UvA ECONOMETRICS

Discussion Paper: 2002/16

Automatic identification and restriction of the cointegration space

Pieter Omtzigt

UvA UNIVERSITEIT VAN AMSTERDAM

Automatic identification and restriction of the cointegration space

Pieter Omtzigt *

January 2, 2003

Abstract

We automate the process of finding the cointegration relations in a cointegrated VAR. There is a rigorous separation between the theory part (search directions must be defined, a final model chosen) and the automated search. The decision rules are set in such a way that a theoretical upper limit can be given to the asymptotic size of recovering all overidentifying restrictions. A Monte Carlo study shows that the algorithm works well, but that the properties of the asymptotic tests are rather poor at times. The software (in Matlab) to execute the algorithm is available.

1 Introduction

The introduction of the concept to cointegration (Engle and Granger, 1987) has led to the development of a wide variety of methods to analyze cointegrated systems. The method of Johansen (1988, 1991) is frequently applied as part of a wider General-to-Specific Modelling strategy which is advocated by the LSE-school in econometrics.

The process of arriving at the specific model is often long and arduous, especially in a VAR with more than 3 variables. Many sequential decisions are taken in the modelling process: which dummy should be included? which accepted restrictions on the cointegration parameters should be tested jointly? Criticism against the LSE methodology often targets these procedures as leaving too many options for the individual researcher and turning the process into an art form instead of science. From the LSE practitioner point of view, the process is tedious and indeed difficult to replicate. Starting with a particular data set once analyzed, one does not necessarily take the same decisions again and it is quite hard to exactly replicate the analysis.

After the article by Lovell (1983) , data mining and automated model selection were seen in a bad light. Recently Hoover and Perez (1999) and Hendry and Krolzig (2003) reran the original experiment, changed the decision rules and found that it was possible

---
*Universiteit van Amsterdam, Roetersstraat 11, 1018WB Amsterdam Email: P.H.Omtzigt@uva.nl

to recover the original DGP with a very high probability. The last article also contains an exhaustive justification of automated modelling and the LSE-framework.

Whereas the three cited articles in the last paragraph are concerned with selecting the right variables in a single equation regression, Brüggeman and Lütkepohl (2000) consider lag length selection in a vector autoregressive model. Davidson (1998a) automates finding the zero-restrictions[1] in the cointegration space. His aim is very close to the one pursued in this paper. Yet his method differs from ours in a number of key elements, which will be fully discussed in section 5.4.

In this chapter we will consider the cointegrated VAR-model and automate the search for any within-equation restrictions in the cointegration space.

In the next section I will outline the model, show how identification and restriction are intimately connected and describe the standard modelling strategy employed. In section 3 the algorithm is stated. A worked-out example, a Monte Carlo simulation and an application to UK money demand data then amply illustrate its use. Extensive comments on the uses of the procedure are given before the concluding remarks in the last section.

2   The cointegrated VAR model

The p-dimensional cointegrated VAR model is given by:

$$\Delta X_t = \alpha\beta' \begin{pmatrix} X_{t-1} \\ Y_{t-1} \\ D_t \end{pmatrix} + \Gamma_0 \Delta Y_t + \sum_{i=1}^{k-1} \Gamma_i \begin{pmatrix} \Delta X_{t-i} \\ \Delta Y_{t-i} \\ \Delta D_{t-i} \end{pmatrix} + \Phi d_t + \varepsilon_t \qquad (1)$$
$$\varepsilon_t \sim iidN(0, \Omega)$$

where α is of dimension $p \times r$ and β is of dimension $q \times r$, Y_t is a vector of exogenous variables and D_t a vector with deterministic variables, which are included in the cointegration space. The lagged differences variables inside the cointegration space are explicitly included outside the cointegration space. Some authors and computer packages take this convention (CATS in RATS,Hansen and Juselius 1994), whereas others do not include ΔD_{t-i}, the lagged differenced deterministic variables, outside the cointegration space (like PcFiml, Doornik and Hendry 1997). The d_t contains other dummies like seasonal and blip[2] dummies.

Neither α nor β is identified, but only their product $\Pi = \alpha\beta'$ is. The likelihood does not change if we take $\alpha^* = \alpha\kappa$ and $\beta'^* = \kappa^{-1}\beta'$ for an arbitrary invertible matrix κ. We assume that all the identifying restrictions are put on β and that there are either no restrictions on α or only restrictions of the kind $\alpha = Ha$. The last kind of restrictions, which corresponds to weak exogeneity of $H'_\perp X_t$ for α and β, does not bring any identification onto the system.

For β we consider only within-equation restrictions:

$$\beta = (H_1\varphi_1, \psi) \qquad (2)$$

---

[1]That is he finds overidentifying restrictions in the cointegration space, but only exclusion-restriction.
[2]They take value 1 in one (or two) periods and are zero elsewhere.

or

$$\beta = (H_1\varphi_1, \ldots, H_r\varphi_r) \tag{3}$$

where H_i are $q \times s_i$ matrices, $s_i \leq q + 1 - r$.

2.1 Identification and restriction

Identification and restriction are so intimately related, that it is impossible to separate them out completely. Let's take the following cointegration space in a DGP as an example:

$$\beta'_{DGP} = \begin{bmatrix} 1 & b & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

If we decide to identify the space by putting an identity matrix in the top of β:

$$\beta'_{M1} = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & c \end{bmatrix}$$

then by successive testing whether $a = 0$ and or $c = 0$ we are unable to recover the DGP. If we had started with the following identification scheme however:

$$\beta'_{M2} = \begin{bmatrix} 1 & d & 0 \\ 0 & e & 1 \end{bmatrix}$$

we would have been able to recover the DGP. As it is impossible to know the DGP before hand, we shall have to treat the problem of identification and restriction contemporaneously in the algorithm.

The algorithm we propose shall mimic the strategy which Juselius (2002) explicitly employs in her papers. First she finds which restrictions of the kind (2) are accepted. She then combines the accepted restrictions to (over)-identify the whole parameter space (3) and reports the final combination which is supported by the data. Usually she will test a great many of these combinations, but for lack of space it is normally only possible to report one or two final models. We shall systematically search for such combinations.

2.2 Modelling Strategy

This paragraph gives a short overview as how a small cointegrated VAR is often modelled in an LSE-type fashion. This is no definite guide, but more an attempt to describe current practice. We refer once again to the work of Juselius as an example. The modelling process can graphically be represented as in figure 1.

1. The selection of variables is based on the theories to be tested or the part of the economy to be modelled. If money demand and monetary transmission mechanisms are the object of study, then logically money, an interest rate, income and prices/inflation should be part of the data vector to get a meaningful model: these variables are dictated by the standard textbook theories on money. Money demand

3

Test ADF/PP

Theory

Data (quality)

Selection of variables and deterministics

Outlier/Misspecification

Economics

History

Dummy variables

Misspecification/ IC

Select VAR lag length

Trace test

Theory

Select rank

χ² test

Theory

Restrictions on β

χ² test

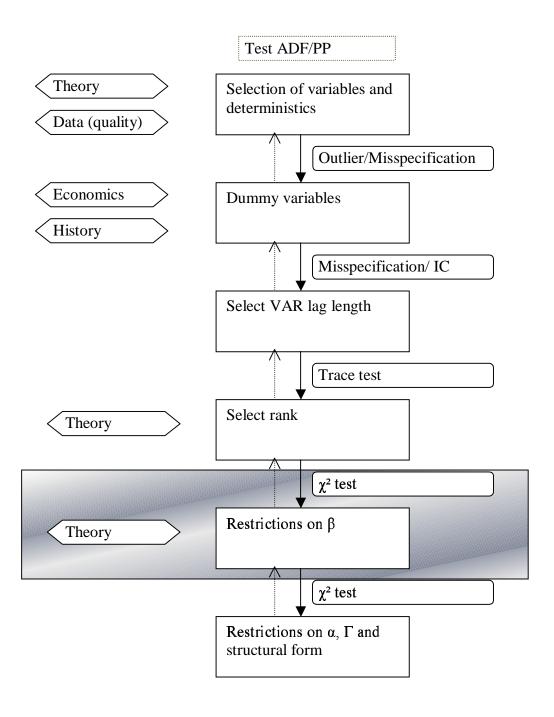Restrictions on α, Γ and structural form

Figure 1: Modelling a cointegrated VAR

can for instance depend on inflation, total income and the relevant opportunity cost, which is measured by means of an interest rate. Often many measures of money (and income) are available. When abstract theories do not offer a convincing choice, data quality often does. Series may contain breaks in the form of changes of definition, changes of bank concerned etcetera. Long series of high quality are often hard to find.

An additional problem is possible I(2)-ness in the data. If an I(1) model is to be preferred, then some pretesting is warranted. This can take the form of unit root tests on series. The augmented Dickey Fuller test and the Phillips-Perron test are often used. Alternatively one can do a full I(2)-rank test. If at least one I(2)-trend is found, a test for a so-called nominal-to-real transformation (Kongsted, 1998, 1999) can be executed to transform the model into an I(1) model.

In choosing the deterministics of the model, preference is often given to so-called star-models, which allow for the same kind of trend in both the stationary and the non-stationary direction. This means either a constant in the cointegration space or a trend in the cointegration space D_t and a constant (differenced trend) outside. This fits in naturally with our formulation of the problem. For these specifications, the rank test is asymptotically similar with respect to the actual trends in the data, such that rank and deterministics determination can be separated (Nielsen and Rahbek, 2000). Restrictions on the trend/constant can be tested in 5 and 6.

2. The procedure we are using is based on a Gaussian likelihood function. If there are large outliers, then these have to be modelled explicitly. Tests on the residuals will reveal whether any misspecification is present. Outliers often have a meaning and several solutions to outlier problems can be employed. If the outliers are caused by the say the oil shocks, then the economic and historical knowledge can lead either to the inclusion of one or two dummies for these shocks or to the inclusion of the oil price as an exogenous variable in the model.

   The dotted arrow between dummy variables and selection of variables on the one hand and deterministics on the other indicates the idea that it is possible to go back one stage in the modelling process and rethink which variables should be included. In fact at any stage in the modelling process the researcher can go back to previous stages.

3. If the model is well-specified, the lag length of the process is set: in the model the lag length is equal on all variables (though this can be changed later). This can be done either on the basis of an information criterion like Aikaike's Information Criterion or by testing successively that the coefficients to the last lag present in the model are zero.

4. The rank of the matrix Π is selected using the trace test Johansen (1995b) . The maximum eigenvalue is not frequently used. If there are strong theoretical priors they can be included in the procedure, see Paruolo (2002) .

5. Restrictions on the cointegration space are tested by means of likelihood ratio test.

These restrictions are usually motivated by economic theory. If real money and real income are the first two variables in our data vector and we want to consider whether velocity of money is stationary, we test an hypothesis of the kind (2) with $H_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}$[3]. Consequently accepted hypotheses are combined to restrict the whole cointegration space.

6. Restrictions on the short run parameters α, Γ or on the so-called structural form, can be tested for known β, as the last are estimated in a superconsistent way and the former are not. If interest only lies in testing theories which involve just the cointegration parameters, then this step can be omitted.

The various parts of the modelling process do not take up equal amounts of time. In applied work especially number 5, the testing of restrictions often requires a large amount of time. This part will be object of study in the next paragraph and then be automated.

3   Identification and restriction of β

Current practice is to find sets of individual restrictions, based on economic theory. One vector is restricted and the other vectors are left to vary freely. We thus have a test of the kind (2). Hypotheses, which are accepted at a set significance level (often 5%) are then tested jointly. To my knowledge, few authors, if any, actually report which combinations they have tested and why and how they have selected the final model. Usually one reports only the final combination of restrictions accepted and the p-value associated with it.

The following algorithm mimics and automates this process. It is graphically represented in figure 2.

1. Let the user specify $q \times 1$ restriction columns of the kind $f_i = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}$, $i = 1, \ldots, F$. These can be interpreted as new variables, that might be stationary themselves or enter into a stationary relationship. Examples include the real rate of interest, velocity and interest rate spread.

2. Letting $e_j = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, j = 1, \ldots, q$ denote the unit vector with 1 in the jth position, we take all possible matrices $H_k, k = 1 \ldots K$ whose columns are combinations of $f_i, i = 1, \ldots, F$ and $e_j, j = 1, \ldots, q$ such that

   (a) The number of columns of H_k is smaller than or equal to $q+1-r$ (this ensures that each of these matrices will put at least one over-identifying restriction).

   (b) H_k is of full column rank

   (c) $sp(H_k) \neq sp(H_l)$ for $k \neq l$ (because they would represent the same restriction).

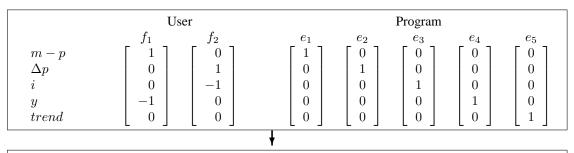[3]For convenience we have assumed 4 variables and at least a constant outside the cointegration space

(d) If the system contains both stochastic and deterministic variables, then that part of H_k which premultiplies the stochastic variables contains at least one non-null element. We thereby avoid testing whether there exists a stationary combination among the deterministics variables only.

3. For $k = 1 \ldots K$ test whether $\beta = (H_k\varphi_1, \psi)$ is accepted or not. Reject if the p-value $p(k)$ is smaller than 1%. Define $C_1 = \{1, \ldots, i, j, \ldots c_1\}$ as the ordered set of accepted restrictions, that is $p(i) > p(j)$ if $i < j$.

4. For every combination $(\{i, j\}, i < j \in C_1)$ test whether $\beta = (H_i\varphi_1, H_j\varphi_2, \psi)$ is accepted or not at the 1% level. Define $C_2 = \{\{i, j\}_l, l = 1, \ldots, c_2\}$ as the set of restrictions accepted.

5. For every combination $(\{i, j, k\}, \{i, j\} \in C_2, k \in C_1, i < j < k)$ test whether $\beta = (H_i\varphi_1, H_j\varphi_2, H_k\varphi_3, \psi)$ is accepted or not at the 1% level. Define $C_3 = \{\{i, j, k\}_l, l = 1, \ldots, c_3\}$ as the set of restrictions accepted.

6. Repeat step 5 until C_r or until one of the C's is an empty set.

7. Select the final model in the following way

   (a) Among the families of sets $B = \{C_1, C_2 \ldots C_r\}$ select all the models, which are accepted at the 5% level, which defines the set B_2.

   (b) Order all the models in B_2 according to the number of accepted restrictions, starting with the one with the highest number of accepted restrictions. In case of parity, rank the one with the highest p-value first.

3.1   Remarks on the algorithm

A number of comments on the algorithm will clarify certain choices inside the procedure and replication of the experiment.

- Every time we take a combination of restrictions (as in step 4,5 and 6), we have to check whether the restrictions are generically identifying. This is done by checking the condition in Johansen (1995a) . If the restrictions implied by say H_1 and H_2 say are not generically identifying, then we can eliminate one column of H_1 or H_2 and still have the same model, see Omtzigt (2002) . As an example consider that in step 3, we have accepted both $H_1 = [e_1, e_2]$ and $H_2 = [e_2]$. Then testing $\beta = (H_1\varphi_1, H_2\varphi_2)$ is equivalent to testing $\beta = (H_3\varphi_1, H_2\varphi_2)$, where $H_3 = [e_1]$. In the algorithm above, this combination is only tested if both H_1 or H_3 are accepted in step 3.

| | User | | | Program | | | |
	f_1	f_2	e_1	e_2	e_3	e_4	e_5
$m-p$	1	0	1	0	0	0	0
Δp	0	1	0	1	0	0	0
i	0	-1	0	0	1	0	0
y	-1	0	0	0	0	1	0
$trend$	0	0	0	0	0	0	1

Take all combinations of vectors $\{f_1, f_2, e_1, \ldots, e_5\}$ to obtain $H_1, \ldots, H_K$
• Max $p - r + 1$ vectors
• No overlap, such that if $sp\,(H_i) = sp\,(H_j)$, then only include one of them:
here include only one of $[e_1, e_4]$, $[f_1, e_1]$ and $[f_1, e_4]$
• Exclude possibility that deterministic term is cointegration relations: exclude $[e_5]$
• H_i must have full column rank: exclude for instance $[e_1, e_4, f_1]$

Test $\beta_i = [H_i \varphi, \psi]$ for $i = 1, \ldots, k$
Accept if p-value ≥ 0.01
Create ordered (highest p-value first) set of accepted restrictions $C_1 = \{1, \ldots, c_1\}$

Test $\beta_i = [H_i \varphi, H_j \varphi, \psi]$ for $i \in C_1, j \in C_1, i < j$ and only if $\{H_i, H_j\}$
generically identify a bi-dimensional space (use test Johansen 1995a)
Accept if p-value ≥ 0.01
Create ordered (highest p-value first) set of accepted restrictions $C_2 = \{\{i,j\}_l\, l = 1, \ldots, c_2\}$

Test $\beta_i = [H_i \varphi, H_j \varphi, H_k \varphi, \psi]$ for $\{i, j\} \in C_2, k \in C_1, k > \max(i, j)$ and only if $\{H_i, H_j, H_k\}$
generically identify a tridimensional space (use test Johansen 1995a)
Accept if p-value ≥ 0.01
Create ordered (highest p-value first) set of accepted restrictions $C_3 = \{\{i,j\}_l\, l = 1, \ldots, c_3\}$

....

Test $\beta_i = [H_i \varphi, H_{r-1} \varphi, \ldots, H_r \varphi]$ for $\{i, j, \ldots, r-1\} \in C_{r-1}, r \in C_1, r > \max(i, j, \ldots, r-1)$
and only if $\{H_i, H_j, \ldots, H_r\}$ generically identify an r-dimensional space
Accept if p-value ≥ 0.01
Create ordered (highest p-value first) set of accepted restrictions $C_r = \{\{i, j, \ldots, r\}_l\, l = 1, \ldots, c_r\}$

Combine family of sets $B_1 = \{C_1, \ldots, C_r\}$ select all models, which are accepted at 5% level: B_5
Rank models in B_5 according to the following criteria:
• highest number of overidentifying restrictions
• in case of parity highest p-value first

Figure 2: Algorithm for automatically restricting the cointegration space

- There is often more than one way to impose the same restrictions. If we define $f_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}$ and in the DGP the fist two variables are stationary and the other two are not (and not cointegrated), then we may accept three different combinations of H's:$\{[f_1],[e_1]\},\{[f_1],[e_2]\}$ and $\{[e_1],[e_2]\}$. In the final step one should check whether models are equivalent.

  Furthermore if one of the restrictions is rejected at an intermediate stage (say $[e_1]$), then the other restrictions could still be accepted (say both $[f_1]$ and $[e_2]$) such that it is still possible to recover at least one equivalent model. Thus the existence of equivalent models can increase the chance of accepting the DGP as the preferred model.

It is possible to prove the following theorem on the asymptotic size and power of the complete procedure

Theorem 1 *The asymptotic size of the algorithm is smaller than or equal to* $0.03 + 0.02r$, *while the asymptotic power is* 1.

The last theorem sets an upper limit to the size of the test procedure. It is very likely that the true size is considerably lower for a number of reasons:

1. The results of the individual tests in the procedure are most likely positively correlated.

2. There may be more than one set of H's which identifies the model, as was seen in one of the previous comments: just one of them needs to be accepted.

3. In 5, we took $(\{i,j,k\}, \{i,j\} \in C_2, k \in C_1, i < j < k)$ on purpose. If i,j and k are the three are the three restrictions of the DGP, which we want to recover, then we really want to arrive at the last step, where we test them jointly. To maximize this chance and minimize the number of calculations (taking all possible triples would take too much time, as would testing all couples and then adding any restriction k, not just $i < j < k$), we test the combination of i and j, the two restrictions, which were most easily accepted by the data, first. Conditional on that combination being accepted, we add restriction k, the restriction with the lowest p-value of the three when tested individually.

   We can in fact proof the following theorem, which is not operational, as the number of restrictions is not known a priori, but shows that in special cases the asymptotic size of the algorithm is in fact 5%

Theorem 2 *In the special case that only one relation contains over-identifying restrictions and in the special case that only two relations contain only 1 over-identifying restriction each, the asymptotic size of the algorithm equals* 0.05

In fact it is possible to lower the intermediate rejection probability, such that the asymptotic size of the test procedure becomes 5%:

Theorem 3 *If the intermediate values at steps (3) and (5) of the algorithm the critical value is set at $\kappa = \chi^2_{0.95}(qr - r^2)$ the asymptotic size of the algorithm is 0.05 while the asymptotic power is* 1.

In the following we shall not use theorem 3, because the critical value κ can become so large in practice, that virtually none of the tested restrictions are rejected at the intermediate levels. Consequently the computational costs explodes and only a small benefit is obtained in size if is it used.

3.2 An example

Let's consider a trivariate DGP with a trend inside the cointegration space. We have rightly determined the cointegration rank to be two and try to identify the parameter space. In the DGP we have that the first variable is trend-stationary, whereas the second one is stationary and the third one non-stationary, as in

$$\beta' = \left[\begin{array}{cccc} 1 & 0 & 0 & 0.023 \\ 0 & 1 & 0 & 0 \end{array} \right]$$

1. Based on economic theory, the following two vectors are indicated as possible combinations in a stationary relationship:

$$f'_1 = \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \end{array} \right]$$
$$f'_2 = \left[\begin{array}{cccc} 0 & 1 & -1 & 0 \end{array} \right]$$

2. As possible matrices H we take all matrices consisting of one column, choosing from the set $\{f_1, f_2, e_1, e_2, e_3, e_4\}$ and any combination of two columns with the following exceptions:

   (a) We do not test e_4 on its own, because then we would test whether a linear trend in stationary by itself

   (b) We test only one out of the combinations $\{[e_1, e_2], [f_1, e_1], [f_1, e_2]\}$ as they are equivalent restrictions. We choose the first combination. Similarly we only take one of the following set $\{[e_2, e_3], [f_2, e_2], [f_2, e_3]\}$

   In total we thus have 16 matrices H which we test individually in the next step

3. We find that the following restrictions are accepted. The p-values are between brackets:

$$\begin{array}{llll}
C_{1,1} & H_1 := \left[\begin{array}{cc} e_1 & e_2 \end{array} \right] & (0.80) & \qquad C_{1,4} \quad H_4 := \left[\begin{array}{cc} e_1 & f_2 \end{array} \right] \quad (0.25) \\
C_{1,2} & H_2 := \left[\begin{array}{cc} e_2 & e_3 \end{array} \right] & (0.40) & \qquad C_{1,5} \quad H_5 := \left[\begin{array}{c} e_2 \end{array} \right] \qquad\;\;\; (0.06) \\
C_{1,3} & H_3 := \left[\begin{array}{cc} e_2 & e_4 \end{array} \right] & (0.35) & \qquad C_{1,6} \quad H_6 := \left[\begin{array}{cc} f_1 & e_4 \end{array} \right] \quad (0.04)
\end{array}$$

We note that H_4 has erroneously been accepted, whereas the hypothesis $H_x := \left[\begin{array}{cc} e_1 & e_4 \end{array} \right]$ has been wrongly rejected. Its p-value was lower than 0.01.

4. There are 15 potential combinations to be tested. However we do not test $\{H_1, H_5\}$ as this combination does not satisfy the conditions for generic identification in Johansen (1995a) : it is easily seen that an equivalent identified model would be $\{[e_1], H_5\}$. The same comment applies to the combinations $\{H_2, H_5\}$ and $\{H_3, H_5\}$

We find that the following combinations are accepted:

$C_{2,1}$	$\{H_1, H_3\}$	(0.52)	$C_{2,5}$	$\{H_2, H_4\}$	(0.23)
$C_{2,2}$	$\{H_1, H_6\}$	(0.52)	$C_{2,6}$	$\{H_5, H_6\}$	(0.05)
$C_{2,3}$	$\{H_3, H_6\}$	(0.52)	$C_{2,7}$	$\{H_2, H_4\}$	(0.03)
$C_{2,4}$	$\{H_2, H_6\}$	(0.36)	$C_{2,8}$	$\{H_1, H_4\}$	(0.03)

We note that the last two have been erroneously accepted, but that no error of type I was made in this round.

(7) We take $C_1, \ldots, C_5$ and $C_{2,1}, \ldots, C_{2,6}$ as they were accepted at 5% or more. First we note that $C_{2,1}, C_{2,2}$ and $C_{2,3}$ share the same p-value. It turns out that the models are equivalent, such that they are the same model. We give the first five models according to our selection criteria:

$Model$	d.o.f.	$p-value$
$C_{2,6}$	3	(0.05)
$C_{2,1}, C_{2,2}, C_{2,3}$	2	(0.52)
$C_{2,4}$	2	(0.36)
$C_{2,5}$	2	(0.23)
$C_{1,5}$	2	(0.06)

Closer inspection shows that model $C_{2,6}$ is an equivalent model to the DGP: Even though H_x was rejected at step 3, we find an equivalent representation. The next four models are accepted all impose 2 out of the 3 restrictions we could have found. This is also a relative success as the restrictions we found actually hold. We just failed to get the last one in those cases.

4 Monte Carlo evidence

We used the following 5-variable DGP with k lags and a trend in the cointegration relation to test the algorithm:

$$\prod_{i=1}^{k} (1 - \phi_i L) X_{1t} = \alpha_{11} (X_{2t} + X_{3t} + 0.02t) + \varepsilon_{1t}$$

$$\prod_{i=1}^{k} (1 - \phi_i L) X_{3t} = \alpha_{32} (X_{4t} + X_{5t}) + \varepsilon_{3t}$$

$$\Delta X_{2t} = \varepsilon_{2t}$$
$$\Delta X_{4t} = \varepsilon_{4t}$$
$$\Delta X_{5t} = \varepsilon_{5t}$$

where

$$\alpha_{11} = \alpha_{32} = \prod_{i=1}^{k} (1 - \phi_i)$$

$$\varepsilon_t \sim iidN\left(0, I_5 \times 10^{-4}\right)$$

There are two cointegration relations in the DGP and the roots of each stationary equations are equal to ϕ_1 to ϕ_k. The main reason for proposing this DGP is that when adding a lag, all the existent roots can be kept constant.

The cointegration space is equal for any number of lags:

$$\beta' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0.02 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

In the experiments that follow we vary three parts of the procedure:

1. The series of roots $\{\phi_i\}_{i=1}^{k}$. In all but one case we model the right number of lags, but in one case we model only 2 lags, whereas the DGP contains three lags.

2. The economic theory input: either we do not specify any f_i vectors, in which case we should be able to find the 5 zeros in the DGP, which is equivalent to 3 restrictions or we specify only one, namely:

$$f_1' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

   In the last case recovering the DGP is equivalent to finding five over-identifying restrictions.

3. The length of the time series: we take $T = 100$ as our bench mark case, but have one experiment with $T = 1000$ to check the asymptotic behaviour of the procedure.

The success of the procedure is measured in two different ways:

1. Taking the first model, we look whether

   (a) the restrictions of the model exactly identify the DGP

   (b) the model found is a submodel of the DGP (that it all the restrictions of the model are present in the DGP), but the model misses one restriction

   (c) the model found is a submodel of the DGP, but it misses two restrictions

   (d) the model found has all the over-identifying restrictions of the DGP plus one restriction too many.

   (e) the model found has all the over-identifying restrictions plus two.

   (f) a completely different model is selected

12

(g) no model at all is selected (as all of them are rejected)

a is an outright success, but especially b can also be classified as a success as the DGP is nearly recovered. f is complete failure.

2. As in the example above, it is possible to give the first five selected model to the researcher and let her make the choice. We thus measure how often the model, which exactly identifies the DGP is among the those first five.

The degree of identification is conditional on the f-vectors specified. In our particular example three overidentifying restrictions can be found, if no f-vectors are specified, whereas 5 restrictions can be found if the $f-$vector (4) is given as input. In the first case recovering the DGP is equivalent to accepting $\{[e_1, e_2, e_3, e_6], [e_3, e_4, e_5]\}$, whereas in the second case it is $\{[f_1, e_6], [e_3, e_4, e_5]\}$.

In the last column of the table we report one minus the rejection probability of a straight test of these last models, when testing at the nominal 5% level. These numbers give an indication of the small sample performance of these tests and also provide an upper bound to how often the DGP can be among the first five models or indeed how often the DGP can be the first selected model.

The results of the Monte Carlo simulations are based on 1000 replications. For each replication the first 100 observations were discarded. The results are reported in table 1.

In our benchmark case, number 1, the systems contains two residual roots of 0.6 and two roots of 0.2. We model the right number of lags, namely two and do not provide any input in the form of an f-vector. In 53.6% of the cases the proposed algorithm is able to recover the restrictions in the cointegration space, whereas in 7.8% it only misses one. If we consider the first five selected model, then the DGP is among them in fully 72.8% of the cases. This is a high success rate when one considers than the estimated size of testing just all the right restrictions is 19.2% (see the last column of the table) for a nominally 5% test. 80.8% is an upper limit to how often the selected model can be recovered (column a) or how often it can be among the first five.

The second and third DGP show that increasing the roots of the process leads to a complete breakdown of the algorithm. On the other hand DGP's with one lag (number 4 and 5) have a relatively good performance. In the case of number 4, in fully 98% of cases do we either recoup the original DGP or miss just one restriction.

It is fairly common to select a low lag length even if it is believed that the true lag length of the DGP is longer, possibly even infinite: to our knowledge no asymptotic results are known for the consistency of the estimated cointegration parameters in this case.(obviously the short run parameters are not estimated unbiased any more in that case) In experiment 6 and 7 we check the effect the effect of underselecting the true lag length and remarkably the algorithm does better by all measures when the lag length is underestimated. This lends support to the view that a short model should be fitted to the data.

In experiment 8 we give some quasi-economic input in that we specify the vector $f_1' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$.

	DGP $\{\phi_i\}$	Lags Mod	T	f ?	The first model a	b	c	d	e	f	g	DGP in 5?	1-Size
1	$\{0.6, 0.2\}$	2	100	no	53.6	7.8	0.6	11.2	2.0	24.6		72.8	80.8
2	$\{0.8, 0.2\}$	2	100	no	5.8	0.8	0.2	9.4	10.2	73.6		19.6	72.2
3	$\{0.8, 0.6\}$	2	100	no	0.4	0	0	1.4	6.6	91.6		2.4	56.4
4	$\{0.6\}$	1	100	no	91.0	7.0	1.2	0.4	0	0.4		91.6	91.6
5	$\{0.8\}$	1	100	no	44.8	4.2	0.2	13.4	3.4	34.0		66.4	83.0
6	$\{0.6, 0.2, 0.2\}$	3	100	no	27.4	6.4	0.6	9.0	4.6	51.8		42.6	68.0
7	$\{0.6, 0.2, 0.2\}$	2	100	no	37.8	5.2	0.2	13.4	5.2	38.2		61.4	82.0
8	$\{0.6, 0.2\}$	2	100	yes	65.0	15.0	1.4	3.2	0.6	14.8		74.6	76.2
9	$\{0.6, 0.2\}$	2	1000	no	93.8	5.4	0.8	0	0	0		93.8	94.2

Table 1: Monte Carlo study of automated model selection

The results should be compared to those of experiment 1: even though the size distortion of this test is large (23.8%!), the true restrictions (5 in this case) or all but one of them are recovered in fully 80% of all cases. Furthermore the model with the true restrictions is among the first five selected in 74.6% of cases and thus almost reaches its upper bound.

Experiment 9 shows that the procedure works asymptotically. The fact that the true model is among the first five in only 93.8% of cases, compared to an upper bound of 94.2% is caused by the fact that one of the two vectors individually can be rejected at the 1% level, whereas the combination of the two restrictions can still be accepted at the 5% level. This happened in our example above, but in that case there were equivalent models, whereas there are none in the Monte Carlo study.

4.1 An empirical example

We take the data on money demand in the UK as analyzed by Hendry and Doornik (1994) and many others. The data consist of log of real output (TFE), money (M1), inflation and an interest rate differential. Furthermore there are two dummies: one for output shocks (Dout) and one for oil shocks (Doil). For full details of the variables and the data the reader is referred to the original paper. The documented data and the programs to run the original analysis are available at the web site of David Hendry[4]. The cointegration space they find is the first one reported in table 2.

After formal testing the authors decide that the model should have 2 lags and 2 cointegration relations. They then decide that the oil dummy is outside the cointegration space and the output dummy inside. We shall put them both inside the space and thereby test their possible exclusion. Their reported test is a combined restriction of the short run matrix (four zero restrictions), a restriction that the second vector is completely specified

---

[4]http://www.nuff.ox.ac.uk/users/hendry/

	H&D		Ex1, M1		Ex 2, M5		Ex 3, M1	
output	-1	1	-1	1	-1	1	-1	1
money	1	0	0.89	0	1	0	1	0
inflation	6.91	-3.4	7.23	-1.98	6.90	-1.86	7.12	-1.90
Rnet	6.91	1.8	6.39	1.34	6.90	1.36	7.12	1.31
trend/100	0	-0.63	0	-0.65	0	-0.66	0	-0.67
Dout	1.46	-0.40	0	-0.26	0	-0.27	0	-0.18
Doil	n/a	n/a	0	0	0	0	0	0
p-value	0.38^5		0.65		0.48		0.30	

Table 2: Automated model selection in UK money demand

and the following H-matrix on the first vector:

$$H_1' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We leave the second vector completely unrestricted throughout and in our first run of the algorithm we impose no $f-$vectors, that is we ran the algorithm without any theory input. The first model, the algorithm selects is reported in table 2.

In the second run we use the theory input from Hendry and Doornik by specifying the following two vectors:

$$f_1' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$f_2' = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

and feeding them into the algorithm. In the table we report only the fifth model. The first four models all have p-values of between 5% and 6%. They have different second cointegration vectors, which consists of changing combinations of inflation, output dummy and trend. In view of the outcome of the previous run of the algorithm (and the next run), the considerably higher p-value and the economic interpretability of the vector as an demand equation, we report model 5.

In the third run we still have the two f-vectors, two cointegration vectors, but choose only one lag, as the previously reported simulation suggested this may help. We report the first model selected by the algorithm.

We thus found a model with one more over-identifying restriction, namely that the output dummy should be zero in first cointegration vector. The total analysis has only taken a few minutes on a Pentium-II computer.

5 Use of the algorithm

5.1 Practical advantages

The algorithm can be used to simply check whether the final result obtained by the traditional way of finding an identified cointegration space is good. There are however more promising uses: the procedure can be directly applied by the practitioner or referee to select the final model.

But the speed of the algorithm (in most cases the answer is given in a few minutes) allows for more possibilities. Sometimes the outcome of the lag length selection procedure or the rank-test is unclear and a choice has to be made. With the new procedure it is possible to find an identified model for both possibilities. Sensitivity check procedures of all kinds are now possible: if the inclusion of an extra dummy leads to a completely different model, then a problem of stability certainly exists in the data.

5.2 Methodological advantages

The presently used procedure to find identifying restrictions is long and cumbersome and many rules of the thumb have to be applied: not every combination of two vectors, which are accepted can possibly be tested in a reasonable time. Replication, even by the very same researcher who did the original study, is often difficult. With a standard, thorough procedure, like the here proposed algorithm, replication becomes possible.

The LSE-methodology itself is often attacked for being partially 'art' or worse, 'alchemy'. By clearly spelling out the modelling process, formalizing and automating it, the methodology itself is being strengthened. The number of key decisions in the identifying/restriction process is brought down to three:

1. What economic theory should be tested in the form of f-vectors?

2. Which of the top models should be chosen as the preferred model?

3. Which sensitivity checks should be done?

All three decisions can be discussed and reported in a paper, whereas it is currently often impossible to report all tests on the cointegrated space done in the modelling process.

Once more the dichotomy between on the one hand selecting the right variables, being explicit in each step which economic theory is being used and selecting the final model and on the other hand the testing procedures is stressed. The second part can be automated (and should be) to leave more time and space to do the first carefully.

5.3 Improving the methods themselves

The present algorithm is but a first proposal on how to automate a part of the modelling process. It can no doubt be improved, both in terms of computational speed and - far

more importantly - in terms of internal decision rules. One possibility is to use sub-samples and sub-sets of the variables in the algorithm to find the building blocks for the total cointegration space.

The small Monte-Carlo study has already pointed to two possible improvements in the methods: keeping in restrictions, which are accepted at the 1% level, but rejected at the 5% level and selecting fewer lags than are believed present in the DGP. The first rule can be applied in automated modelling only, but the second is certainly relevant for traditional modelling as well.

The very poor actual size performance of the applied tests is of great concern in the development of these methods. The only theorem underlying the algorithm relies on the asymptotic size. Omtzigt and Fachin (2002) clearly show that the currently available methods Bartlett corrections and Bootstrapping, do not offer satisfactory solutions, so new ones will have to improve the algorithm. The size of this problem had been somewhat hidden in the literature due to the reliance on small DGP, whose parameters can easily be controlled. The large number of parameters in the Monte Carlo DGP used here means that no effective exploration of the size in the whole parameter space could be executed. Yet this particular example shows that at least in some cases size distortion is an extremely serious problem.

Finally no evidence is yet available about how likely the modelling procedure is able to recover the original DGP. We show it only for part of the process (disregarding rank and lag selection for instance). Of course it is unlikely that the DGP falls within the class of models considered, but still measures on how well the modelling process does when it is close should be available.

5.4   Comparison with Minimal

Minimal by Davidson (1998a) follows a different approach: it tries to find all the smallest subsets of variables which are cointegrating. It does so by using the Wald test statistics for (2) as developed by the Davidson (1998b) . His procedure is certainly faster, as the Wald test does not require restricted optimization. Yet the method proposed carries the following advantages:

1. It allows for theory input in the form of f-vectors.

2. The combinations of restrictions are tested in the algorithm presented, whereas in minimal the accepted cointegration vectors are just combined

3. In minimal, the combination of accepted vectors can and does lead to conflicting evidence: it is perfectly possibly for minmax to accept $[e_1, e_2]$ and $[e_1, e_3]$ but reject $[e_2, e_3]$. Thus the space spanned by the cointegration vectors is not unambiguously defined.

The speed of the algorithm remains a cause for concern. With 2 cointegration vectors and 5 variables minimal or step 3 will consists of doing (a maximum of) $W = \sum_{i=1}^{3} \binom{5}{i} =$

25 tests. With 8 variables (either because variables are added or f-vectors defined), there are up to 246 tests. If many of them are accepted, the number of calculations explodes in the step 4 of the algorithm: with $r = 2$, there could be up to $W(W-1)/2$ combinations in that step. This will be a major concern in future development of the algorithm.

6 Conclusions

We have presented an algorithm for the automatic identification and restriction of the parameter space of a cointegrated VAR. Its used was demonstrated both by means of a Monte Carlo study and an empirical example. One remarkable solution of the former is that underselection of the VAR lag length may lead to a higher probability of recovering the original cointegration space.

Throughout we have argued the methodological advantages of automation: separation of economic theory decision, which will have to be performed by the researcher and search procedures, which can be automated.

References

Brüggeman, R. and H. Lütkepohl (2000). Lag selection in subset VAR models with an application to a U.S. monetary system. Technical report, Humbold Universität, Berlin.

Davidson, J. (1998a). Structural relations, cointegration and identification: Some simple results and their applications. *Journal of Econometrics 87*, 87–113.

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

Doornik, J. and D. Hendry (1997). *Modelling Dynamic Systems Using PcFiml 9 for Windows*. Timberlake, London.

Engle, R. and C. Granger (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica 55*, 251–276.

Hansen, H. and K. Juselius (1994). *CATS for RATS4: Manual to Coinegration Analysis to Time Series*. Estima.

Hendry, D. and J. Doornik (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy 41*, 1–33.

Hendry, D. and H. Krolzig (2003). New developments in automatic general-to-specific modelling. In B. Stigum (Ed.), *Econometrics and the Philosophy of Economics*. MIT Press.

Hoover, K. and S. Perez (1999). Data mining reconsidered: Encompassing and the general -to-specific approach to specification search. *Econometrics Journal 2*, 167–191.

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 processes. *Econometrica 59*, 1551–1580.

Johansen, S. (1995a). Identifying restrictions of linear equations: With applications to simultaneous equation and cointegration. *Journal of Econometrics 69*, 111–132.

Johansen, S. (1995b). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Juselius, K. (2002). *The Cointegrated VAR Model: Econometric Methodology and Macroeconomic Applications*. Unpublished manuscript.

Kongsted, H. (1998). An I(2) cointegration analysis of small country import price determination. Technical Report 98-22, Institute of Economics, University of Copenhagen.

Kongsted, H. (1999). Testing the nominal-to-real transformation. Technical report, Institute of Economics, University of Copenhagen.

Lovell, M. (1983). Data mining. *The Review of Economics and Statistics 65*, 1–12.

Nielsen, B. and A. Rahbek (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics 62*(1), 5–22.

Omtzigt, P. (2002). Automatic identification of simultaneous equations models. Technical Report 2002/01, University of Insubria, Varese, Italy.

Omtzigt, P. and S. Fachin (2002). Bootstrapping and Bartlett corrections in the cointegrated VAR model. Technical report.

Paruolo, P. (2002). LR cointegration tests when some cointegrating relations are known. *Journal of the Italian Statistics Society forthcoming*.

Appendix: Proofs

Proof of Theorem 1. If $r = 1$ we test all possible restrictions $\{H_i\}$. They fall in three categories:

1. $\{H_i\}_t$ The true DGP (there could be more equivalent reprentations) This is accepted at the asymptotic 5% level. It has the maximum (true) number of overidentifying restrictions

2. $\{H_i\}_s$ The restrictions hold true, but are less in number than in $\{H_i\}_t$. This means that if the model is accepted, it is classified below $\{H_i\}_t$ (if that last one is accepted)

3. $\{H_i\}_f$ The restrictions do not hold true. The asymptotic power of the test is 1 (Johansen 1991):$L\left(\{H_i\}_f\right) \to \infty$ as $t \to \infty$ such that this test in always rejected asymptotically.

   The result for $r = 1$ follows.

For $r > 1$ consider what happens to $\{H_1, \ldots, H_r\}$ the DGP combination. In step 3, we test all possible restrictions of the kind (2) and so each of the DGP restrictions is tested individually at the 1% level, that is accepted if $L(H_i) \leq c_{0.99}$ where c_r is the 99%th percentile of the χ^2 distribution. In step 4, we test whether $L(H_1, H_2) \leq c_{0.99}$ whereas in step 5 we test $L(H_1, H_2, H_3) \leq c_{0.99}$ until $L(H_1, H_2, \ldots, H_{r-1}) \leq c_{0.99}$. In step 7, we then test whether $L(H_1, H_2, \ldots, H_r) \leq c_{0.95}$. To find the asymptotic size of this procedure, we use Bonferroni inequality:

$$P\left(\cap_{i=1}^{r}\left(L(H_i) \leq c_{0.99}\right) \cap \cap_{i=2}^{r}\left(L(H_1, \ldots, H_i) \leq c_{0.99}\right) \cap L(H_1, H_2, \ldots, H_r) \leq c_{0.95}\right)$$

$$\geq 1 - \sum_{i=1}^{r} P\left(L(H_i) \geq c_{0.99}\right) - \sum_{i=2}^{r-1} P\left(L(H_1, \ldots, H_i) \geq c_{0.99}\right) - P\left(L(H_1, H_2, \ldots, H_r) \geq c_{0.95}\right)$$

$$= 1 - 0.01r - 0.01(r - 2) - 0.05$$

$$= 1 - 0.02r + 0.03$$

The comments of point 2 and 3 still hold true in this case. $\blacksquare$

Proof of Theorem 2. If only one relation contains overidentifying restrictions, the DGP can be written as

$$\beta = (H_x \phi, \psi)$$

and this is tested at round 3 and included in set C_1 if accepted. Then in step 7 it is tested again, such that we have:

$$P\left((L(H_x) \leq c_{0.99}) \cap (L(H_x) \leq c_{0.95})\right)$$
$$= P\left(L(H_x) \leq c_{0.95}\right)$$
$$= 0.95$$

If two relations contain exactly overidentifying relations, then we have the following DGP:

$$\beta = (H_x \phi_1, H_z \phi_2, \psi)$$

and we find

$$P\left((L(H_x) \leq c_{0.99}) \cap (L(H_y) \leq c_{0.99}) \cap (L(H_x, H_y) \leq c_{0.99}) \cap (L(H_x, H_y) \leq c_{0.95})\right)$$
$$= P\left((L(H_x) \leq 6.63) \cap (L(H_y) \leq 6.63) \cap (L(H_x, H_y) \leq 5.99)\right)$$
$$= P\left(L(H_x, H_y) \leq 5.99\right)$$
$$= 0.95$$

where we have used that $L(H_y) \leq L(H_x, H_y)$ and $L(H_x) \leq L(H_x, H_y)$. $\blacksquare$

Proof of Theorem 3. The proof proceeds as the proof of theorem 1 until

$$P\left(\cap_{i=1}^{r}\left(L\left(H_{i}\right)\leq\kappa\right)\cap\cap_{i=2}^{r}\left(L\left(H_{1},\ldots,H_{i}\right)\leq\kappa\right)\cap L\left(H_{1},H_{2,},\ldots,H_{r}\right)\leq c_{0.95}\right)$$
$$= P\left(L\left(H_{1},H_{2,},\ldots,H_{r}\right)\leq c_{0.95}\right)$$
$$= 0.95$$

where in the second line we have used the fact that all hypotheses $\{H_{i}\}, i = 1, \ldots, r$ and $\{(H_{1},\ldots,H_{i})\}, i = 1, \ldots, r$ are subhypotheses of $\{(H_{1}, H_{2,}, \ldots, H_{r})\}$ such that $L\left(H_{i}\right) \leq L\left(H_{1},H_{2,},\ldots,H_{r}\right), i = 1, \ldots, r$ and $L\left(H_{1},\ldots,H_{i}\right) \leq L\left(H_{1},H_{2,},\ldots,H_{r}\right), i = 1, \ldots, r$. Furthermore the maximum number of overidentifying restrictions on the space is $qr - r^{2}$ (at which point each matrix H_{i} has only one column), such that $\kappa \geq c_{0.95}$. ∎