KB)	0.008768 (0.00203)	-0.004565 (0.00142)
D(UNPE)	0.005830 (0.00196)	-0.000825 (0.00137)
2 Cointegrating equation(s):		Log likelihood
		16878.78

3. Conclusion

On the financial market, on one hand, the price of an individual stock may fluctuate randomly, as the efficient market hypothesis postulates. On the other hand, a certain combination of those stocks may share a long-term equilibrium when they are cointegrated. In this case, such relationship can be used to develop a statistical arbitrage profit. To determine the linkages between prices of the most liquid stocks on Czech stock market, cointegration analysis is a powerful tool and was used to detect a long-term equilibrium among them. For this purpose, the most liquid Czech stocks with long time series were chosen. They are stocks of Telefónica O₂ (previously named Český Telecom) (abbreviated SPTT), Komerční banka (KB), České energetické závody (CEZ) and Unipetrol (UNPE).

Figure 2: Behaviour of combination of CEZ, KB a UNPE over the course of time



To find out the cointegration relationship among them, first, all time series of stocks prices on the Czech stock market of our interest have been tested for unit roots by an ADF test to determine whether they are $I(1)$ processes. The presence of a unit root in all chosen time series was confirmed statistically. After that, all combinations of chosen stocks from 2 to 4 stocks were made, they are 11 in all and the Johansen's procedure of the cointegration test was carried on them. As far as the results of the cointegration test are concerned, cointegration in time series of stocks prices of CEZ, KB and UNPE was detected both by the trace test and the maximum eigenvalues test at the level of 5% of significance. This result enables us to state that even though the time series of stock price of CEZ, KB and UNPE are $I(1)$ processes and they may fluctuate randomly, however, their combination is definitely a $I(0)$ process which is mean reverting and can be used for an arbitrage trading strategy. This combination is following:

$$\text{MIX} = -3.722893 \text{ CEZ} + 1.780315 \text{ KB} + 4.155830 \text{ UNPE},$$

where CEZ, KB and UNPE are prices of these stocks respectively as stated before. To be understood the long-term equilibrium equation with ease, the term -3.722893 CEZ is

moved from the righthand side to the lefthand side of the equilibrium equation and it can be written as following:

$$\text{MIX} + 3.722893 \text{ CEZ} = 1.780315 \text{ KB} + 4.155830 \text{ UNPE}.$$

The long-term equilibrium value of combination MIX should be 0. It means that if the value of $1.780315 \text{ KB} + 4.155830 \text{ UNPE}$ is higher than the value of 3.722893 CEZ , then the value of combination MIX is positive. In this case, it will be falling to zero in the future. Therefore, we buy an amount of 3.722893 CEZ stocks and sell an amount of $1.780315 \text{ KB} + 4.155830 \text{ UNPE}$ and an arbitrage profit can be made in the future. If the value of our combination MIX is negative now, it will be increasing to zero later and we should buy the righthand side of the equation and sell the lefthand side given the original pattern of the individual stock behaviour is unchanged over the course of time. The behaviour of time series MIX during the course of time is shown in Figure 2. A long period from 500th trading day to 1850th trading day, when the value of the portfolio was fluctuating around zero, can be seen at the figure.

To effectively exploit this mean reverting behaviour of combination MIX, some challenges should be dealt with. First of all, it would be more useful for investors if time series revert to their mean so quickly as possible because in this case, investors can realize the trading strategy sooner rather than later. Secondly, it will be more comfortable to them if the variation of the combined time series is large enough so that trading opportunities could be executed when the returns cover all transactional and informational costs. The last thing to be mentioned is the stability of this pattern in the future. We have only investigated the in-sample behaviour of stock prices of our interest. Of course, they may not follow this pattern in the future. Then this pattern may not be useful to investors in the future at all.

References

- [1] DICKEY, D. A.; FULLER, W. A. (1979). Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Associations*, 1979, vol. 74, no. 366, pp. 427–431.
- [1] DICKEY, D. A.; FULLER, W. A., (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 1981, vol. 49, no. 4, pp. 1057–72.
- [2] ENGLE, R. F.; GRANGER, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica*, 1987, vol. 55, no. 2, pp. 251–276.
- [3] JOHANSEN, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 1988, vol. 12, no. 2–3, pp. 213–254.
- [4] JOHANSEN, S.; JUSELIUS, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration With Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, 1990, vol. 52, no. 2, pp. 169–210.
- [5] MACKINNON, J. G.; HAUG, A.; MICHELIS, L. (1999). Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration. *Journal of Applied Econometrics*, 1999, vol. 14, no. 5, pp. 563–577.

- [6] MACKINNON, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 1996, vol. 11, no. 6, pp. 601–618.
- [7] OOMEN, R. (2006). *Forecasting financial markets*. London : London School Summer School notes, 2006
- [8] STOCK, J. S.; WATSON, M. W. (1988). Variable Trends in Economic Time Series. *Journal of Economic Perspectives*, 1988, vol. 2, no. 3, pp. 147–174.

Testování kointegrace na českém akciovém trhu

Tran Van Quang

Abstrakt

Na základě výsledků kointegrační analýzy časových řad cen nejlikvidnějších akciových titulů bylo nalezeno dlouhodobé rovnovážné vztahy mezi cenami akcií Komerční banky, Českých energetických závodů a Unipetrolu. Časové řady cen těchto akcií mají jednotkový kořen a jsou korelované. Byly nalezena jedna kombinace těchto akcií, která je mean reverting a může být použita ke statistické arbitráži. Nicméně, k úspěšnému využití této možnosti v praxi je třeba vzít v úvahu další aspekty. Je nutné ještě zkoumat rychlost reverze k průměru, míru variability a chování mimo zkoumaný vzorek této kombinace zmíněných akcií.

Klíčová slova: jednotkový kořen; kointegrace; časové řady cen českých akcií; mean reverting.

Testing Cointegration for Czech Stock Market

Abstract

Based on cointegration analysis of daily data of the most liquid Czech stock from September 1, 1997 to February 28, 2007, a long run equilibrium relationship was revealed to exist between prices of stocks of Komerční banka (KB), České energetické závody (CEZ) and Unipetrol (UNPE). Prices time series of these stocks have a unit root and are cointegrated. There is a unique combination of these stocks which is mean reverting and can be used to achieve statistical arbitrage. However, in order to exploit this possibility, a number of challenges need to be dealt with. Investors should take into account the speed of the mean reversion rate, the size of the variation and the stability of the out of sample behaviour of this combination of these stocks.

Key words: unit root; cointegration; Czech stocks prices time series; mean reverting.

JEL classification: C32, E32.