ELSEVIER Journal of Econometrics 87 (1998) 87–113

JOURNAL OF Econometrics

Structural relations, cointegration and identification: some simple results and their application

James Davidson[a,*]

[a] *Cardiff Business School, Colum Drive, Cardiff CF1 3EU, UK*

Received 1 June 1995; received in revised form 1 October 1997

Abstract

This paper presents and applies some results on the interpretation of cointegrating regressions. The key concept is the irreducible cointegrating (IC) relation, one from which no variable can be omitted without loss of the cointegration property. Extending earlier results, it is shown that under certain circumstances, IC relations are identified structural forms. It is possible, at least in principle, to learn about the structure of simultaneous long-run relations directly from cointegration analyses, in contrast with the well-known fact that no such knowledge can be obtained from the correlations between stationary variables. IC relations can also be estimated by asymptotically mixed Gaussian and median unbiased estimators, permitting standard inference. MINIMAL, an algorithm for extracting the IC subsets of a data set, is applied to variety of artificial and actual data.

JEL classification: C32

Keywords: Structural; Identification; Cointegration; Irreducible cointegrating vector

1. Introduction

Considerable interest has recently been shown in the problem of identifying the long-run relationships in a linear cointegrating model with $I(1)$ variables; see for example Pesaran and Shin (1994), Johansen (1995), Boswijk (1995), and

* Corresponding author. E-mail: davidsonje@cardiff.ac.uk

PII S0304-4076(98)00007-4

Hsiao (1997). These papers focus on the problem of imposing and testing restrictions on the space of the cointegrating vectors in the context of a vector error correction model (VECM), so as to determine long-run behavioural parameters such as supply and demand elasticities, propensities to consume or save, etc. We will call such parameters, and the relations containing them, *structural*. Following on from a recent note (Davidson, 1994) this paper addresses the structural identification issue from a slightly different perspective compared to the above-mentioned work. One feature of our analysis is that it provides an interpretation of single cointegrating regressions of the familiar Engle–Granger (1987) type, although as we subsequently show, it is relevant to the system estimation problem too.

There is a risk of confusion in the use of the word structure, because of the many different uses to which it has been put by different authors. We note for example that Hendry (1995) invokes the properties of constancy across interventions and regime shifts in his definition of structural, and these are implicit here, but our analysis does not consider such shifts explicitly. We use the term to mean, simply, parameters/relations which have a direct economic interpretation, and may therefore satisfy restrictions based on economic theory. In the context of simultaneous equations, our usage is the same as in the familiar Cowles Commission analysis (see e.g. Johnston, 1984, Chapter 11). This warns us, for example, that a regression involving jointly determined variables typically estimates a hybrid of different structural relations; for example, the regression of price on quantity and other variables will in general estimate a mixture of the demand and supply schedules. However, a different estimation method (e.g. instrumental variables) can yield consistent estimates of a structural equation provided the well-known conditions for identification are satisfied. If we focus on the 'cointegrating structure' (cointegrating relations between $I(1)$ variables) and ignore the short-run dynamics of the system (involving differences of the $I(1)$ variables and other $I(0)$ variables), the identification problem is formally similar to the Cowles Commission case.[1] The main differences are, first, that the distinction between endogenous and exogenous variables disappears and normalisations are arbitrary; and second, as we show, the characterisation of consistent estimators is notably different in the $I(1)$ case.

Section 2 of the paper gives a number of simple, non-technical results on the interpretation of cointegrating regressions. The basic idea is to show that it may be possible to deduce features of the long-run structure direct from the data analysis. Thus, we show that least squares consistently estimates an identified

[1] As Johansen (1995) argues, the identification of the dynamic components may be considered separately. Ignoring these is appropriate to single-equation cointegration analysis, but even in the context of system estimation, a priori restrictions on the dynamics which might aid identification of the cointegrating structure are rare, and we lose little by neglecting this possibility.

structural equation, but also that in certain favourable circumstances, it is possible to establish the status of the equation from the data alone. The key concept, developed in Section 2.1, is the *irreducible cointegrating* (IC) relation; that is to say, a cointegrating relation, which ceases to be so if any of the variables it contains are suppressed. Section 2.2 gives some additional results concerning IC relations, and Section 2.3 illustrates the results by means of a simple example. A closely related question concerns the feasibility of efficient estimation and standard inference. The standard errors and t-ratios in an OLS cointegrating regression do not generally have the conventional interpretation, but procedures such as the fully modified least squares estimator of Phillips and Hansen (1990) are available. Under certain circumstances these are asymptotically mixed normal, median unbiased to order T^{-1}, and allow standard inference. Section 2.4 gives a theorem showing that the Phillips–Hansen type of estimator is always appropriate for the estimation of IC equations.

Section 3 of the paper considers the problem of detecting and estimating the IC vectors for a data set. Section 3.1 describes MINIMAL, an algorithm implemented in the GAUSS language which tests sequentially for the IC property in every subset of variables, employing a Wald test of restrictions on the Johansen (1988, 1991) estimator of the cointegrating space. Section 3.2 summarises some simulation findings on MINIMAL. Section 3.3 reports applications of the method to two economic data sets, a collection of quarterly series of UK interest rates and related variables, and the data on the US economy used in the study of King et al. (1991). Section 4 concludes the paper.

2. Structure and cointegration

2.1. The background

The starting point for the analysis is the pair of results proved in Davidson (1994). However, to provide the necessary background, we first briefly review the textbook theory of the identification of simultaneous equations. Suppose

$$\beta' x_t = z_t \quad (s \times 1) \tag{2.1}$$

is a vector of equations connecting a vector of variables $x_t\,(m \times 1)$ where β is a $m \times s$ matrix of coefficients. In the usual Cowles Commission type of simultaneous equations model, $z_t \sim \text{iid}(0, \Sigma)$. At most s of the variables in x_t are thought of as being determined by this system, the remaining $m - s$ being given exogenously. If we know nothing about the correlation of the disturbances, the system

$$d\beta' x_t = dz_t \tag{2.2}$$

where $d(s \times s)$ is any non-singular matrix, is observationally equivalent to the original one. For consistent estimation, enough restrictions on the columns of β must be known to rule out all possibilities except $d = I_s$.[2] Suppose we are interested in estimating the first column of β, say β_1, normalised on the first element, given the prior information that the last $m - g_1$ elements are zero; in other words,

$$\beta_1 = (1, \beta_{21}, \dots, \beta_{g_1 1}, 0, \dots, 0)'. \tag{2.3}$$

Let β be row-partitioned conformably with β_1 as

$$\beta = \begin{bmatrix} \beta_a \\ \beta_b \end{bmatrix} = \begin{bmatrix} \beta_{1a} & \beta_{2a} \\ 0 & \beta_{2b} \end{bmatrix} \begin{matrix} (g_1 \text{ rows}) \\ (m - g_1 \text{ rows}) \end{matrix} \tag{2.4}$$

where β_{2b} has dimension $(m - g_1) \times (s - 1)$. If and only if $\mathrm{rank}(\beta_{2b}) = s - 1$, the only choice of r $(s \times 1)$ such that the vector βr satisfies the known prior restrictions on β_1 is $r = e_1 = (1, 0, \dots, 0)'$. This is the well-known rank condition for identification, and in the usual analysis, it is normally a necessary condition for consistent estimation of the equation by instrumental variable or maximum likelihood methods.[3]

Now consider the cointegrating VAR, the system of reduced-rank dynamic equations which has been analysed by Johansen (1988, 1991) *inter alia*:

$$A(L)x_t = \alpha\beta' x_t + A^*(L)\Delta x_t = u_t \quad (m \times 1), \tag{2.5}$$

where L is the lag operator, $A(L) = \alpha\beta' + A^*(L)(1 - L)$ such that $A(1) = \alpha\beta'$, and α and β are $m \times s$ matrices, the loadings matrix and the matrix of cointegrating vectors, respectively.[4] When $s < m$ it can be shown that $x_t \sim I(1)$, and the system embodies a set of long-run equilibrium relations of the form (2.1) where

$$z_t = (\alpha'\alpha)^{-1}\alpha'(u_t - A^*(L)\Delta x_t) \sim I(0). \tag{2.6}$$

---

[2] For simplicity we confine attention to the case of zero restrictions on the coefficients.

[3] Johansen (1995) points out that the test of this condition is not strictly operational because elements of β are unknown a priori, and presents a version of the rank test in terms solely of the restrictions being imposed. Note however that the usual test yields a valid result for all possible values of β except for a set of Lebesgue measure zero.

[4] It is customary to write this representation with the maximum order of lag on the 'levels' term, which merely redefines $A^*(L)$. Eq. (2.5) is more convenient for the present analysis. The representation also assumes that the variables have zero mean, and that there are no deterministic trends. The elaborations required to relax these assumptions are straightforward and do not alter the essentials of the problem. For simplicity of exposition, they are omitted here.

In this model, there are s distinct linearly independent cointegrating vectors, the columns of β. As in the static model, without restrictions on β the long run of the system is observationally equivalent to Eq. (2.2) with loadings matrix αd^{-1}. Johansen's methodology does not attempt to solve this problem and estimates a collection of orthonormalised vectors spanning the same space as β.

However, the usual considerations apply in identifying the cointegrating vectors. The simplest result is the following, which, although it directly relates to single equation estimation, has wider implications.

Theorem 1 (Davidson, 1994). *If a column of β (say, β_1) is identified by the rank condition, the OLS regression which includes just the variables having unrestricted (non-zero) coefficients in β_1 is consistent for β_1.*

Note that if another variable is added to this cointegrating regression, its coefficient does not converge to zero and the other coefficients to the same limits as before, as we would expect of an 'irrelevant' variable in a stationary regression. The regression coefficients will generally converge to some other element of the cointegrating space. The situation is loosely analogous to what happens if an instrument whose exclusion from a simultaneous structural equation is helping to identify it is added to the equation. The resulting IV estimator is of course inconsistent. Theorem 1 is given in terms of OLS, but it holds in respect of any estimator that has been shown consistent for β_1 when OLS is consistent, such as 2-stage least squares, or the Phillips–Hansen (1990) fully modified estimator. We say more about this in Section 2.4.

If a collection of $I(1)$ variables is found to be cointegrated, it does not follow that the estimated vector can be interpreted as structural. However, the following concept is relevant in this connection.

Definition 1. A set of $I(1)$ variables will be called irreducibly cointegrated (IC) if they are cointegrated, but dropping any of the variables leaves a set that is not cointegrated.

We will also speak of an IC relation, or IC vector, to refer to the corresponding cointegrating coefficients for these variables. We immediately note the following important property of these vectors.

Theorem 2. An IC vector is unique, up to the choice of normalisation.

Proof. This is by contradiction. Suppose there exists for the IC variables a set of cointegrating vectors of rank at least two. Any linear combination of these vectors lies in the cointegrating space, and by choosing the weights appropriately, we can always generate such a combination having a zero element. Since the variable in question can therefore be dropped from the set without losing cointegration, the IC assumption is contradicted. □

The following is the result that explains our interest in IC relations.

Theorem 3 (Davidson, 1994). *If and only if a structural cointegrating relation is identified by the rank condition, it is irreducible.* □

This tells us that at least some IC vectors are structural. Theorem 3 does not eliminate all ambiguity in the interpretation of a particular cointegrating regression, but it has a number of useful implications. In fact, it is easy to see on reflection that an IC vector must have a certain status with regard to the underlying cointegrating structure. When the cointegrating rank of the system is s, an IC relation can contain at most $m-s+1$ variables. There are between s and $\binom{m}{m-s+1}$ of these vectors in total, the actual number depending on the degrees of overidentification of the relations of the system. But in addition to up to s identified structural relations, which are among the IC vectors by Theorem 3, there are also generally a number of *solved relations* (or equivalently, solved vectors).

Definition 2. A solved vector is a linear combination of structural vectors from which one or more common variables are eliminated by choice of offsetting weights such that the included variables are not a superset of any of the component relations.

A solved vector lies in the cointegrating space by construction, and if its components are identified structural vectors, it may also be irreducible.[5] The largest number of solved relations arises with a just-identified system. There are none at all under maximal overidentification, such that each variable appears in one and only one structural relation. In the sense of being solved from the structure, these relations are comparable to the reduced form equations of the conventional simultaneous equations analysis, and may indeed coincide with them (see the example in Section 2.3). However, the reduced forms are defined with respect to a particular normalisation, based on the endogenous–exogenous distinction, which is irrelevant here. Moreover, the reduced forms with respect to a given normalisation all contain $m-s+1$ variables by definition. They need not be irreducible even if the model is identified.

2.2. *Implications of irreducible cointegration*

A common research methodology is to construct a putative cointegrating regression in the light of some economic theory, the theory being deemed to

---

[5] Indeed, it seems to be necessarily irreducible, but we do not need to prove this, simply to note that IC vectors can be of this form.

receive support if the hypothesis of non-cointegration is rejected in this equation. However, according to Theorem 3, a cointegrating vector that contains redundant elements can be of no interest to us. The theory could be wrong, in which case this is just an arbitrary element of the cointegrating space. If the theory is correct, the relation is revealed to be underidentified, and the estimate is inconsistent, representing a hybrid of different structural equations, just as in the conventional analysis of simultaneity. Irreducibility is therefore an important diagnostic property of a cointegrating regression. Although simple considerations of parsimony would often lead to IC vectors being reported in practice, no one, to the author's knowledge, has previously considered the implications of testing for irreducibility.

If an IC relation is found, interest focuses on the problem of distinguishing between structural and solved forms. Of course, the theoretical model might answer this question for us, as the example below will illustrate. However, important clues may also be provided by the data themselves. The following lemma has inherent interest, and is also useful in proving our first main result. (All proofs for this section are given in Appendix A).

Lemma 1. Provided β is restricted only by zero and normalisation restrictions, a solved IC relation contains at least as many variables as each of the identified structural relations from which it derives.

An example of an additional restriction on β which would invalidate the result of Lemma 1 is where two rows of the matrix are the same, such that two variables take the same coefficients in every cointegrating relation. Such restrictions are clearly not impossible since the economics of the model might imply them. However, we can rule them out from arising "by chance" in the same way as we are willing to rule out a failure of identification from the same cause. (See footnote 3.)

In general, therefore, the fewer variables an IC relation contains, and the fewer it shares with other IC relations, the better the chance that it is structural and not a solved form. In the extreme cases, we can actually draw definite conclusions, as the following pair of results show.

Theorem 4. If an IC relation contains strictly fewer variables than all those others having variables in common with it then, subject to the condition of Lemma 1, *it is an overidentified structural relation.*

Theorem 5. If an IC relation contains a variable which appears in no other IC relation, it is structural.

Thus, it is possible, in the context of simultaneous cointegrating relations, to discover structural economic relationships directly from a data analysis, without the use of any theory. To take a very simple example, suppose our system (assumed complete) consists of four $I(1)$ variables, x, y, z and w. If the pairs (x, y) and (z, w) are found to be cointegrated (but not the pairs (x, z) or (y, w)) these two cointegrating relations, necessarily irreducible of course, are also necessarily structural. Neither can have arisen as a result of solving out some more fundamental relationships. This is a case of maximal overidentification. If, on the other hand, the pairs (x, y) and (x, z) are cointegrated, it follows necessarily that the pair (y, z) is also cointegrated. The cointegrating rank of these three variables is 2, and one of these three IC relations necessarily exists by being solved from the other two; but it is not possible to know which, without a prior theory.[6]

It needs to be emphasised that Theorems 4 and 5 are no more than possibility theorems. There is no guarantee that the conditions will be satisfied in any particular case. Nonetheless, the mere possibility of a 'free lunch', of gaining some direct knowledge of the cointegrating structure without any prior theory, may be found surprising and counter-intuitive, given our strong preconceptions about structural modelling derived from the stationary data case. The fundamental implication of the present results is that models with stochastic trends are different in this respect.

2.3. An example

The results can be illustrated in more depth with reference to the standard market model discussed in Davidson (1994), that we write here as

$$\begin{aligned} &q_t - \beta_{21}p_t - \beta_{31}w_t - \beta_{41}r_t \sim I(0) \quad \text{(demand)},\\ &q_t - \beta_{22}p_t - \beta_{32}w_t - \beta_{42}r_t \sim I(0) \quad \text{(supply)},\\ &w_t \sim I(0),\\ &r_t \sim I(0), \end{aligned} \tag{2.7}$$

where p_t denotes price, q_t quantity, and w_t and r_t are autonomous variables.[7] Both schedules are cointegrating but w_t and r_t are *not* cointegrated with each

[6] The empirical example of US output, consumption and investment discussed in Section 3.3 below is a case in point. The consumption and investment functions are commonly thought of as structural, directly reflecting the economic activities generating the data. They jointly imply a cointegrating relationship between consumption and investment. Of course, a theory which says the consumption/investment relation is 'structural' cannot be contradicted by the data. The point we make here is that in the maximally overidentified case (pairwise cointegration) no such ambiguity exists. The structure is *explicit* in the data.

[7] The notation '$\sim I(0)$' indicates the possible existence of short-run dynamics in these relationships which, however, are explicitly excluded from consideration. It is only the cointegrating structure which concerns us here.

other, otherwise we would have to augment the system with this extra relation. Here, $m=4$ and $s=2$. If there are no restrictions on the system (so that it is underidentified by the usual criteria) there are four IC relations each involving three variables, obtained by dropping the variables p_t, q_t, w_t, r_t from the set one at a time. Of course, none of these corresponds to the structural relations. If

$$\beta_{41}=\beta_{32}=0$$

so that both schedules satisfy the rank condition, the demand schedule is consistently estimated by regressing q_t on p_t and w_t, and the supply schedule is consistently estimated by regressing q_t on p_t and r_t. There are two other IC relations, what we usually call the reduced forms, containing respectively (q_t, w_t, r_t) and (p_t, w_t, r_t). To distinguish the reduced (i.e. solved) forms from the structural relations we can use our economic knowledge that only the demand and supply schedules contain both p_t and q_t. We do not even need to know in advance which schedule contains w_t and which r_t, since the theoretical signs of the slopes, $\beta_{21}<0$ and $\beta_{22}>0$, are sufficient prior knowledge to distinguish them.[8]

If the restrictions are $\beta_{31}=\beta_{41}=0$, the supply relation is underidentified and the demand relation is overidentified. In this case p_t and q_t are cointegrated so there are three IC relations, of which one is uniquely smallest. We thereby know from Theorem 4 (without *any* economic knowledge being required) that this is an identified structural relation. There are also the same two solved forms as before. Of course, the supply relation is not IC.

If the restrictions are $\beta_{22}=\beta_{41}=\beta_{32}=0$, so that supply is determined exogenously, the whole system is again identified, but q_t and r_t are a cointegrated pair. There are again only three IC relations, of which only one is a solved form. The supply relation is known to be structural, independent of theory, by Theorem 4 and similarly the demand relation, being the only relation containing p_t is known to be structural by Theorem 5. We could actually learn about the existence of the supply and demand functions empirically, without having any economic theory to structure the investigation! A similar argument would apply in the c ase of inelastic demand.[9]

[8] Of course, we need to know that the system *is* identified to make this interpretation, and a situation in which we know that each schedule contains at most one autonomous variable, but not which one, is arguably unlikely! Nonetheless the point of principle remains valid, that such information could not suffice to reveal the correct specification in the stationary data case, but could do so here.

[9] Note that this analysis assumes, crucially, that all relevant variables are included in the data set being analysed. If any are omitted the application of the results could lead to incorrect conclusions. This case is explored more fully in the working version of the paper, see Davidson (1996).

2.4. Estimation of IC vectors

We showed in Theorem 2 that if a set of variables have the IC property, the IC vector is unique. This fact is of some practical importance, because it means that single equation estimators are available for IC vectors which are asymptotically mixed Gaussian and median unbiased to $O(T^{-1})$, so that standard tests of hypotheses are available asymptotically. This is true whether or not the vectors are structural, but it obviously has special relevance in the structural case, because it implies that identification by the rank condition is a sufficient condition for these properties to hold. This result follows from the fact that any IC vector can be embedded in a closed dynamic system for the relevant subset of included variables, solved out of Eq. (2.5). By construction, this system has cointegrating rank 1, and after a suitable transformation, it can be rearranged into the triangular form assumed in Phillips (1991) and Phillips and Hansen (1990).

These authors work with systems having the form (apart from possible deterministic trends that we omit here)

$$x_{1t} = \gamma' x_{2t} + u_{1t},$$
$$\Delta x_{2t} = u_{2t}, \tag{2.8}$$

where x_{1t} is $s \times 1$, x_{2t} is $(m-s)\times 1$, γ is $(m-s)\times s$, and u_{1t}, u_{2t} are vectors of stationary processes. This structure implies the partition $\beta' = [I_s : -\gamma']$, such that the regressions of the elements of x_{1t} onto x_{2t} yield consistent estimates of the structural coefficients. The marginal submodel for x_{2t} necessarily carries the full complement of $m-s$ unit roots. The 'reduced form' structure of the conditional submodel in Eq. (2.8) is highly restrictive for cases with $s > 1$, but if $s = 1$, every $\beta(m \times 1)$ satisfies the requirement after merely being normalised on its first element. The following theorem shows that a solved model having this structure can always be invoked for an IC vector.

Theorem 6. Let $x_t = (x'_{at}, x'_{bt})'$ define a partition of the variables into subsets of dimension g_1 and *$m - g_1$ respectively. If x_{at} is an IC subset, the closed dynamic subsystem for x_{at} solved out from* Eq. (2.5), *having the form*

$$B(L)x_{at} = v_t \sim I(0) \quad (g_1 \times 1) \tag{2.9}$$

has cointegrating rank 1.

(See Appendix A for the proof.)[10] Given Eq. (2.9), it is straightforward to reparameterise the dynamics of the subsystem into the form of Eq. (2.8). Letting

[10] The condition of Theorem 6 is sufficient but not necessary. A non-IC vector having a solved dynamic form of rank 1 would arise if the 'droppable' variable(s) do not appear in *any* cointegrating relations, which is not ruled out (i.e., rows of β can be zero).

$B(1) = \alpha_{1a}\beta'_{1a}$ (see the proof of Theorem 6 for details of the notation) construct a full-rank matrix $G(g_1 \times g_1)$ with a partition into the first column, $G_1 = (\alpha'_{1a}\alpha'_{1a})^{-1}\alpha_{1a}$, and the remaining $g_1 - 1$ columns G_2 orthogonal to α_{1a}, with the property $G'_2\alpha_{1a} = 0$. Then define the partition

$$G'_2 B^*(z) = [C_1(z), C_2(z)] \tag{2.10}$$

where $C_1(z)$ is $(g_1 - 1) \times 1$ and $C_2(z)$ $((g_1 - 1) \times (g_1 - 1))$ has all its roots outside the unit circle by assumption. It is now easy to see that after transforming system (2.9) by premultiplication by G', it has the partition

$$x_{a,1t} = \gamma'_1 x_{a,2t} + w_{1t},$$

$$\Delta x_{a,2t} = w_{2t}, \tag{2.11}$$

where $(1, \ -\gamma'_1) = \beta'_{1a}$,

$$w_{1t} = v_{1t} - (\alpha'_{1a}\alpha_{1a})^{-1}\alpha_{1a}B^*(L)\Delta x_{at} \sim I(0) \tag{2.12}$$

and

$$w_{2t} = C_2(L)^{-1}(v_{2t} - C_1(L)\Delta x_{a,1t}) \sim I(0). \tag{2.13}$$

While model (2.11) is not a finite-order VAR in general, the short-run dynamics may (for example) be approximated non-parametrically by the estimation methods cited.

3. Testing for irreducibly cointegrated subsets

3.1. The MINIMAL Algorithm

The analysis of Section 2 suggests that a good starting point for an empirical investigation of a set of $I(1)$ variables might be to catalogue the IC subsets. To know whether the long-run structural equilibrium relations suggested by economic theory actually belong to this collection, in the available data, could often be useful in, for example, pointing to potentially fruitful avenues of research and ruling out others. There are $2^m - m - 1$ possible subsets of two or more out of m variables, and even with a moderate value of m, an ad hoc data exploration could not be relied on to examine every one of these. Except in the smallest data sets, a systematic search procedure is required to be sure of detecting all the relevant subsets.

One could, in principle, perform cointegration tests on all the possible subsets systematically, and merely study the full printout. Once a set is judged on the evidence to be cointegrated, all supersets of this set can be excluded from the analysis on the basis that only IC vectors are interesting. Once all non-cointegrated sets, and cointegrated supersets, have been eliminated, the remaining sets of variables can be categorised as IC and subjected to further study. Pursuing this approach leads naturally to the idea of automating the elimination procedure. MINIMAL is an algorithm that performs this task, implemented in the GAUSS language.

The essential component of MINIMAL is a self-calling procedure whose code is listed in Appendix B. The procedure takes as its input a set of 3 or more $I(1)$ variables, which at the initial call is the complete data set. It returns the value of a logical variable set to 'True' if the input set contains a cointegrated subset, according to the test criteria set by the user. The procedure itself consists of a loop that drops each variable in turn. If more than two then remain, the procedure calls itself with the remaining variables as input, and if 'True' is returned, the loop is incremented directly. Otherwise, the cointegration test is performed on the subset,[11] and if cointegration is found this subset is necessarily IC on the criteria, and is printed out. When execution of the loop is complete, the procedure returns with the value 'True' unless none of the subsets was found cointegrated, in which case 'False' is returned.

Thus, in effect, sequences of tests are conducted as follows. Start with a pair of variables, test these for cointegration, and then add variables one at a time to the set until a cointegrated subset is found; repeat this sequence of operations in every possible way. When the program terminates, every subset of the variables will have been tested for cointegration unless a subset *of* the subset has previously been found cointegrated. The output lists all the sets of cointegrated variables that have no cointegrated subset. The 'bottom-up' structure of the algorithm ensures that all such sets are detected with the smallest number of tests. It interesting to draw a comparison here with the more usual type of specification search in econometrics. The 'general-to-particular' mode of search, starting with a general model and specialising it until a rejection is obtained on a test for correct specification, is often preferred to the 'particular-to-general' method. MINIMAL may appear to belong in the latter category, but it should be noted that since *every* subset is tested, the ordering of the tests does not affect the outcome. It is easy to verify that unless one or more cointegrating vectors are found (which eliminates the corresponding supersets from the search), every one of the possible $2^m - m - 1$ subsets are tested by the algorithm. The critical rule is that every superset of a cointegrated set (on the chosen criterion) is treated as cointegrated. A 'top-down' search would require many more tests to be

[11] To avoid multiple reporting, the test subset is first checked against the list of IC sets previously found.

performed, but if the superset rule is adopted, could not lead to a different outcome.[12]

The test used to implement MINIMAL could take either non-cointegration or cointegration as the null hypothesis. This only makes a difference when we have to take into account the effect of false inferences, to which different probabilities are then implicitly assigned. The procedure was implemented initially using the usual Engle–Granger (1987) methodology, of running a regression and testing the hypothesis of non-cointegration of the residuals. However, residual-based tests introduce a number of well-known problems. One must choose a normalisation for the regression, and the outcome of the tests is not invariant to this choice. In addition, these methods are known to suffer low power, see for example the study of Kremers et al. (1992).

An alternative approach, assuming the cointegrating rank of the system is known, is to perform Wald tests of the relevant restrictions on the estimates of β in Eq. (2.5) yielded by the Johansen (1988, 1991) maximum likelihood procedure. The hypothesis to be tested takes the form that the cointegrating space contains a vector satisfying certain exclusion restrictions, such that the included variables form a cointegrated subset. The hypothesis of p exclusions can be expressed in the form

$$H\beta a = 0 \quad (p \times 1) \tag{3.1}$$

where (after suitable ordering of the variables) $H = [0 : I_p]$ $(p \times m)$, and a represents the eigenvector corresponding to the smallest eigenvalue of the matrix $\beta' H' H \beta$. In Davidson (1998) it is shown that a suitably constructed quadratic form in the vector $H\hat{\beta}\hat{a}$, where $\hat{\beta}$ is the Johansen MLE, and $\hat{a}$ the sample counterpart of a corresponding to $\hat{\beta}$, is asymptotically chi-squared with $\min(p, m - s)$ degrees of freedom under the null hypothesis.[13] The test can be implemented in practice following a preliminary estimation of the cointegrating rank using (for example) the Johansen trace test. Subject to the null being true, a consistent, asymptotically mixed normal estimator of the restricted cointegrating vector is provided by

$$\hat{b} = G\hat{\beta}\hat{a} \quad ((m - p) \times 1) \tag{3.2}$$

[12] The alternative rule, that if a test on a particular set rejects cointegration, then all subsets are deemed without further testing to be non-cointegrating, could of course yield a different outcome. For consistency one must impose one rule or the other, and the present choice is viewed as the best one for this problem.

[13] We cannot test more than $m - s$ restrictions on a column of β because of the way the Johansen β matrix is normalised for estimation. If $p > m - s$ we can test $m - s$ linear combinations of the restrictions, which in general will hold only if the restrictions themselves do. In practice this is done by taking the Moore–Penrose inverse of the singular covariance matrix. See Davidson (1998) for details.

where $G = [I_{m-p} : 0]$. Moreover, if this vector is IC by virtue of being an identified structural relation, then $\hat{b}$ consistently estimates the structural parameters up to the choice of normalisation. This last result reflects the fact that Theorem 1 is also relevant to system estimators. A vector that is both cointegrating and satisfies the identifying restrictions must be the identified structural form. Illustrative output from MINIMAL applied to artificial data is reported in Appendix C.

3.2. Simulation experiments

This section reports some Monte Carlo replications of MINIMAL. Artificial data were generated from four different long-run models, each containing five variables. Three have cointegrating rank of 2, with differing amounts of overidentification (i.e., sparseness of the β matrix) since preliminary experiments suggested that this is a crucial factor in the exercise. The fourth has cointegrating rank 3. The values of the β matrices are as follows:

$$\underset{\text{Model 1}}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}} \underset{\text{Model 2}}{\begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}} \underset{\text{Model 3}}{\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 1.5 & -1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}} \underset{\text{Model 4}}{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}} \tag{3.3}$$

Model 1 is maximally overidentified, having disjoint cointegrated subsets, and has just the two structural IC vectors. It can also be verified by inspection that Model 2 has four IC vectors, Model 3 (which is just-identified) has five, and Model 4 has six.

For each of the long-run models, alternative dynamic structures were specified, having the triangular form of Eq. (2.9), but with different amounts of noise and short-run dependence. First, five independent AR processes were generated, of the form $u_{it} = \gamma u_{i,t-1} + \varepsilon_{it}$ where $\varepsilon_{it} \sim \mathrm{NI}(0,1)$, $i = 1, \ldots, 5$. Of these, the cases $i = 1, 2, 3$ in Models 1–3, and $i = 1, 2$ in Model 4, were cumulated to produce ARIMA(1,1,0) processes. The remaining variables were then generated by the equations defined in Eq. (3.3), with additive errors σu_{it}. (See Appendix C for an example, corresponding to Model 2). The parameters γ and σ were varied between experiments, the former being assigned the values 0, 0.3, 0.6 and 0.9, and the latter, 0.5, 1, 2 and 4. All the experiments were conducted with 100 observations. The Johansen maximum likelihood estimates were calculated, with deterministic trends being fitted (although not present in the data) and with the order of the VAR set first to 1, and then to 2. The latter corresponds to the actual maximum lag implicit in the structure, although the Johansen method implies some over-parameterisation, and the former was usually the value recommended by the Schwarz information criterion. The full set of experiments was conducted for both cases. Although technically misspecified, the more parsimonious model often gave the better results, especially in the trace tests.

Figs. 1 and 2 report a selection of the test results in graphical form.[14] These plots show the proportion of 'successes' in 1000 Monte Carlo replications, as a function of γ and σ. By a 'success' is meant (except in Fig. 1b) that MINIMAL correctly reported all the true IC vectors for the model, and no false ones. The shading in the figures distinguishes those regions of the response surface where the success rate lies in the different bands of width 0.2, ranging from better than 0.8 (white) down to 0.2 or worse (darkest).

Fig. 1 shows the results of four experiments (in the rows) on models 1, 2 and 3 respectively (shown in the columns). Row (a) shows the proportion of 'successes' in the application of MINIMAL using the Wald test where the true cointegrating rank is treated as known. Row (b) shows the proportion of successes in the sequential application of the Johansen trace test[15] to determine the cointegrating rank, starting with null hypothesis $s \leqslant m-1$, and reducing s until a rejection is obtained. Row (c) shows the results of treating the cointegrating rank as unknown, and running MINIMAL after choosing s according to the outcome of the trace test sequence. Finally, row (d) shows the results of running MINIMAL using a residual-based test for cointegration, the Phillips (1987) nonparametric $\hat{Z}_t$ test.[16] All of the tests in these experiments were carried out at the nominal 1% significance level.

To return a 'success' in these experiments, MINIMAL must make correct decisions in *all* of between 14 and 25 tests. By contrast, to determine the cointegrating rank in this instance requires correct decisions in just three tests. In view of this, the relative performance of MINIMAL revealed in Fig. 1a is noteworthy. Moreover, Fig. 1c shows that the penalty attached to not knowing s, although not negligible, is smaller than we would expect if the two sets of test outcomes were independent of each other. In data sets whose cointegrating rank can be determined, there is a relatively high probability that we can also correctly discover the IC vectors. These findings should not surprise us too

[14] In Fig. 1 the results reported are the VAR(1) estimates, while in Fig. 2 they are for the VAR(2). For the full set of results, in tabular form, see the working version of the paper, Davidson (1996).

[15] The trace test is implemented using the critical values given by Osterwald-Lenum (1992), and also the degrees of freedom correction of Reimers (1991) cited in Banerjee et al. (1993). In other experiments the maximal eigenvalue test was tried, but proved less successful in finding the correct rank.

[16] See Phillips and Ouliaris (1990) for details of this test. The Newey-West (1987) variance estimator was adopted. To overcome the problem of choice of normalisation, the residuals were calculated for every normalisation, and the test statistics averaged. Since no estimate of β is used, there is no need for a prior estimate of s in this implementation of MINIMAL.

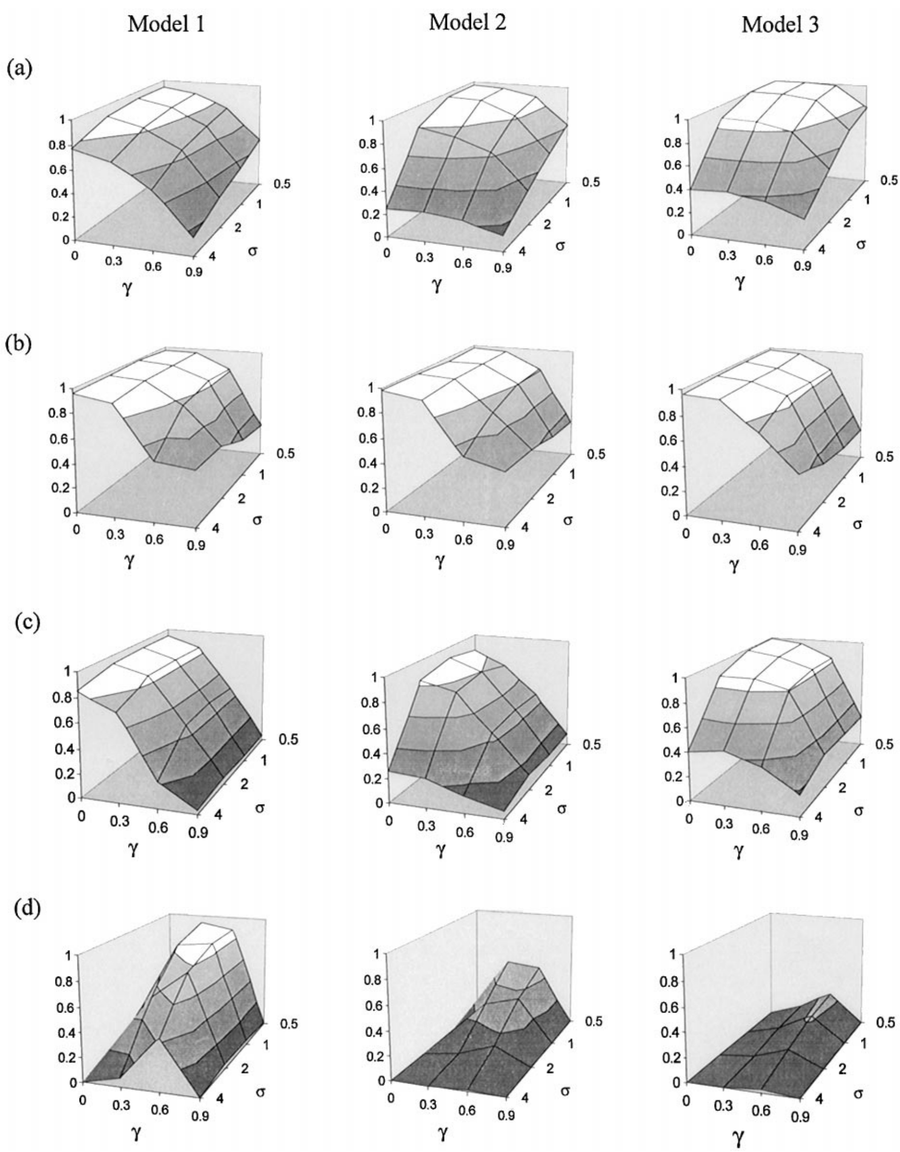

Fig. 1. (a) MINIMAL with known cointegrating rank; (b) cointegrating rank determination by the trace test; (c) MINIMAL with estimated cointegrating rank; (d) MINIMAL implemented with residual-based cointegration test.

much if we note that once s is set correctly the Wald test should be relatively powerful, exploiting the T^{-1}-consistency of the ML estimator. By contrast, MINIMAL implemented with the residual-based test (Fig. 1d) performs poorly.

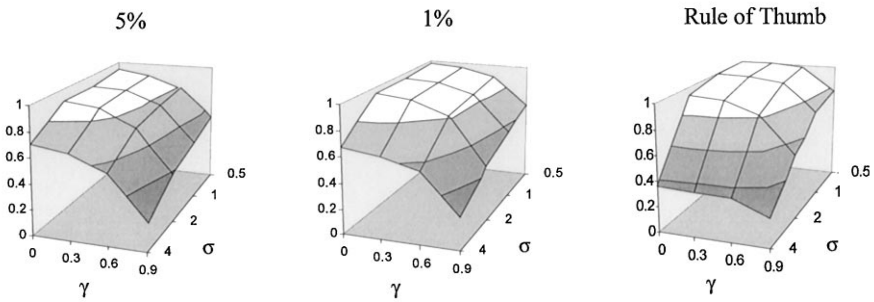

Fig. 2. Model 4: MINIMAL with different Wald test rejection criteria.

Fig. 2 shows the application of MINIMAL to Model 4 (with s treated as known) where in this case the three plots show the effect of changing the rejection criterion of the Wald test. The first two plots show the result of using 5% and 1% nominal significance levels respectively. Note that a 'success' is obtained by both rejecting and not rejecting certain hypotheses in combination, and therefore the probability of success is controlled by adjusting the size of the tests to match the available power. In favourable cases, it was found that choosing a nominal significance level substantially lower than 1% gave the best results. Various rules of thumb for varying the rejection criterion were tried, and the third plot in Fig. 2 shows the result of one of these. The rule of thumb used here is as follows: set the critical value equal to four times the 1% nominal value from the chi-squared tabulation when $p = m - 2$ (the hypothesis of a cointegrated pair), at twice the 1% nominal value for $p = m - 3$, and use the standard 1% test otherwise. This rule is spectacularly successful in favourable cases (low σ and γ) with near 100% success.[17]

For reasons of space we do not report here the corresponding results for Models 1–3, but can summarise the overall finding that there is no 'best' significance level to cover all cases. In general, the low significance levels do best with a high degree of overidentification. On the other hand, in just-identified cases such as Model 3, the test statistics for all the valid (just-identifying) restrictions are identically zero. Since we want to reject all the overidentifying restrictions in these cases, a smaller critical value will necessarily out-perform a larger one. In short, some experimentation with different rejection criteria would seem advisable in practical modelling situations.

[17] A point to note about this model is that when $p = m - 2 = 3$, the limiting null distribution has at most $m - s = 2$ degrees of freedom. See footnote 13.

Table 1

	$\hat{\beta}$		1	2
2.5% Consols rate	0.29	0.19	0.282 (0.095) [10.23]	
3 month Local Authorities rate	− 0.32	− 5.02	—	− 4.9 (0.036)
3 months uncovered Eurodollar rate	− 0.65	0.034	− 0.648 (0.122) [37.32]	
Inter-bank rate	0.19	4.91	—	5 (0.038)
\$/£ exchange rate	3.38	0.18	3.37 (0.0615) [3021]	
U.S. Govt. long bond yield	0.98	− 0.013	0.98 (0.143) [48.95]	
Wald			1.48	6.36
Phillips–Perron			− 4.325	− 6.77

3.3. MINIMAL analysis in practice

In this section, we apply the methodology to two sets of quarterly economic time series. The first is a collection of UK interest rates and related variables for the period 1963(4)–1984(2) (Table 1). The second consists of five US macroeconomic series, the logs of real output (y), consumption (c), investment (i), and real money balances ($m-p$), and the interest rate (R), for 1951(1)–1985(4) (Table 2). The latter are the variables analysed in the well-known study of King et al. (1991) (KPSW), and also by Pesaran and Shin (1994), among others. KPSW also include the rate of inflation in their study, but it is excluded here in view of doubt that it is truly $I(1)$.

The tables show the $\hat{\beta}$ matrices from the Johansen MLE, and then the estimates of the IC vectors, computed as $\hat{b}$ in Eq. (3.2), with 'standard errors' in parentheses.[18] These vectors are not normalised on any variable, but inherit their normalisation from that of $\hat{\beta}$. The numbers in square brackets, for vectors with more than two elements only, are the values of the Wald statistic obtained if the corresponding variable is dropped, as confirmation that the subset is

[18] As noted, these standard errors are stochastic asymptotically, but provide approximate confidence intervals in the sense that the ratios with the deviations of the point estimates from their true values are asymptotically N(0,1).

Table 2

	$\hat{\beta}$			1	2	3	4	5	6
y	− 11.3	− 63.3	− 5.56	—	− 9.27 (5.23)	—	14.7 (6.24)	− 48.5 (2.2) [43.5]	—
c	− 15.3	17.5	22	—	—	− 18.8 (5.94)	− 25.3 (2.22)	—	− 21.6 (1.27) [43.49]
i	− 2.06	3.63	− 14.2	− 11.5 (1.4) [43.49]	12.4 (1.39)	14.6 (1.3)	—	—	—
$m - p$	17.5	40.6	2.73	6.92 (2.36) [4.05]	—	—	—	39.1 (2.74) [110]	10.1 (2.77) [5.33]
R	0.86	0.22	− 0.15	0.486 (0.464) [1.96]	—	—	—	0.803 (0.539) [2.73]	0.826 (0.545) 3.01]
Wald				0	0.857	0.041	1.94	0	0
Phillips–Perron				− 4.44	− 4.48	− 4.52	− 4.83	− 5.98	− 4.48

indeed IC. At the foot of each column there appears both the Wald statistic for the vector itself and the Phillips–Perron test (1988) statistic for non-cointegration, computed from the residuals generated by $\hat{b}$. Note that $\hat{b}$ converges to a fixed limit at rate T^{-1} under both null and alternative hypotheses. Asymptotically the test can be treated as a simple test for a unit root, without the usual complications associated with residual-based tests where the regression coefficients converge to random limits under the null. This statistic provides an independent check that the estimated relation is in fact cointegrating.[19] Failure to reject would point to the possibility that the cointegrating rank is incorrectly chosen, or simply that the sample evidence is weak.

For the interest rate data (Table 1) the maximum lag of the VAR was set to 1 on the basis of the Schwarz information criterion, taking note of the simulation evidence cited above, which cautions against over-parameterisation. A deterministic trend is assumed. The maximal eigenvalue and trace tests both indicate a cointegrating rank of 2. Two IC vectors were found, implying a maximally overidentified structure with two disjoint IC subsets, so that no solved vectors exist. The results shown were obtained by setting the significance level of the Wald tests to 5%. Setting the significance level to 1% results in the exclusion of the Consols rate in vector 1. (The frailty of its role is also indicated by the exclusion test statistic, which is only 10.2.) Vector 1 explains the Eurodollar rate as a weighted average of US and UK government bond rates, also depending positively on the strength of sterling against the dollar. Vector 2 tells the simple story of a close co-movement between two UK short rates.[20]

In the US data, the Akaike and Schwarz criteria both indicate a maximum lag of 2. The sample evidence is equivocal here, since for two of the variables, i and R, the unit root hypothesis is rejected once a deterministic trend is allowed for. It is also not possible to reject the hypothesis that the cointegrating rank is 2. However, in the spirit of attempting to reproduce previous results we over-ride this result, and assume a five-variable model with a cointegrating rank of 3. Perhaps because β is not so well determined in this instance, the Wald test does not reject various theoretically implausible restrictions at the conventional significance levels. However, if we rank the vectors according to the value of the Wald statistic, we find that the six IC vectors which the evidence most favours are precisely those suggested by the KPSW model. Table 2 shows the result of running MINIMAL with (in effect) the nominal significance level set somewhat over 10%. According to the KPSW analysis, the three 'structural' vectors are 2, 4 and 5. Vectors 1, 3 and 6 can be interpreted as solved from 2 and 5, 2 and 4, and

[19] The 5% critical value for $n \approx 100$, and allowing for a deterministic trend, is -3.46.

[20] The gap between these two series rarely exceeds 0.5% but is quite highly autocorrelated, which explains why the cointegration between them is strong but not overwhelming.

4 and 5, respectively. This is the complete set of IC vectors for the model. However, note the fragility of the inclusion of R in all the vectors containing it, especially vector 1, and also the fact that these coefficients are all within two standard errors of zero.

Finally, we note in passing that other structural restrictions on β could easily be tested by the Wald procedure, for example, the unit elasticity restrictions on the (y, i) and (c, y) relations. This avoids the need for iterative estimation under the null hypothesis, but it is more important to note how it would allow such hypotheses to be tested without fully specifying the structural model. We could, for example, test the hypothesis: 'the cointegrating space contains a vector $(1, -1, 0, 0, 0)'$, up to a normalising constant' without specifying anything else, other than the cointegrating rank of the system.

4. Conclusion

In this paper we have defined the notion of an irreducible cointegrating relation, or vector. We show that such vectors have a special relationship with the long-run structural relations of the data generation process, when the latter are subject to identifying exclusion restrictions. This is a feature of cointegration analysis that appears not to have been noted explicitly before. Sequential tests for cointegration can determine whether the IC property holds for a given set of time series. In principle only single equation methods are needed for the analysis, but working with a system estimator proves much more efficient in practice. We show in Monte Carlo experiments that sequential testing using the MINIMAL algorithm to determine the IC relations can have a success frequency comparable to the correct determination of the cointegrating rank by the usual sequential testing. The method may therefore prove useful as an adjunct to conventional methods of analysis, particularly if it is difficult to specify the long-run structure completely on the basis of prior theory.

We show that in certain circumstances, when there is overidentification of the long-run structure, we might be able to learn something about that structure through nothing more than inspection of the irreducible cointegrating regressions associated with a set of data. Such an outcome is by no means guaranteed in practice, since only overidentified models can have this property, and maximum overidentification (no overlap of the cointegrated subsets) is necessary to learn the complete structure in this way. Nevertheless, the result runs counter to the popular intuition about simultaneous structural models, drawing as it does on our experience of the standard stationary-data case, and is of some interest for this reason alone. The possibility that 'incredible assumptions' (in the oft-quoted words of Christopher Sims, 1980) need not always be the price of obtaining structural estimates turns out to be a distinctive feature of models with stochastic trends.

Acknowledgements

I am grateful to Jan Magnus for advice, and to Peter Boswijk, Laurence Copeland, Russel Davidson, Jurgen Doornik, David Hendry, Cheng Hsiao, David Peel, Tom Rothenberg, two anonymous referees, and associate editor for their comments on earlier versions of the paper. The usual disclaimer applies.

Appendix A. Proofs

Proof of Lemma 1. Let a solved IC relation $\delta' x_t$ be defined by $\delta = \beta r$, where r $(s \times 1)$ is a vector of weights which, for irreducibility, must equate at least $s-1$ elements of δ to zero. With no loss of generality, consider the first structural cointegrating relation, having coefficients β_1 $(m \times 1)$. Let this be identified and hence IC by Theorem 3, and assume $r_1 \neq 0$, so that δ depends on it. Next, let r^* $(s \times 1)$ to be equal to r except for its first element, which is 0, and define $\delta^* = \beta r^*$ such that

$$\delta = r_1 \beta_1 + \delta^*.$$

By the assumption of identification, there exists no r, as defined, such that δ^* has the same zero restrictions as β_1. In the absence of linear restrictions across the columns of β, r_1 can be chosen to equate at most one element of δ to zero. It follows that δ contains at least as many nonzero elements as β_1. □

Proof of Theorem 4. First, consider the case where each relation in the system is either just-identified or underidentified. Then every identified structural relation, and every solved relation, contains exactly $m-s+1$ variables, and there is no uniquely smallest IC relation. Next, suppose there are overidentified structural relations. These contain fewer than $m-s+1$ variables, and Lemma 1 shows that they never contain more variables than the solved relations containing them. If there exists a uniquely smallest IC relation, such that all the others contain more variables, it follows that this must be structural.

Next consider an arbitrary subset of the variables, denoted S. The set $A(S)$ of all the IC relations containing at least one member of subset S consists of all the identified structural relations containing elements of S, plus all the solved IC relations derived, in part or wholly, from these. Note that each IC relation that is solved from a particular structural relation must, by construction, have one or more variables in common with it. Suppose $A(S)$ contains a relation, R, which is *strictly* the smallest (has fewest variables) in $A(S)$. R cannot be a solved relation according to Lemma 1, and hence must be structural. Note that this is true even if there are equally small or smaller IC relations not belonging to $A(S)$, since R cannot be solved from any of these by definition. □

Proof of Theorem 5. Suppose there is a structural relation, Equation (1) (say), containing a variable x_{1t} (say) which appears in no other structural relation. First, let equation (1) be identified, and hence IC by Theorem 3. Since x_{1t} appears in only the one equation, any solved relation involving equation (1) must contain x_{1t} by definition. Given the existence of such a relation, x_{1t} would appear in at least two IC relations, contrary to supposition.

Second, suppose equation (1) is underidentified. This means there is a linear combination of equations (2)–(5) that satisfies the same zero restrictions as equation (1). This vector lies in the cointegrating space, and by the supposition that x_{1t} appears uniquely in equation (1), its first element must be zero. It follows that x_{1t} can be dropped from equation (1) without losing cointegration, and hence, that x_{1t} is not cointegrated with any combination of $x_{2t},\dots,x_{mt}$. It therefore cannot appear in any IC relation, whether structural or solved. □

Proof of Theorem 6. Assume initially that the vector in question is structural and identified. Re-order the system such that the coefficients are $\beta_{1a} = (1,\beta_{21},\dots,\beta_{g_1 1})'$ as in Eqs. (2.3) and (2.4), and such that after partitioning Eq. (2.5) as

$$A_{aa}(L)x_{at} + A_{ab}(L)x_{bt} = u_{at} \quad (g_1 \times 1)$$

$$A_{ab}(L)x_{at} + A_{bb}(L)x_{bt} = u_{bt} \quad ((m - g_1) \times 1) \tag{A.1}$$

the polynomial matrix $A_{bb}(z)$ has full rank. That is to say, $A_{bb}(z)$ is non-singular at all points $z \in C$ other than the solutions of $|A_{bb}(z)| = 0$. The second block in Eq. (A.1) may be solved for x_{bt} as

$$x_{bt} = A_{bb}(L)^{-1}u_{bt} - A_{bb}(L)^{-1}A_{ba}(L)x_{at}. \tag{A.2}$$

Since β spans the cointegrating space we must have, from Eqs. (2.4) and (A.2),

$$\beta'_a x_{at} = -\beta'_b x_{bt} + I(0) = -D(L)x_{at} + I(0) \tag{A.3}$$

where $D(z) = \beta'_b A_{bb}(z)^{-1}A_{ba}(z)$ $(s \times g_1)$. We cannot assume that $A_{bb}(1)^{-1}$ exists, but the top row of $D(z)$ is zero since the top row of $\beta_b{}'$ is zero from Eq. (2.4). If $D_2(z)$ denotes the remaining rows, the decomposition $D_2(z) = D_2(1) + D_2^*(z)(1 - z)$ exists by construction. By the irreducibility hypothesis and Theorem 3, $\mathrm{rank}(\beta_{2b}) = s - 1$, and hence none of the columns of β_{2a} are independently cointegrating for x_{at}. It is therefore clear from Eq. (A.3) that $\beta'_{2a} = D_2(1)$.

Substituting for x_{bt} in the first block of Eq. (A.1) we obtain Eq. (2.9), where $B(z) = A_{aa}(z) - A_{ab}(z)A_{bb}(z)^{-1}A_{ba}(z)$ and $v_t = u_{at} - A_{ab}(L)A_{bb}(L)^{-1}u_{bt}$. Using $A_{aa}(1) = \alpha_a\beta'_a$ and $A_{ab}(1) = \alpha_a\beta'_b$, where α_a is $g_1 \times s$, we find

$$B(1) = \alpha_a(\beta'_a - D(1)) = \alpha_a\begin{bmatrix}\beta'_{1a}\\0\end{bmatrix} = \alpha_{1a}\beta'_{1a}, \tag{A.4}$$

where α_{1a} ($g_1 \times 1$) is the first column of α_a. Two conclusions may be drawn from this. First, applying the usual decomposition $B(z) = \alpha_{1a}\beta'_{1a} + B^*(z)(1-z)$ as in Eq. (2.5), we have

$$v_t = B(L)x_{at} = \alpha_{1a}\beta'_{1a}x_{at} + B^*(L)\Delta x_{at} \sim I(0) \tag{A.5}$$

so that Eq. (2.9) holds. Second, $B(1)$ has rank 1, and β_{1a} is the unique cointegrating vector of this system.

This completes the proof for the identified structural case. If the vector in question is a solved form, let β^* denote the 'true' structure, and assume that Eq. (2.5) represents a reparameterised structure with $\beta = \beta^* d$ where d is $s \times s$ non-singular, of which the first column $\beta_1 = \beta^* d_1$ is the solved form in question. Since by assumption the cointegrating vector β_{1a} is irreducible, it must be 'identified' in the reparameterised model, in the sense that rank $(\beta_{2b}) = s - 1$, by Theorem 3. The previous analysis can now be performed on this case, completing the proof. □

Appendix B. GAUSS code for the MINIMAL elimination procedure

```
proc elim(a,anvar,ia);
local ib,b,bnvar,coivec,result;
  coivec = 0;
  ib = ia;
  bnvar = anvar - 1;
  do while ib < = anvar;
    result = 0;
    b = getvec(a, anvar, ib);
    if bnvar > 2;
      if elim(b,bnvar,ib) = = 0;
        result = testb(b,bnvar);
      else;
        result = 1;
      endif;
    else;
      result = testb(b,bnvar);
    endif;
```

```
if result = = 1;
  coivec = 1;
endif;
    ib = ib + 1;
  endo;
retp(coivec);
endp;
```

Notes:

1. `ia` is a scalar, set to 1 when procedure `elim` is called from the main program. `anvar` is a scalar set initially to m, the number of variables in the data set. `a` is a vector of dimension `anvar`, containing the storage locations of variables in the data set. On the initial call, `a` $= \{1, \dots, m\}$.
2. Procedure `getvec` removes the `ib`th element from a, returning `a` vector of dimension anvar $-$ 1.
3. Procedure `testb` returns the result of the cointegration test. Before doing the test it checks if the current set has been found to be IC at a previous stage, in which case a positive result is returned directly. Otherwise, if the test result is positive the set is added to the list of IC subsets, and printed out.

Appendix C. An example of MINIMAL output, with simulated data

Let variables X_{1t}, X_{2t}, X_{3t} be generated by $X_{it} = X_{i,t-1} + U_{it}$, with $X_{i0} = 0$, and then X_{4t} and X_{5t} be generated by

$$X_{4t} = X_{1t} + X_{2t} + X_{3t} + U_{4t}, \tag{C.1}$$

$$X_{5t} = X_{1t} - X_{2t} + U_{5t}, \tag{C.2}$$

where $U_{it} \sim \mathrm{NID}(0,1)$ for $i = 1, \dots, 5$. In this model, the cointegrating rank (s) is 2, and the number of IC vectors is 4. These are the 'structures' (C.1) and (C.2), and two solved relations obtained as the sum and difference of (C.1) and (C.2), respectively.

Table 3 shows the output from MINIMAL applied to a sample of size 100 generated from this DGP, following Johansen estimation and using the trace test to set the cointegrating rank (correctly) at 2. The nominal 5% significance level was used in the Wald tests. The first two columns of the table contain the cointegrating vectors estimated by the Johansen MLE, and the next four are the reported IC vectors, obtained with the formula in Eq. (3.2), with standard errors in parentheses. The vectors â are shown, transposed, in the top row. The numbers in the bottom row are the Wald statistics. Note that these equal 0 identically for fewer than $s = 2$ restrictions. The third and fourth columns may be recognised as estimates of Eqs. (C.1) and (C.2) respectively (with arbitrary

Table 3

			$\hat{a}'$ 1	2	3	4
	$\hat{\beta}$		(1, − 0.003)	(− 0.13, − .99)	(.61, − .78)	(.74,.67)
X_1	− 0.0007	− 1.47	—	1.46 (0.014)	1.16 (0.020)	− 0.98 (0.015)
X_2	− 1.61	0.22	− 1.61 (0.011)	—	− 1.16 (0.014)	− 1.05 (0.010)
X_3	− 0.83	− 0.63	− 0.83) (0.022)	0.74 (0.018)	—	− 1.04 (0.020)
X_4	0.75	0.604	0.749 (0.023)	− 0.70 (0.019)	—	0.96 (0.021)
X_5	− 0.76	0.85	− 0.763 (0.024)	− 0.741 (0.020)	− 1.14 (0.022)	—
		Wald statistics	0	0	0.491	0

normalisation, although in this instance it is close to the original normalisation) and first two are the solved forms.

References

Banerjee, A., Dolado, J.J., Galbraith, J.W., Hendry, D.F., 1993. Co-integration, error correction and the econometric analysis of nonstationary data. Oxford University Press, Oxford, London.

Boswijk, H.P., 1995. Efficient inference on cointegration parameters in structural error correction models. Journal of Econometrics 69 (1), 133–158.

Davidson, J., 1994. Identifying cointegrating regressions by the rank condition. Oxford Bulletin of Economics and Statistics 56 (1), 103–108.

Davidson, J., 1996. Structural relations, cointegration and identification: Some simple results and their application. Cardiff Business School Economics Discussion Paper 96, 081.

Davidson, J., 1998. A Wald test of restrictions on the cointegrating space based on Johansen's estimator. Economics Letters (59) 2 (1998), 183–187.

Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction: Representation, estimation and testing. Econometrica 55 (2), 251–276.

Hsiao, C., 1997. Cointegration and dynamic simultaneous equations model. Econometrica 63 (3), 647–670.

Hendry, D.F., 1995. Dynamic Econometrics. Oxford University Press, Oxford, London.

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. Econometrica 59 (6), 1551–1580.

Johansen, S., 1995. Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. Journal of Econometrics 69, 111–132.

Johnston, J., 1984. Econometric Methods. 3rd ed, McGraw Hill, New York.

King, R.G., Plosser, C.I., Stock, J.H., Watson, M.W., 1991. Stochastic Trends and Economic Fluctuations. American Economic Review 81 (4), 819–840.

Kremers, J.J.M., Ericsson, N.R., Dolado, J.J., 1992. The power of cointegration tests. Oxford Bulletin of Economics and Statistics 54 (3), 325–348.

Newey, W.K., West, K., 1987. A simple positive definite heteroskedasticity and correlation consistent covariance matrix. Econometrica 55, 703–708.

Osterwald-Lenum, M., 1992. A note with fractiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics: Four cases. Oxford Bulletin of Economics and Statistics 54, 461–472.

Pesaran, M.H., Shin, Y., 1994. Long-run structural modelling. Working Paper. Department of Applied Economics. University of Cambridge (September).

Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55 (2), 277–301.

Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. Econometrica 59(2) (March), 283–306.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical Inference in Instrumental Variables Regression with I(1) Processes. Review of Economic Studies 57, 99–125.

Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic Properties of residual based tests for cointegration. Econometrica 58 (1), 165–193.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Reimers, H.E., 1991. Comparisons of tests of multivariate cointegration. Discussion Paper 58. Christian-Albrechts University, Kiel.

Sims, C., 1980. Macroeconomics and Reality. Econometrica 48 (1), 1–48.