TIME SERIES ANALYSIS OF THE EMPLOYMENT RATE IN THE CZECH REPUBLIC IN 2005–2016

DAGMAR BLATNÁ
University of Economics, Prague, Faculty of Informatics and Statistics,
Department of Statistics and Probability,
W. Churchill Sq. 4, Prague, Czech Republic
e-mail: blatna@vse.cz

Abstract

When analysing labour market trends, the employment rate serves as a key social indicator, defined as a ratio measuring the proportion of the employed working-age population. The rate of employment is one of the headline indicators in the Europe 2020 strategy for smart, sustainable and inclusive growth. The development of the employment rate both in the EU and the Czech Republic in 2005–2016 is examined with reference to achieving the objectives set in the Europe 2020 strategy, the Czech Republic having already exceeded the 2020 national target. The aim of the paper is to analyse the relationship between the employment rate and selected socio-economic indicators such as real per capita GDP growth rate, inflation rate, educational attainment, level of lifelong learning, social benefits, etc., in the Czech Republic. The regression analysis was used as the main tool of analysis. The employment rate regression analysis was carried out applying multivariate non-stationary time series methodology. The analysis covers the period between 2005 and 2016.

Keywords*: employment rate, non-stationary time series, regression analysis, cointegration*

JEL Codes*: C32, E24*

1. Introduction

The employment rate (ER) is one of the headline objective indicators of the Europe 2020 strategy for smart, sustainable and inclusive growth. The employment target is to increase the rate of employment of the 20–64-age group to at least 75 percent by 2020, the national objective for the Czech Republic being set at the same level.

According to the Eurostat definition, the employment rate is calculated by dividing the number of employed people aged 20 to 64 by the total population of the same age. The indicator is based on the EU Labour Force Survey (EU LFS) covering the entire population living in private households, excluding those in collective households such as boarding houses, residential homes and hospitals. The employed population consists of those persons who do any work for pay or profit for at least one hour during the reference week, or do not perform any work but have a job (are in employment) from which they are temporarily absent; see EU LFS. The employment rate is defined as a ratio in which existing labour resources (people available to work) are being used.

The present paper aims to provide the results of the regression analysis of the time series relationship of the employment rate and selected socio-economic indicators in the Czech Republic. The regression analysis was conducted employing the theory of multivariate non-stationary time series.

2. Methodology

Using time series data during the estimation of the regression model, it is necessary to know whether the variables are stationary or non-stationary. Stationary time series are those with an autoregressive model of zero order I(0), also known as short memory ones, the stationary process oscillating around a constant long-term mean and having a constant variance independent of time. Non-stationary time series, on the other hand, have means, variances and covariances that change over time. Difference time series are integrated, denoted as I(d), where d is the order of integration indicating the number of unit roots contained in the series or the number of differencing operations needed to make the series stationary. Non-stationary time series are generated by an autoregressive process of order one I(d) and referred to as long memory ones.

It is important to check the stationarity before using the time series in regression analysis. For stationarity testing, the so-called unit root tests of the autoregressive parameter ϕ_1 are performed, the most common being the Dickey-Fuller test. There are several variations of the basic Dickey-Fuller test with or without a drift and trend, respectively. The Augmented Dickey-Fuller (ADF) test takes into account the possibility that the process is a higher-order autoregressive one.

The ADF test was used to verify the null hypothesis

H_0: $\phi_1 = 1$ for non-stationary I(1) time series and

H_1: $|\phi_1| < 1$ for stationary I(0) time series.

ADF test statistic is defined as

$$t = \frac{\hat{\phi}_1 - 1}{S_{\hat{\phi}_1}}, \tag{1}$$

where $\hat{\phi}_1$ is an estimate of the autoregressive parameter of the model $y_t = \phi_1 y_{t-1} + a_t$, $S_{\hat{\phi}_1}$ is an estimate of the standard error of $\hat{\phi}_1$ and a_t is a non-systematic component with white noise characteristics, i.e. the series of uncorrelated random variables $\operatorname{cov}(a_t, a_{t-k}) = 0$, probability distribution $N(0, \sigma_a^2)$ with zero mean and constant variance $D(a_t) = \sigma_a^2$. The test statistic follows the Dickey-Fuller distribution; for critical values, see Dickey and Fuller (1979). The null hypothesis of a unit root is rejected in favour of the stationary alternative if the test statistic is more negative than the critical value. For details, see Arlt and Arltová (2009), Hušek (2007), Caner and Kilian (2001), Dickey and Fuller (1981), Elliot *at al*, (1996) or Phillips (1987).

In the regression analysis, the time series of both variables must be of the same order integrated process. The “classical” regression model can be employed when the analysed time series are zero order stationary ones. Applying the unit root tests, it can be concluded that the series are not of the same order, showing no relationship.

When unit root tests of a linear combination of two non-stationary time series do not reject the non-stationarity of the non-systematic component, there is only spurious regression, suggesting relationships even when there are none. A standard approach to building regression models for non-stationary variables is to differentiate the series in order to achieve stationarity and analyse the relationship between the stationary variables, the information on the long-run relationship being lost. The relation between the two non-stationary time series exists only if their linear combination is stationary. Then, there is a cointegration relationship considered as a long-term one.

Validation of the calculated regression model is performed using diagnostic tests of the non-systematic component of the model. To verify the normality, the Jarque-Bera test is carried out. This is a goodness-of-fit test of whether the sample data show the skewness and kurtosis matching a normal distribution; the null hypothesis is a joint one assuming that both the skewness and excess kurtosis are zero. The Jarque-Bera statistic has a χ^2 distribution with two degrees of freedom under the null hypothesis of normally distributed errors; see, e.g., Jarque and Bera (1980). Testing for autocorrelation in the errors in a regression model, Breusch-Godfrey Serial Correlation LM test was conducted with the null hypothesis that there is no serial correlation of any order up to p; see Breusch and Godfrey (1986). Autoregressive conditional homoscedasticity in the residuals was verified by ARCH(1) LM test; see Engle (1982).

When the non-systematic component is auto-correlated, the Autoregressive Distributed Lag (ADL) model with time-shifted variables is utilized for estimation. It can be written as

$$Y_t = c + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{i=1}^{q} \boldsymbol{\beta}_j (B^{j-1}) \mathbf{X}_t + a_t \,. \tag{2}$$

where Y_t is the dependent variable in time t, Y_{t-i} is the dependent variable in delay t-i, where i = 1,…, t-p, $\mathbf{X}_t$ is the matrix of explanatory variables in time t and delay t-j, where j = 1, t-q, α and $\boldsymbol{\beta}$ are parameters in the model and c is a constant and a_t has a white-noise characteristic, i.e. it is the series of uncorrelated random variables with $\mathrm{cov}(a_t, a_{t-k}) = 0$, the probability distribution $\mathrm{N}(0, \sigma_a^2)$ with zero mean and constant variance $D(a_t) = \sigma_a^2$. For more details, see Arlt (1998), Arlt and Arltová (2009), Hušek (2007), Hendry *at al*. (1984).

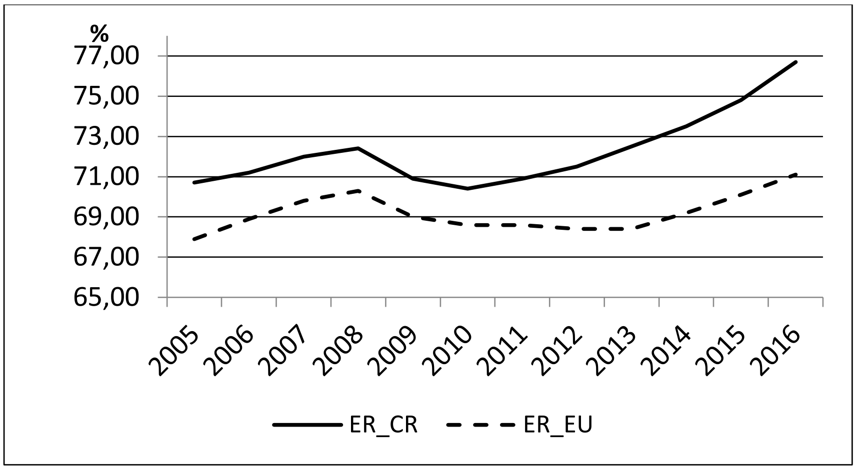
3. Analysis of indicator ER development from 2005 to 2016

The analysis is based on 2005–2016 data adopted from the Eurostat database. Calculations were performed using Excel, E-Views 9 and Statgraphics statistical software.

As you can see in Figure 1, employment rates are sensitive to the economic cycle. The EU employment rate was growing steadily during the years prior to the economic downturn, peaking at 70.3 percent in 2008. The crisis, however, reversed the positive trend. Having returned to the 2006 level in 2009, the employment rate in the EU-28 continued to decline to 68.4 percent in 2013. Since then, the trend has changed, starting to rise continuously to 71.1 percent in 2016, showing an average annual growth of 1.3 percent. Nevertheless, the EU-28 employment rate still remains below the target value of 75 percent.

The employment rate in the Czech Republic was higher than in the whole European Union throughout the analysed period. Its development before the crisis was similar in the Czech Republic to that in the EU, reaching 72.4 percent in 2008. But it was affected by the economic downturn of 2008 and the following recession. The crisis impact, however, was not long-term, the ER indicator falling to the lowest value (70.4 percent) in 2010. The Czech Republic returned to an upward trend, its 2013 employment rate being higher than before the crisis and increasing continuously to 76.7 percent in 2016. Thus, the 2020 national target was exceeded as early as in 2016.

Figure 1: ER indicator development in the EU-28 and the CR, 2005–2016



Source: Eurostat data; The autor's work.

Table 1 displays the basic statistics of the ER indicator. A significant difference between the rate of employment in the Czech Republic and the EU-28 was confirmed using the non-parametric Wilcoxon rank-sum test (W = 4.0000, p = 0.0001).

Table 1: Basic employment rate statistics for the Czech Republic and the European Union

	Average	Median	Variance	Standard deviation	Minimum	Maximum
ER_EU	69.1917	68.95	0.8917	0.9443	67.9	71.1
ER_CR	72.2917	71.75	3.5572	1.8860	70.4	76.7

Source: Eurostat data; The autor's work.

4. Analysis of relationship between the employment rate in the Czech Republic and the EU-28

Using the Augmented Dickey-Fuller test, non-stationarity of both ER time series has been identified; see Table 2. Since the time series are of the same type of integration, their linear combination type of integration is to be carefully examined. The Augmented Dickey-Fuller test of a non-systematic component of the linear combination does not reject non-stationarity (t_{ADF} = -1.3196, p = 0.1611), and so the relations between ER_CR and ER_EU represent the case of spurious regression. In the next step of the analysis, the correlation between differenced time series was studied. Both differenced ER time series are stationary (d(ER_CR): t_{ADF} = -0.6725, p = 0.4002; d(ER_EU): t_{ADF} = -1.6721, p = 0.0881), their regression model taking the form of

$$d(ER\_CR)_t = 0.2489 + 1.0192\, d(ER\_EU)_t\,. \tag{3}$$

The non-systematic component of this model is stationary (t_{ADF} = -1.9935, p = 0.0489), the Rsq coefficient of determination being 0.6940. Diagnostic tests show that the model is statistically acceptable, comfirming that the non-systematic component is not autocorrelated (Breusch-Godfray LM test F = 1.4406, p = 0.2992), has a normal distribution (Jarque-Bera test JB = 0.7539, p = 0.6859) and is homoscedastic (ARCH = 1.1962, p = 0.6695). This regression model, however, captures only the short-time relationship between the employment rate in the Czech Republic and the EU-28.

Table 2: Diagnostic tests of d(ER_CR)~ d(ER_EU) model's non-systematic component

Test	Test statistic	Prob.	p-value
Breusch-Godfrey serial correlation LM test	1.4406	Prob F(2,7)	0.2992
Normality test: Jarque-Bera	0.7539	Prob	0.6859
Heteroscedasticity test: ARCH test	1.1962	Prob F(1,8)	0.6695

Source: Eurostat data; The autor's work.

5. Regression analysis of relationship between the employment rate and selected indicators in the Czech Republic from 2005 to 2016

Regression analysis of multivariate time series was done, the employment rate representing the dependent variable. The following socio-economic indicators were employed as explanatory variables: the real per capita GDP growth rate, inflation rate (HICP), unemployment rate, tertiary/secondary/lower secondary education attainment, lifelong learning, household saving rate, social benefits, average salary growth, people living in households with very low work intensity, and young people neither in employment nor in education and training.

Initial analysis applying a unit root test, namely the Augmented Dickey-Fuller one, identified stationary and non-stationary time series, respectively. In Table 3, ADF test values and relevant p-values are presented.

Table 3: Unit root tests of selected time series

Indicator	Abbreviation	t_{ADF}	p-value	Stationarity/ non-stationarity
Employment rate in the CR	ER	0.9522	0.8948	N
Employment rate in the EU	ER_EU	- 1.7499	0.3803	N
Real GDP per capita growth rate	GGDP	-2.0883	0.0375	S
Inflation rate (HICP)	IR	-7.1295	0.0022	S
Unemployment rate	UR	-1.2280	0.1721	N
Tertiary education attainment	TE	-3.5853	0.0858	N
At least secondary education attainment	YE	-3.1793	0.1396	N
Lower secondary education attainment	LSE	-1.4020	0.5422	N
Lifelong learning	LLL	-1.5884	0.4349	N
Household saving rate	HSR	-2.7565	0.0987	N
Social benefits	SB	-1.6926	0.4078	N
Average salary growth	ASG	-0.1337	0.2175	N
People living in households with very low work intensity	LWI	-1.2258	0.1885	N
Young people neither in employment	YNE	-3.4720	0.0339	S

nor in education and training				

Source: Eurostat data; The autor's work.

Upon performing the unit root tests, the non-stationarity of dependent variable ER time series was identified (t_{ADF} = 0.9522, p = 0.8948). Because the time series are of the same integrated process type, it can be concluded that there is no relationship between the employment rate and indicators whose time series are stationary, such as GDP growth, inflation rate and the proportion of young people who are neither in employment nor in education and training. Since most of the selected explanatory variables are non-stationary too, they were included in the regression analysis with the dependent variable ER.

A lot of regression models with different combinations of explanatory variables were calculated. We, however, present only the two best regression models producing statistically satisfactory results. In Table 4, the regression model with explanatory variables TE (tertiary education attainment) and ASG (average salary growth) is shown.

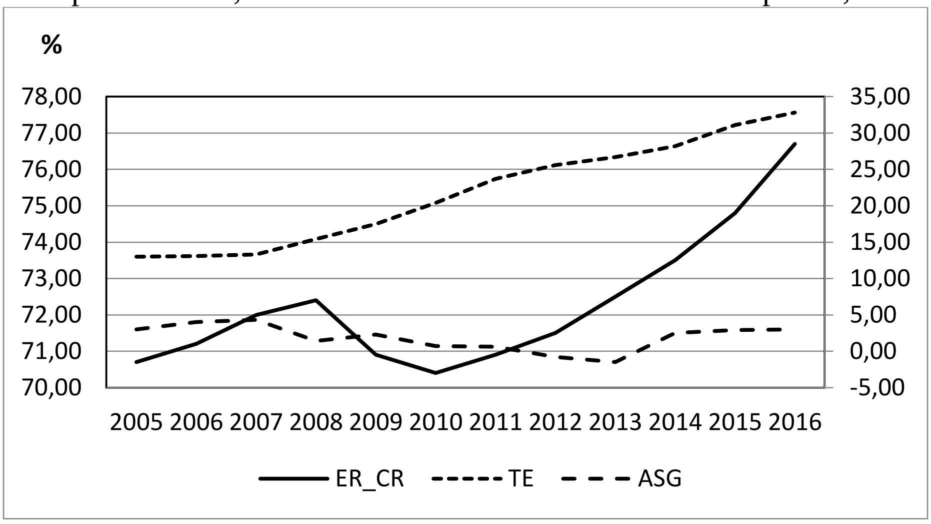
Table 4: Regression model ER~ TE + ASG

Variable	Coefficient	Std. error	t-Statistic	p-value
C	66.1578	1.1560	57.2262	0.0000
TE	0.2353	0.0443	5.3092	0.0005
ASG	0.5465	0.1766	3.0934	0.0129

Source: Eurostat data; The autor's work.

Owing to the non-stationarity of dependent and explanatory variables, the non-stationarity of the non-systematic component of their linear combination was tested applying the ADF test. Because the ADF test rejects non-stationarity (t = -2.7258, p = 0.0114), the time series are to be cointegrated, the model expressing the long-term relationship. Thus, we can say that the employment rate is positively correlated with TE and ASG, an increase in the proportion of people with tertiary education attainment (TE) and average salary growth (ASG) leading to the growth of the employment rate. For the development of the dependent variable ER and explanatory variables TE and ASG, see Figure 2.

Figure 2: Development of TE, ASG and ER indicators in the Czech Republic, 2005–2016



Source: Eurostat data; The autor's work.

The model explains 76.95 percent of time series ER dynamics, its determination index is 0.7695, the F-test being significant (F = 15.0258, p = 0.0013). It is a statistically acceptable model, diagnostic tests comfirming that the non-systematic component has a normal distribution (Jarque-Bera test JB = 1.1866, p = 0.5525), it is homoscedastic (ARCH = 1.3312, p = 0.2783) and not autocorrelated (Breusch-Godfray LM test F = 0.8138, p = 0.4811); see Table 5.

Table 5: Diagnostic tests of *ER~ TE + ASG* model's non-systematic component

Test	Test statistic	Prob	p-value
Breusch-Godfrey serial correlation LM test	0.8138	Prob F(2,7)	0.4811
Normality test: Jarque-Bera	1.1866	Prob	0.5525
Heteroscedasticity test: ARCH test	1.3312	Prob F(,9)	0.2783

Source: Eurostat data; The autor's work.

In the second model for dependent variable ER, explanatory variables SB (social benefits) and YE (at least secondary education attainment) are included. This model is presented in Table 6, the development curves for SB, YE and ER indicators for the given period being plotted in Figure 3. Both the dependent and explanatory variables are non-stationary. Applying the ADF test, the stationarity of the non-systematic component of the regression model has been confirmed (t_{ADF} = -3.4733, p = 0.0032). Again, it is cointegration and the relationship can be considered as long term.

Table 6: Regression model ER ~ SB + YE

Variable	Coefficient	Std. error	t-Statistic	p-value
C	-52.5796	12.9601	-4.0570	0.0029
SB	-2,2008	0.3372	-6.5256	0.0001
YE	1,6580	0.1637	10.1234	0.0000

Source: Eurostat data; The autor's work.

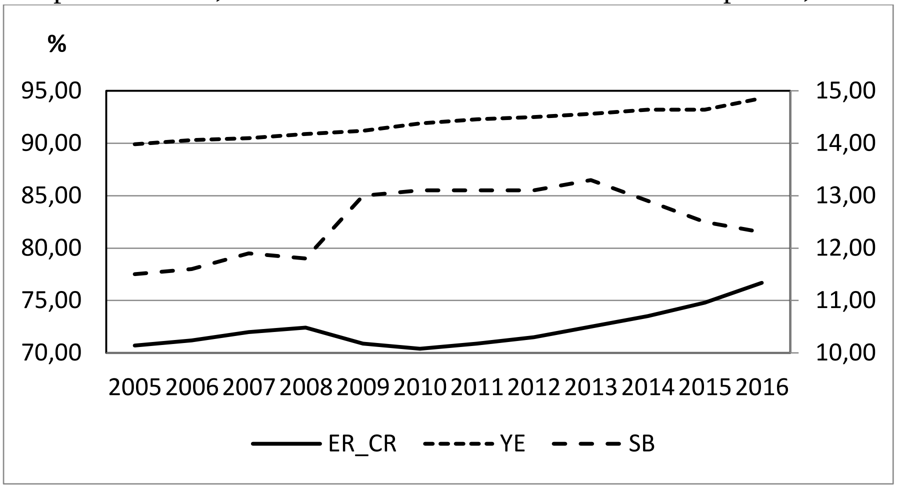
Thus, we can say that the employment rate is negatively correlated with social benefits (SB) and positively with the proportion of people with at least secondary education attainment. The model explains 91.9 percent of time series ER dynamics, the F-test being significant (F = 51.3867, p = 0.0000). This model is statistically acceptable, diagnostic tests comfirming that the non-systematic component is not autocorrelated (Breusch-Godfray LM test F = 0.1189, p = 0.8896), has a normal distribution (Jarque-Bera test JB = 0.9244, p = 0.8299) and is homoscedastic (ARCH = 0,1614, p = 0,6972); see Table 7.

Table 7: Diagnostic tests of ER~ SB + YE model's non-systematic component

Test	Test statistic	Prob	p-value
Breusch-Godfrey serial correlation LM test	0.1189	Prob F(2,7)	0.8896
Normality test: Jarque-Bera	0.9244	Prob	0.8299
Heteroscedasticity test: ARCH test	0.1614	Prob F(1,9)	0.6972

Source: Eurostat data; The autor's work.

Figure 3: Development of SB, YE and ER indicators in the Czech Republic, 2005–2016



Source: Eurostat data; The autor's work.

6. Conclusion

The rate of employment is defined as the proportion of labour resources (people available to work) being employed. According to the Eurostat definition, it is calculated by dividing the number of employed persons aged 20 to 64 years by the total population of the same age group.

The development of the employment rate in the European Union and the Czech Republic in 2005–2016 was examined in view of the objectives specified in the Europe 2020 strategy, the Czech Republic having already exceeded the 2020 national target.

The employment rate trend in both the EU and the CR was sensitive to the economic cycle. The CR rate of employment was higher than that in the EU-28 throughout the period, significant difference being verified with the use of the non-parametric Wilcoxon rank-sum test.

Because the unit root tests of the linear combination of non-stationary time series ER_CR and ER_EU do not reject non-stationarity of the non-systematic component, the spurious regression exists. The regression model based on differenced time series expresses only the short-time relationship between the employment rate in the Czech Republic and the EU 28 countries.

Many regression models with different combinations of socio-economic explanatory variables having been calculated for the dependent variable ER in the Czech Republic, two most appropriate statistically satisfactory regression models were presented, namely those for TE (tertiary education attainment) and ASG (average salary growth) explanatory variables and for SB (social benefits) and YE (people with at least secondary education) variables, respectively. A growing proportion of people with tertiary education and average growth of salary leads to an increasing employment rate, which is negatively correlated with social benefits (SB) and positively correlated with the proportion of people who attained at least secondary education.

In both models, the time series must be cointegrated. The regression models represent the long-term relationship between the rate of employment and explanatory variables examined.

Acknowledgment

The paper was written with the help of long term institutional support by the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Arlt, J. (1998). Časové řady typu I(0) a I(1). Praha: Acta oeconomica pragensia, 6(2), pp. 7-11, Praha, ISSN 0572-3043.

[2] Arlt,J., Arltová, M. (2009). Ekonomické časové řady. Praha: Professional Publishing. ISBN 978-80-247-1319-9.

[3] Breusch, T., Goldfray, L.G. (1986). Data transformation Tests. Economic Journal, 96, pp. 47-58.

[4] Caner, M., Kilian, L. (2001). Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate. Journal of International Money and Finance, 20, pp. 639-657.

[5] Dickey, D., Fuller,W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association, 74, pp. 427-431.

[6] Dickey, D., Fuller, W.(1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, Econometrica, 49, 1057-1072.

[7] Elliot, G., Rothenberg, J., Stock, J.H. (1996). Efficient Tests for an Autoregressive Unit Root. Econometrica, 64, 813-836

[8] Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50,987-1007.

[9] Granger, C. W. J., Newbold, P. (1974). Spurious Regression in Econometrics. Journal of Econometrics, 2(2), s. 111–120. ISSN 0304-4076.

[10] Hendry, D., Pagan, A., Sargan, J. (1984). Dynamic Specifications. Chapter 18 in Handbook of Econometrics, Vol.II (ed., Z. Griliches and M. Intriligator), North Holland

[11] Hušek, R. (2007). Ekonometrická analýza. Praha: Oeconomica. ISBN 978-80-245-1300-3.

[12] Jarque, C., Bera, A. (1980). Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals, Economics Letters 6, pp.255-259.

[13] Phillips, P.C.B. (1987). Time Series Regression with a Unit Root. Econometrica 55, pp. 277-301.