

## HOW TO ESTIMATE LONG-RUN RELATIONSHIPS IN ECONOMICS: AN OVERVIEW OF RECENT DEVELOPMENTS

Utku Utkulu<sup>(\*)</sup>

### ABSTRACT

*This paper provides an overview of an important and relatively recent approaches to estimate long-run economic relationships using 'cointegration', a technique becoming widely used in macroeconomic modelling. The appeal of the cointegration analysis is that it simply provides an effective formal framework for estimating (also testing and modelling) long-run economic relationships from time-series data. The paper is selective in nature. In particular, it describes the estimation procedures of those which have become increasingly popular in the recent applied literature.*

### I. Introduction

Nonstationary (trended) time-series data can be regarded as potentially a major problem for empirical econometrics. It is well known that trends, either stochastic or deterministic, may cause spurious regressions, uninterpretable student-t values and other statistics, goodness of fit measures which are 'too high' and, as a rule, make regression results rather difficult to evaluate. However, most macroeconomic time-series are subject to some type of trend. Some researchers have suggested a remedy, namely to difference a series successively until stationary is achieved. Nevertheless, it has been proved that 'differencing' results in a loss of some valuable long-run information in the data.

A real breakthrough in the time-series econometrics came with the concept of 'cointegration' in the early 1980s. The concept was first introduced by Granger (1981). Afterwards, Engle and Granger (1987), in their seminal paper, provided a firm theoretical base for representation, testing, estimating and modeling of cointegrated nonstationary time-series variables. Since then, there has been an explosion of research on cointegration and related fields (for a recent survey on the subject, see e.g. Utkulu (1994)). Cointegration analysis allows nonstationary data to be used so that spurious results are avoided. It also provides applied econometricians an effective formal framework for testing and estimating long-run models from actual time-series data.

Some recent developments in estimating long-run economic relationships are discussed in the next section briefly. Section III examines some of the alternative approaches. Final section draws some conclusions.

<sup>(\*)</sup> Dr., Ass. Prof. Dokuz Eylül University, Economics Department.

## II. Recent Developments

The existence of cointegration between, say, two macroeconomic variables implies "a true long-run economic relationship" which prevents the residuals (see equation 1) becoming larger and larger in the long-run. A number of different methods for estimating the long-run equation and the short-run error-correction model (ECM) are suggested in the literature. Such models currently represent the most attracted approach to cases where researchers seek to incorporate both the economic theory and relating to the long-run (equilibrium) relationship between variables, and short-run adjustment (disequilibrium) behavior. Among these various different approaches, the Engle-Granger (EG) type of static long-run regression has become a widely applied method since it was introduced by Engle and Granger (1987). Some suggest that the estimates of the EG type static long-run ordinary-least-squares (OLS) regression parameters are both consistent and highly efficient (see e.g. Stock (1987)). The EG static long-run regression has not gone unchallenged. For instance, Banerjee *et al.* (1986) stress that ignoring the lagged terms in small samples is likely to create a bias in the estimated parameters.<sup>1</sup>

In an attempt to estimate alternative cointegrating regressions, many have been interested in adding dynamic (either differences or lags) components (see e.g. Charemza and Deadman (1992), Cuthbertson *et al.* (1992), Inder (1993), Phillips and Loretan (1991), Saikkonen (1991), Wickens and Breusch (1988)). Others have been more concerned with the appropriate corrections and modifications to the static parameter estimates (see e.g. Engle and Yoo (1991), Park and Phillips (1988), Phillips and Hansen (1990), West (1988)). Since these two groups of critics emphasize different aspects of the problem, they naturally lead to different solutions. Some (see e.g. Banerjee *et al.* (1986)) favor for estimating long-run parameters in an unrestricted error-correction model (ECM) form incorporating all the dynamic components, while others (see e.g. Phillips and Hansen (1990)) advocate the use of some corrections to the OLS estimator to eliminate the bias (Phillips and Hansen call this 'the modified OLS estimator').

In a recent Monte Carlo study, Inder (1993) contributes to the debate by comparing various estimators of the long-run parameters. It is suggested that

---

<sup>1</sup> The findings of Banerjee *et al.* (1986) have two important implications. First,  $R^2$  is only important as an indicator of the degree of bias of the estimates. They show that the bias is large when  $R^2$  is considerably less than one. They also suggest that the EGM should be completed in an extra confirming step by that the residuals of the estimated short-run equation are stationary. Second,  $R^2$  might have a role as a guide for choosing the appropriate correction for the bias. The closer the  $R^2$  to one is the less bias and the more appropriate correction. Also Blough (1988) is concerned about the low power of the cointegration tests in small samples. The results of Blangiewicz and Charemza (1990) are, however, more promising as regards the power of cointegration tests in small samples.

estimates which include the dynamics are much more reliable, even if the dynamic structure is overspecified. In addition, Inder criticizes the Phillips-Hansen paper (Phillips and Hansen (1990)), which puts a strong case for the modified-OLS (i.e. the semiparametric correction applied to OLS estimates) in preference to the unrestricted error-correction model estimator. Instead, he proposes the fully-modified unrestricted error-correction model estimator which gives precise estimates and valid t-statistics even in the presence of endogenous explanatory variables (for full details of the estimators, see Inder (1993), Phillips and Hansen (1990)).

In addition to these single equation-based approaches to estimate the long-run equilibrium, Johansen (1988, 1991) and Johansen and Juselius (1990) provide a systems-based approach. The main advantage of the Johansen Maximum Likelihood (ML) method is that it enables one to determine the number of existing cointegrating (i.e. long-run) relationships among the variables in hand. It is important to note that single equation-based approaches assume the uniqueness of the cointegrating vector. Some of the considerable methods used in the literature are assessed in the following section.<sup>2</sup>

### III. Some Alternative Approaches to Estimate Long-run Relationships in Economics

#### A. The Engle-Granger Two-Step Modeling Method (EGM)

Among a number of alternative methods, the EGM, originally suggested by Engle and Granger (1987), has received a great deal of attention in recent years. One of its benefits is that the long-run equilibrium relationship (i.e. the cointegrating regression) can be modeled by a straightforward regression involving the levels of the variables. In the first step, all dynamics are ignored and the cointegrating regression is estimated by the OLS. Let us now write the long-run (cointegrating) regression:<sup>3</sup>

$$C_t = \beta Y_t + u_t \quad (1)$$

where both  $C_t$  and  $Y_t$  are nonstationary variables and integrated of order one (i.e.  $C_t \sim I(1)$  and  $Y_t \sim I(1)$ ). In order for  $C_t$  and  $Y_t$  to be cointegrated, the necessary condition is that the estimated residuals from Eq. (1) should be stationary (i.e.  $u_t \sim I(0)$ ). Since the variables in Eq. (1) are nonstationary (which causes the famous 'spurious regression problem!'), one should place little faith in the standard error estimates (and thus t-statistics) in the cointegrating regression.

<sup>2</sup> For extensive reviews on the field, see especially Inder (1993), Phillips and Loretan (1991).

<sup>3</sup> Here, for simplicity, we are only concerned with the two-variable case. Extension for multivariate case is straightforward.

Therefore, little importance can be attributed to the standard statistical tests on  $R^2$  or t-statistics of the estimated coefficients unless a correction procedure is employed to eliminate the bias. Different types of corrections are reported by Engle and Yoo (1991), Park and Phillips (1988), Phillips and Hansen (1990) and West (1988).

The second step involves estimating a short-run model with an error-correction mechanism (ECM) by the OLS. According to the Granger Representation Theorem (GRT), if a number of variables, such as  $C_t$  and  $Y_t$ , are cointegrated, then there will exist an ECM relating these variables and *vice versa*. Conditional on finding cointegration between  $C_t$  and  $Y_t$ , the estimate of  $\beta$  from the first step long-run regression (1) may then be imposed on the following short-run model with the remaining parameters being consistently estimated by the OLS. In other words, we retrieve the estimate of  $\beta$  from Eq. (1), and insert it in place of  $\beta$  in the error-correction term ( $C_t - \beta Y_t$ ) in the following short-run equation:

$$\Delta C_t = \alpha_1 \Delta Y_t + \alpha_2 (C_t - \beta Y_t)_{t-1} + \varepsilon_t \quad (2)$$

where  $\Delta$  represents first-differences and  $\varepsilon_t$  is the error term. Alternatively, in practice, since  $C_t - \beta Y_t = u_t$ , one can substitute the estimated residuals from Eq. (1) in place of the error-correction term, as the two will be identical. Note that the estimated coefficient  $\alpha_2$  in the short-run Eq. (2) should have a negative sign and be statistically significant. Note also that, to avoid an explosive process, the coefficient should take a value between -1 and 0. According to the GRT, negative and statistically significant  $\alpha_2$  is a necessary condition for the variables in hand to be cointegrated. In practice, this is regarded as a convincing evidence and confirmation for the existence of cointegration found in the first step. It is also important to note that, in the second step of the EGM, there is no danger of estimating a spurious regression because of the stationarity of the variables ensured. Combinations of the two steps then provides a model incorporating both the static long-run and the dynamic short-run components.

To summarize the EGM, estimate Eq. (1) by the OLS and test for stationarity of the error terms in the first step. In the second step, if the null hypothesis of no cointegration is rejected, estimate Eq. (2) by replacing  $\beta$  by its previously computed OLS estimate  $\beta$  in the error-correction term ( $C_t - \beta Y_t$ ) or simply substituting the estimated residuals ( $u_t$ ) in place of ( $C_t - \beta Y_t$ ). In practice, most practitioners seem to prefer the latter one due to its simplicity. In the second step, all variables and the residuals are supposed to be stationary provided the model properly specified.

### B. The Engle-Yoo Three-Step Modeling Method (EYM)

Engle and Yoo (1991) propose a three-step estimation technique to overcome two of the main disadvantages of the classical two-step EGM. The two major shortcomings of the EGM are: i) although the long-run static regression gives consistent estimates, they may not be fully efficient, ii) due to nonnormality of the distribution of the estimators of the cointegrating vector, no sensible judgment can be made about the significance of the parameters.

The third step corrects the parameter estimates of the first step so that standard tests, such as t-test, can be applied (for further details, see Engle and Yoo (1991), Cuthbertson *et al.* (1992)). The three steps are then: first, estimate a standard cointegrating regression of the form Eq. (1), where  $u_t$  is the OLS residual to give first-step estimates of  $\beta$ , i.e.  $\beta^*$ . Then, estimate a second-step dynamic model of the form Eq. (2) using the lagged residuals from the cointegrating regression as an error-correction term. The third step, then consists of the regression

$$\varepsilon_t = \eta(-\alpha_2 Y_t) + v_t \quad (3)$$

The appropriate correction for the first-step estimates is, then, simply

$$\beta_{\text{cor}} = \beta^* + \eta \quad (4)$$

and the correct standard errors for  $\beta_{\text{cor}}$  are given by the standard errors for  $\eta$  in the third-step regression.

Engle and Yoo (1991) compare the EGM with the Johansen ML procedure. They emphasize that although the Johansen method has some advantages over the standard EGM, this can be reached at the cost of computational complexity. However, the three-step estimator achieves the same limiting distribution as the Johansen approach in an additional OLS regression from the two-step estimate.

### C. The Saikkonen Method

Banerjee *et al.* (1986) stress that ignoring the lagged terms in small samples is likely to create a bias in the estimated parameters. As noted earlier, in an attempt to estimate alternative cointegrating regressions, many have been interested in adding dynamic components (in the form of lags, leads or differences) to avoid the bias (for details, see Inder (1993), Phillips and Loretan (1991), Saikkonen (1991); also for the use of Autoregressive Distributed Lag (ADL) models in estimating the long-run relationships, see Charemza and Deadman (1992, 157-8)).

Among those, Saikkonen (1991) suggests a new asymptotically efficient estimator which is quite straightforward to compute using the OLS without any initial estimation. In practice, the proposed long-run estimator would take the following structure (note that the following is a simplified version of the Saikkonen method) as far as the regression (1) is concerned:

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 \Delta Y_{t-1} + \beta_3 \Delta Y_{t+1} + e_t \quad (5)$$

A time domain correction is reached by adding  $\Delta Y_{t-1}$  and  $\Delta Y_{t+1}$  to the classical Engle-Granger type static long-run regression of Eq. (1) where  $\Delta$  is the first-difference operator. In Saikkonen's words (Saikkonen (1991, p.15): "...The idea is essentially to remove the asymptotic inefficiency of the OLS estimator by using all the stationary information of the system to explain the short-run dynamics of the cointegration regression. Increasing the amount of such stationary information may reduce the relevant error covariance matrix of the cointegration regression and there by improve the asymptotic efficiency...".

D. The Johansen Maximum Likelihood (ML) Vector Autoregressive (VAR) Method

Due to the existence of VAR modeling within the Johansen method (Johansen (1988, 1991), Johansen and Juselius (1990)), the entire concept of cointegration becomes more complicated, not only conceptually but also computationally. Thus, here, we present a simplified version. Let us assume that the vector of variables  $Z$  has the following representation:

$$Z_t = \sum_{i=1}^m A_i Z_{t-i} + E_t \quad (6)$$

where  $Z_t$  contains all  $n$  variables of the model and  $E_t$  is a vector of random errors. This model can also be represented in the form of

$$\Delta Z_t = \sum_{i=1}^{m-1} \Gamma_i Z_{t-i} + \Pi Z_{t-m} + E_t \quad (7)$$

where

$$\Gamma_i = -I + A_1 + \dots + A_i \quad (I \text{ is a unit matrix})$$

$$\Pi = -(I - A_1 - \dots - A_m).$$

Let us now focus on matrix  $\Pi$ . Matrix  $\Pi$  can be represented in the following form:

$$\Pi = \alpha \cdot \beta$$

where  $\alpha$  and  $\beta$  are both  $n \times r$  matrices.

Matrix  $\beta$  is called the *cointegrating matrix* whereas matrix  $\alpha$  is referred to as the *adjustment matrix* or the *feedback matrix*. The Johansen method provides not only the direct estimates of the cointegrating vectors but also enables us to construct tests for the order (or rank) of cointegration,  $r$ . It is worth noting that, in a VAR model explaining  $N$  variables, there can be at most  $r = N-1$  cointegrating vectors. It is commonly acknowledged that the statistical properties of the Johansen procedure are generally better and the cointegration test is of higher power compared to the EG one. However, it is important to point out that they are grounded within different econometric methodologies and thus cannot be directly compared. In this regard, the Johansen method can be used for single-equation modeling as an auxiliary tool, testing the validity of the endogenous variable division. This may also be used as a confirmation test of the single-equation model. Following Charemza and Deadman (1992), we believe that single-equation-based and systems-based methods should be seen complementaries rather than substitutes. Let us assume that the Johansen results suggest the existence of unique cointegrating vector. Then, if the estimated cointegrating coefficients have economically sensible signs and are roughly similar in size to those estimated by, say, the EG method, this could be taken some confirmation of the single-equation model to which the EG method was applied (for more details, see Charemza and Deadman (1992, 201-2)).

Despite its theoretical advantages and superiority, the Johansen estimating procedure is, in practice, also subject to some shortcomings. First, given the small sample size, the method cannot be accepted as an appropriate one since the point estimates obtained for cointegrating vector,  $\beta$ , may not be particularly meaningful. Second, some additional problems occur if we do not have a unique cointegrating vector. The problem of multiple long-run relationship is presumably best seen as an identification problem (Granger (1986)), and can be resolved in, basically, two ways: either rejecting all but one such cointegrating vectors as economically meaningless or if the model is consistent with the underlying economic theory, it should consist of not one but two or more single equations. In this respect, Phillips and Loretan (1991) favor for the use of equation-by-equation approach of the single-equation error-correction model since such a possibility is not available in complete systems-methods such as the Johansen approach.<sup>4</sup>

---

<sup>4</sup> The multitude of available methods for estimation and inference in cointegrated systems is confusing and is still no overall agreement about the prescriptions for applied econometric research. However, many of these methods give particular emphasis on error-correction model representation, and such models currently represent the most common approach to modeling cointegrated variables.

#### IV. Some Concluding Remarks

Economic theory is mostly interested in equilibrium conditions and has little to say about the nature of economic configurations in disequilibrium. While economic theory proposes that certain macro variables have equilibrium relationships with each other, the data does not confirm that these hold at all times. To overcome this difficulty, economists make a distinction between the short-run and the long-run. The appeal of cointegration is that it provides a formal framework for testing long-run models from actual time-series data. The cointegration technique allows nonstationary data to be used so that spurious regression results are avoided. It also gives the chance to test the validity of an economic theory. If a postulated economic relationship exists, then the variables under consideration should be cointegrated. Testing for cointegration is, thus, a test for the existence of the equilibrium relationship postulated. In a sense, this is a test whether or not a model is well specified. Contrary to the popular belief, however, the concept of cointegration does not suggest any easy shortcuts in the construction and estimation of dynamic time-series models in economics (for various advantages and limitations with the employment of cointegration analysis in macroeconomic time-series modeling, see e.g. Muscatelli and Hum (1992), Perman (1991), Maddala (1992, 601-3)).

#### ÖZET

Bu makale uzun dönem ekonomik denge ilişkilerinin ekonometrik tahmin ve modellemesinde son yıllarda yoğun olarak başvurulan 'cointegration' yöntemini kullanan son yaklaşımları gözden geçirmekte ve değerlendirmektedir. Çalışmada ilgili literatürün genişliği nedeniyle kapsam açısından seçici olunmuştur.

#### REFERENCES

- BANERJEE, A., J. DOLADO, D.F. HENDRY and G. SMITH (1986) "Exploring Equilibrium Relationships in Econometrics Through Static Models", *Oxford Bulletin of Economics and Statistics*, 48, 253-77.
- BLANGIEWICZ, A. and W.W. CHAREMZA (1990) "Cointegration in Small Samples", *Oxford Bulletin of Economics and Statistics*, 52, 303-15.
- BLOUGH, S.R. (1988) "On the impossibility of Testing for Unit Roots and Cointegration in Finite Samples", *Working Paper*, No. 211, Department of Economics, John Hopkins University.
- CHAREMZA, W.W. and D.F. DEADMAN (1992), *New Directions in Econometric Practice*, Edward Elgar, England.



- CUTHBERTSON, K., S.G. HALL and M.P. TAYLOR (1992) *Applied Econometric Techniques*, Philip Allan, New York.
- ENGLE, R.F. and C.W.J. GRANGER (1987) "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-76.
- ENGLE, R.F. and B.S. YOO (1991) "Cointegrated Economic Time Series: An Overview with New Results" in R.F. Engle and C.W.J. Granger (eds.), *Long-run Economic Relationships: Readings in Cointegration*, Oxford University Press, New York.
- GRANGER, C.W.J. (1981) "Some Properties of Time Series Data and Their Use in Econometric Model Specification", *Journal of Econometrics*, 16, 121-30.
- GRANGER, C.W.J. (1986) "Developments in the Study of Cointegrated Economic Variables", *Oxford Bulletin of Economics and Statistics*, 48, (3), 213-28.
- INDER, B. (1993) "Estimating Long-run Relationships in Economics", *Journal of Econometrics*, 57, (1-3), 53-68.
- JOHANSEN, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-4.
- JOHANSEN, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, 55, 1551-80.
- JOHANSEN, S. and K. JUSELIUS (1990) "Maximum Likelihood Estimation and Inference on Cointegration: with Application to the Demand for Money", *Oxford Bulletin of Economics and Statistics*, 52, 169-210.
- MADDALA, G.S. (1992) *Introduction to Econometrics*, 2nd Ed., Macmillan.
- MUSCATELLI, V.A. and S. HURN (1992) "Cointegration and Dynamic Time Series Models", *Journal of Economic Surveys*, 6, 1-43.
- PARK, J.Y. and P.C.B. PHILLIPS (1988) "Statistical Inference in Regressions with Integrated Processes: Part I", *Econometric Theory*, 4, 468-97.
- PERMAN, R. (1991) "Cointegration: An Introduction to the Literature", *Journal of Economic Studies*, 18, (3), 3-30.

- PHILLIPS, P.C.B. and B.E. HANSEN (1990) "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, 57, 99-125.
- PHILLIPS, P.C.B. and M. LORETAN (1991) "Estimating Long-run Economic Equilibria", *Review of Economic Studies*, 58, 407-36.
- SAIKKONEN, P. (1991) "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometric Theory*, 7, 1-21.
- STOCK, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 56, 1035-56.
- UTKULU, U. (1994), "Cointegration Analysis: Introductory Survey with Applications to Turkey", in M. Güneş, Ş. Üçdoğruk and M.V. Pazarlıoğlu (eds.), *I. Ulusal Ekonometri ve İstatistik Sempozyumu Bildirileri, 11-12 Kasım 1993 (Papers at the I. National Symposium of Econometrics and Statistics)*, 303-24, İzmir.