COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1003

UNIDENTIFIED COMPONENTS IN REDUCED RANK REGRESSION ESTIMATION OF ECM'S

Peter C. B. Phillips

October 1991

UNIDENTIFIED COMPONENTS IN

REDUCED RANK REGRESSION ESTIMATION OF ECM's

by

Peter C. B. Phillips*

Yale University

0. ABSTRACT

Reduced rank regression procedures in error correction models (ECM's) permit consistent estimation of the cointegration space but do not provide consistent estimates of individual structural relations when the dimension of the cointegration space is greater than one. Indeed, individual structural cointegrating equations are unidentified without additional *a priori* restrictions, just as in the conventional simultaneous equations framework. The effect of this lack of identification is explored by considering the distributions and limit distributions of reduced rank regression estimates of unidentified components of the cointegrating matrix in a typical VAR formulation of the ECM. Some recommendations are made for empirical practice.

JEL Classification: 211

Keywords: Cauchy distribution; Identification; Rotation; Structural model; Unidentified component; Uniform distribution; VAR.

1. INTRODUCTION

Reduced rank regression methods are growing in popularity in applied econometric research. The statistical methodology of this approach is developed in Ahn-Reinsel (1988, 1990) and Johansen (1988, 1991). Empirical applications of the method with economic time series data for many different countries are now proliferating and these generally follow the spirit and detail of the applications reported in Johansen-Juselius (1989).

The object of the present paper is to make some simple but empirically important observations concerning the proper statistical interpretation of estimates from reduced rank regression methods. First, it seems not to be generally appreciated, although it was recently pointed out in a paper by Park (1990), that individual cointegrating relations are unidentified in error correction models of the type formulated by Johansen (1988) and used in reduced rank regression. What this means is that while the cointegration space is identified individual cointegrating relations generally are not. *A fortiori* individual coefficients in an estimated cointegrating matrix are also not identified. This lack of identification affects the statistical properties of the estimates that are obtained by reduced rank regression as well as hypothesis tests about the coefficients. This paper shows that the individual coefficient estimates in a reduced rank regression converge to random variables whose distributions are scaled versions of the Fisher (1915) product moment correlation coefficient distribution. Statistical tests about individual coefficients are generally inconsistent.

These results suggest that reduced rank regression estimates should be interpreted with care in empirical work. In cases where the dimension of the cointegration space (r) is found to be greater than one, we recommend:

(i) Individual eigenvectors from the reduced rank regression procedure should not be interpreted as estimates of individual structural equations. When $r > 1$ the structural equations are unidentified unless further identifying restrictions are imposed prior to estimation. The usual normalizing restrictions of reduced rank regression do not identify individual equations. This is in contrast to the restrictions involved in the triangular representation of cointegrated systems used by Phillips (1991), which do identify the individual structural equations *and* the individual coefficients in each equation.

(ii) Interpretations of individual coefficients from reduced rank regressions, no matter how tempting these are in practice, should be avoided. For example, the fact that certain components of an eigenvector from a reduced rank regression are close to zero should not be construed to mean that the corresponding coefficients in a structural equation are zero.

(iii) Ordering the fitted eigenvectors according to the size of the corresponding fitted eigenvalues in reduced rank regression does not provide a valid system of coordinates for interpreting empirical results when $r > 1$. Unfortunately, this is a procedure that is now being used frequently in empirical practice, partly because it seems natural and partly because of the way the regression output is presented. When $r > 1$, the ordering the of eigenvalues and hence the eigenvectors is random and remains so in the limit as the sample size $T \to \infty$. Use of the fitted eigenvalue ordering leads only to an empirical determination of a basis for the cointegrating space. Individual structural equations are still not identified and the coefficients obtained by this procedure are still random in the limit, thereby representing by their limiting variability the uncertainty that is inherent in their lack of identification.

(iv) Some hypotheses about the cointegrating space may be validly tested in reduced rank regression. For example, Johansen (1988, 1991) discusses the hypothesis that the cointegration is spanned by an explicit set of r cointegrating vectors. While certainly of interest, we recommend that such hypotheses also be treated with care in empirical work because although they isolate an explicit set of vectors that span the cointegration space they still do not identify individual structural relations. Arbitrary rotations of these vectors span the same space and are observationally equivalent. Only when all coefficients of a certain variable are zero is this configuration retained under rotation. This latter case is especially interesting because it corresponds to a special case where one variable has no causal effect on another variable or group of variables. This is a case studied explicitly in recent work by Toda and Phillips (1991a, b). It will be mentioned again later in the paper.

2. ISSUES OF IDENTIFICATION

Suppose the n-vector time series y_t satisfies an ECM of the form

(1)
$$J(L)\Delta y_t = \Gamma A' y_{t-1} + \varepsilon_t \quad (t = 1, \ldots, T);\ J(L) = \Sigma_0^{k-1} J_i L^i,\ J_0 = I$$
$$\varepsilon_t \equiv \text{iid } N(0, \Sigma_\varepsilon).$$

In (1) A' is an $r \times n$ matrix of cointegrating vectors and Γ $(n \times r)$ is a matrix of loading coefficients (essentially factor loadings) that are sometimes interpreted as adjustment coefficients in empirical work (c.f. Johansen and Juselius, 1990).

Note that neither Γ nor A are identified in (1). Indeed $\Gamma A' = \Gamma F^{-1}FA'$ for any nonsingular matrix F $(r \times r)$, as emphasized in the recent discussion by Park (1990). In empirical work based on the application of reduced rank regression methods a normalization condition is employed in the estimation of the matrix A. However, this normalization rule, as we discuss below, is not sufficient to identify A when $r > 1$. Indeed, under the usual normalization rule of reduced rank regression (see equation (6) below), A is determined only up to an arbitrary rotation of coordinates, i.e. up to multiplication on the right by an arbitrary $r \times r$ orthogonal matrix. This leads to a lack of identification in reduced rank regression and has important consequences for empirical practice.

An alternate way of writing the long-run component of (1) is

$$A'y_t = u_{1t} \tag{2}$$

where u_{1t} is a stationary process whose spectrum $f_{11}(\lambda)$ has the property that at the origin $f_{11}(0) > 0$, i.e. is positive definite. Interpreted as a system of r structural equations, (2) is identified only up to the space that is spanned by the columns of A, as pointed out by Park (1990). In the usual simultaneous equations framework we would require r restrictions per equation of (2), including the normalization in this count, in order to identify (2). System identification would then require r^2 restrictions. For instance, suppose we partition A' as $A' = [A_1, A_2]$ where A_1 is $r \times r$ with a conformable partition of y_t and write (2) as

$$A_1 y_{1t} + A_2 y_{2t} = u_{1t}, \; \Delta y_{2t} = u_{2t}. \tag{3}$$

Suppose we also require that $f_{uu}(\lambda)$, which is the spectrum of $u_t = (u_{1t}', u_{2t}')'$, has the property that $f_{uu}(0) > 0$. Then the rule just mentioned of r^2 restrictions *in toto* for the identification of A' applies directly to (3). In the model (3) the first r equations are in "structural form" and the final $n-r$ equations are in "reduced form." Note that the specification of (3), with the attendant restriction that $f_{uu}(0) > 0$, implicitly requires that $n-r$ variables of y_t, viz. y_{2t}, be isolated as a set of full rank integrated regressors. This condition as it appears in (3) is the analogue of the separation of the jointly dependent and predetermined variables in a structural simultaneous equations model together with the full rank condition on the observation matrix of the predetermined variables.

In the triangular system representation studied in Phillips (1991) the matrix A_1 in (3) is set to the identity, i.e. $A_1 = I$. With this formulation there are r^2 restrictions (r per equation of (3)) and (3) is identified.

In the reduced rank regression approach only an eigenvector normalization condition is employed in practice (see (6) below) and this condition involves only $r(r+1)/2$ restrictions, which are insufficient to identify either individual structural equations of (3) or individual elements of the cointegrating matrix A when $r > 1$. This lack of identifiability has important empirical consequences. In particular, it means that empirical estimates of the cointegration space that are obtained by reduced rank regression do not support a valid empirical determination of individual economic relationships when $r > 1$. Moreover, it is not possible, at least using this approach, to validly test hypotheses concerning individual elements of the matrix A. If such tests are conducted then they are inconsistent. Unfortunately, this point seems so far to have gone unrecognized in all of the empirical work that has been conducted with the reduced rank regression procedure. In practice when this method is employed, empirical eigenvectors are often presented and upon examination seen to have some elements that are close to zero, from which individual long-run relationships (such as purchasing power parity relations or money demand relations) are deduced as being supported by the data. As we will show in this note such statistical inference is generally invalid. Estimates of individual elements of the cointegrating matrix in unidentified models like (1) and (3) when $r > 1$ converge to random variables not constants and tests that are based on these individual estimates are inconsistent. As in the conventional simultaneous equations model, identifying prior information on structural relationships is required for valid inference about those relationships. In the absence of such prior information hypothesis testing must be confined to explicit hypotheses about the cointegration space itself, such as

$$\mathrm{H}_0 : A = A_0\Phi, \tag{4}$$

where A_0 is a known $n \times r$ matrix and Φ is some $r \times r$ matrix. Note that even in the case of hypotheses such as H_0 which are quite explicit about a set of vectors (viz. the columns of A_0) that span the cointegration space, the linear combinations (i.e. the columns of Φ in H_0) that do so are still not identified and are subject to arbitrary rotations.

3. CONSEQUENCES OF LACK OF IDENTIFICATION

The reduced rank regression estimator of A in (1) satisfies the optimization problem

$$\hat{A} = \mathrm{argmin}_A |S_{00} - S_{01}A(A'S_{11}A)^{-1}A'S_{10}| \tag{5}$$

(see Johansen (1988), p. 234) subject to a normalization condition on $\hat{A}$. The normalization rule employed by Johansen is:

(6) $$\hat{A}'S_{11}\hat{A} = I_r.$$

The S_{ij} in (2) are moment matrices of the residuals from the regression of Δy_t ($i = 0$) and y_{t-1} ($i = 1$) on the lagged differences Δy_{t-j} ($j = 1, ..., k-1$) that appear in (1). The notation is the same as Johansen's (1988) except that we use A' in place of his β' and the index 1 in place of his k (the latter since we use y_{t-1} rather than y_{t-k} in the ECM formulation (1)).

The normalization rule (6) imposes $r(r+1)/2$ restrictions on $\hat{A}$. Note that (6) remains invariant under the following arbitrary rotation of the rows of $\hat{A}'$

(7) $$\hat{A}' \rightarrow H\hat{A}'$$

where H is any matrix in the orthogonal group $O(r)$ of $r \times r$ matrices. Since

$$S_{00} - S_{01}\hat{A}\hat{A}'S_{10} = S_{00} - S_{01}\hat{A}H'H\hat{A}'S_{10}$$

it is clear that the criterion function of (5) is also invariant under the rotation (7). It follows directly that $\hat{A}$ and $H\hat{A}$ cannot be empirically distinguished and indeed these matrices have identical distributions for all sample sizes T. Moreover, the distributional equivalence of $\hat{A}$ and $H\hat{A}$, which we write as

(8) $$\hat{A}' =_d H\hat{A}', \text{ for all } H \in O(r),$$

applies in the limit as $T \rightarrow \infty$. Taking a partition of $A' = [a_1, ..., a_n]$ into individual columns we see that

$$\hat{a}_i =_d H\hat{a}_i.$$

In conventional terminology (cf. Muirhead, 1982, p. 32) the distribution of $\hat{a}_i$ is spherically symmetric. Suppose we now partition $\hat{a}_i$ as $\hat{a}'_i = (\hat{a}_{i1}, \hat{a}'_{i2})$ and consider the ratio $\hat{b} = \hat{a}_{i2}/\hat{a}_{i1}$. Note first that if $X =_d N(0, I_r)$ is an arbitrary r-vector of independent standard normal variates partitioned as $X' = (X_1, X_2')$ conformably with $\hat{a}_i$ then $\hat{a}_i/(\hat{a}'_i/\hat{a}_i)^{1/2} =_d X/(X'X)^{1/2}$. Next observe that

$$\hat{b} = \frac{\hat{a}_{i2}}{\hat{a}_{i1}} = \frac{\hat{a}_{i2}/(\hat{a}_i'\hat{a}_i)^{1/2}}{\hat{a}_{i1}/(\hat{a}_i'\hat{a}_i)^{1/2}} =_d \frac{X_2/(X'X)^{1/2}}{X_1/(X'X)^{1/2}} = \frac{X_2}{X_1},$$

which is multivariate Cauchy (of dimension $r-1$) for all sample sizes. Thus $\hat{b}$ is multivariate Cauchy in the limit as $T \rightarrow \infty$.

Clearly the effect of lack of identifiability of the cointegrating matrix A in (1) is that elements of A cannot be consistently estimated. Functions like $\hat{b}$ of the elements of $\hat{A}$ converge to random variables in the limit even when the normalization rule (6) is applied. The same is true of the individual elements of $\hat{A}$, which we now consider.

Let $\hat{a}_i = \hat{h}(\hat{a}'_i\hat{a}_i)^{1/2} = \hat{h}\hat{s}^{1/2}$, say. Under the normalization rule (6) $\hat{s} = \hat{a}'_i\hat{a}_i$ is invariant and as $T \rightarrow \infty$ $\hat{s}$ tends in probability to a well defined finite limit value s, say. The vector $\hat{h}$ lies on the sphere $\hat{h}'\hat{h} = 1$ and is spherically distributed for all T Thus, $\hat{h}$ is uniformly distributed on the sphere $S_r = \{h \in \mathbf{R}^r : h'h = 1\}$, see e.g. Phillips (1985). The limiting distribution of $\hat{a}_i$ is therefore given by that of the vector $hs^{1/2}$ where h is uniform on the sphere S_r and s is a constant.

The distribution of an individual element, say $\hat{a} = \hat{a}_{ij}$, of $\hat{a}_i$ is therefore given in the limit by the variate

$$a = h_j s^{1/2}$$

where h_j is the j'th element of the uniform vector variate h. As shown in the lemma in the Appendix, the density of h_j is given by

$$\mathrm{pdf}(h_j) = [B(1/2), (r-1)/2]^{-1}\left(1 - h_j^2\right)^{(r-3)/2}, \quad -1 \le h_j \le 1,$$

where $B(\cdot,\cdot)$ is the beta function, and this is the density of the product moment correlation coefficient (when the population correlation $\rho = 0$) discovered by R. A. Fisher (1915). The density of a is therefore

$$\mathrm{pdf}(a) = [B(1/2), (r-1)/2]^{-1} s^{-(r-4)/2}\left(s - a^2\right)^{(r-3)/2}, \; -s^{1/2} \le a \le s^{1/2}.$$

For $r = 1$ the density is zero and the distribution is discrete with mass points $\pm s^{1/2}$ (i.e., $h_j = \pm 1$) each of which is assigned a probability of one half. For $r = 2$ the density is U shaped with infinite ordinates at the points $a = \pm s^{1/2}$. For $r = 3$ the distribution is uniform over the interval $[-s^{1/2}, s^{1/2}]$. For $r \ge 4$ the distribution is symmetric and unimodal with mode at $a = 0$.

Since $\hat{a} \rightarrow_d a$ and the limit variate is not degenerate when $r > 1$, any attempt to test hypotheses about individual coefficients in the matrix $\hat{A}$ will lead to inconsistent tests.

4. CONCLUDING REMARKS

Reduced rank regression methods produce consistent estimates of the dimension of the cointegration space and the space itself. However, without further identifying information this class of methods cannot be used to validly estimate individual structural relations or to test hypotheses about the coefficients in such relations when the dimension of the cointegration space $r \ge 2$. If these methods are used for such purposes then the resulting estimates converge weakly to random variables not constants and statistical tests are inconsistent.

Models constructed using a triangular representation as in (3) with $A_1 = I$ are identified. Individual structural equations may be consistently estimated in such systems and valid hypothesis testing may be conducted using conventional chi-squared criteria (see Phillips (1991) for details).

In reduced rank regression models like (1) the only hypotheses that may be validly tested are those that remain invariant under arbitrary coordinate rotations. One such class of hypothesis is given by (4) wherein an explicit basis for the cointegration space is hypothesized. In this case, of course, we have

$$HA' = H\Phi' A_0', \text{ for arbitrary } H \in O(r)$$

and the hypothesis (that the rows of A_0' span the cointegration space) still applies. Another class of hypothesis that is invariant under rotation concerns the length of the columns of $A' = [a_1, ..., a_n]$. In particular the hypothesis

$$\mathrm{H}_0 : a_k = 0, \text{ for some particular } k,$$

is equivalent to

$$\mathrm{H}_0 : a_k' a_k = 0$$

and this hypothesis is invariant under rotation of A'. This hypothesis is especially important because it arises in a natural way in testing the noncausal effects of one variable in the system (here y_k) on other variables in the system. Such hypotheses and tests of them are systematically explored in Toda and Phillips (1991a, b). Even in this case, however, valid test procedures based on the error correction model (1) are not straightforward. This is because the noncausality null hypothesis involves not H_0 but rather the nonlinear hypothesis

$$\mathrm{H}'_0 : \gamma_j' a_k = 0, \text{ for some } j \text{ and } k,\ j \neq k$$

where $\Gamma' = [\gamma_1, ..., \gamma_n]$. Clearly, H_0 is sufficient but not necessary for H_0'. As shown in Toda and Phillips (1991a,b), attempts to test H'_0 directly are generally frustrated by parameter dependencies and nonstandard distributions in the limit theory. These complications would seem to make the error correction model (1) and reduced rank regression method rather less useful for empirical work in practice than has generally been recognized.

Finally, we remark that it is possible to impose restrictions like those in (3) with $A_1 = I$ *ex post*. Suppose for example that (1) is estimated by reduced rank regression. Then the matrix $B = -A_1^{-1}A_2$ may be estimated by $\hat{B} = -\hat{A}_1^{-1}\hat{A}_2$ where $\hat{A}' = [\hat{A}_1, \hat{A}_2]$ is the reduced rank regression estimator of A. Note that B is invariant to coordinate rotations in the cointegrating space. This estimate $\hat{B}$ is entirely analogous to the maximum likelihood estimate $\tilde{B}$ of B in the triangular system representation studied in Phillips (1991). However the finite sample properties of these estimates are very different. In fact, $\hat{B}$ has no finite sample moments of integer order, whereas $\tilde{B}$ has moments to order $T-n+r$. This explains, in part, the outlier behavior of the reduced rank regression estimator $\hat{B}$ observed in some earlier studies, e.g. by Park and Ogaki (1991). The reader is referred to Phillips (1991) for further details on this point.

REFERENCES

Ahn, S. K. and G. C. Reinsel (1988). "Nested reduced rank autoregressive models for multiple time series," *Journal of the American Statistical Association*, 83, 849-856.

Ahn, S. K. and G. C. Reinsel (1990). "Estimation for partially nonstationary multivariate autoregressive models," *Journal of the American Statistical Association*, 85, 813-823.

Fisher, R. A. (1915). "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, 10, p. 507.

Johansen, S. (1988). "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," University of Copenhagen, to appear in *Econometrica*, 1991.

Johansen, S. and K. Juselius (1989). "Maximum likelihood estimation and inference on cointegration -- with applications to the demand for money," *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.

Park, J. Y. (1990). "Maximum likelihood estimation of simultaneous cointegrated models," mimeographed, Cornell University.

Park, J. Y. (1991). "Canonical cointegrating regressions," *Econometrica* (forthcoming).

Park, J. Y. and M. Ogaki (1991). "Inference in cointegrated models using VAR prewhitening to estimate short run dynamics," Rochester Center for Economic Research Working Paper No. 281.

Phillips, P. C. B. (1986). "The distribution of FIML in the leading case," *International Economic Review*, 27, 239-243.

Phillips, P. C. B. (1989). "Spherical matrix distributions and Cauchy quotients," *Statistics and Probability Letters*, 8, 51-53.

Phillips, P. C. B. (1991a). "Optimal inference in cointegrated systems," *Econometrica*, 59, 283-306.

Phillips, P. C. B. (1991b). "The tail behavior of maximum likelihood estimators of cointegrating coefficients in error correction models," Cowles Foundation Discussion Paper No. xxx.

Toda, H. and P. C. B. Phillips (1991a). "Vector autoregressions and causality," Cowles Foundation Discussion Paper No. 977.

Toda, H. and P. C. B. Phillips (1991a). "Vector autoregression and causality: A theoretical overview and simulation study," mimeographed, Yale University.

TECHNICAL APPENDIX

LEMMA. *Let h be a random m-vector that is uniformly distributed on the sphere $S_m = \{h \in \mathbb{R}^m : h'h = 1\}$. Let h_j be an arbitrary element of h. Then the density of h_j is given by*

$$\mathrm{pdf}(h_j) = B(1/2,\ (m-1)/2)^{-1}\left(1 - h_j^2\right)^{(m-3)/2}$$

where $B(\cdot,\cdot)$ is the beta function.

PROOF. Without loss of generality choose $j = 1$. Since $h =_d X(X'X)^{-1/2}$ where $X =_d N(0, I_m)$ we may work with the random variate

$$Z = X_1(X'X)^{-1/2} = X_1\left(X_1^2 + X_2'X_2\right)^{-1/2} = t\left(t^2 + 1\right)^{-1/2}$$

where $t = X_1(X_2'X_2)^{-1/2}$ and $X = (X_1, X_2')'$ partitions X into the scalar X_1 and $(m-1)$-vector X_2. Observe that

$$t =_d (m-1)^{-1/2}t_{m-1}$$

where t_{m-1} is a t-variate with $m-1$ degrees of freedom. The density of t is

$$\mathrm{pdf}(t) = B(1/2,\ (m-1)/2)^{-1}\left(1 + t^2\right)^{-m/2}.$$

Noting that $t = z(1 - z^2)^{-1/2}$ and changing variables to Z we find that

$$\mathrm{pdf}(z) = B(1/2,\ (m-1)/2)^{-1}\left(1 - Z^2\right)^{(m-3)/2},$$

giving the stated result. □

REMARK. The density obtained in the lemma above is the canonical form of the density of the product moment correlation coefficient. To see this, let X and Y be independent $N(0, I_m)$ vectors. Define $r = X'Y/(X'X)^{1/2}(Y'Y)^{1/2}$. Observe that $k = X(X'X)^{1/2}$ and $h = Y(Y'Y)^{1/2}$ are independent and uniform on the sphere S_m. Now condition on k and construct the orthogonal matrix $C = [k, k_\perp] \in O(m)$. Note that $\bar{h} = C'h$ is also uniform on the sphere (by the invariance of this distribution under rotations) and is independent of k. Thus $\bar{h}$ is unconditionally uniform on S_m. Next write

$$r = k'h = k'CC'h = \bar{h}_1$$

so that r is distributed as the first element of $\bar{h}$. Thus, the density of r is the same as the density of h_j as given in the lemma. This provides an alternate derivation of the exact distribution of the product moment correlation coefficient for independent normal populations, originally given by Fisher (1915).