COMMON TRENDS AND COMMON CYCLES IN LATIN AMERICA: A 2-STEP VS AN ITERATIVE APPROACH

Alain Hecq *
University of Maastricht,
Department of Quantitative Economics

October 8, 2004

Abstract

We are interested in determining the number of common trends and common cycles in the output of a set of Latin American countries. In order to work with homogeneous and reasonably good series however, we should rely on annual data. Consequently, the number of variables is relatively large compared to the number of observations to blindly trust the asymptotics. For several years, the panel data literature proposes tools to tackle this problem, mainly for the study of long-run co-movements. We take another road here and we test for cointegration and common cyclical features in a time series framework using an iterative strategy that maximizes the likelihood function by successively imposing long and short-run restrictions until convergence is achieved. Monte Carlo simulations stress advantages of this approach over the two-step one. In practice however, the cost of implementing this more complicated procedure must be evaluated with the expected benefits. Overall, simple adjustments for the degrees of freedom and the use of information criteria are helpful "cheap" alternatives.

Keywords: Common trends, Common cycles, Monte Carlo.

JEL Classification: C32

*Alain Hecq, University of Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: a.hecq@ke.unimaas.nl, homepage: www.personeel.unimaas.nl/a.hecq.
1 Introduction

Cointegration techniques are now routinely applied by economists to extract $r$ meaningful long-run relationships among a set of $n$ non stationary time series $y_t = (y_{1t}, \ldots, y_{nt})'$. Interpreting the issue in its dual form, this means there also exist $n-r$ common trends, hence only $n-r$ common permanent shocks driving the economy. To evaluate the presence of such long-run co-movements, the Johansen maximum likelihood approach (1995 \textit{inter alia}) based on Anderson’s results (e.g. 1984), is widely used. In practice, after choosing the relevant series one needs (i) to select $p$, the lag length of the finite order VAR generating the multivariate process $y_t$, (ii) to find the deterministic terms, (iii) to determine the cointegrating rank and quite often (iv) to restrict the cointegrating space. The literature has shown that none of these steps is trivial.

Next, the non stationary VAR is rewritten in its VECM representation, ready for the discussion about its economic content, for the study of short and long-run causality, for the analysis of the reduced number of shocks, for forecasting or for impulse response analyses.

Beyond cointegration, a number of papers have concentrated on modelling the common serial correlation feature among stationary time series. Indeed, alike cointegration is associated with long-run relationships, common dynamics is a sign of co-movements in the short-run. These common sources of transmission mechanisms, namely the so called common cycles, allow to extract common transitory shocks that can often be linked to business cycle co-movements.\footnote{At least this is the case for the multivariate Beveridge-Nelson decomposition. Indeed, the reduced number of transitory shocks is, in the Gonzalo-Granger (1995) decomposition, implied by the presence of cointegration. However it can be shown that this is due to the fact that the Gonzalo-Granger decomposition underlies the presence of common cycles (Proietti, 1997; Hecq et al., 2000).}

Additional advantages of considering these short-run restrictions are the large reduction of the number of parameters that need to be estimated and their role for forecasting (see Vahid and Issler, 2002 and Hecq, Palm and Urbain, 2004 on this latter issue).

This being said, this note stresses another advantage of considering short-run co-movements, namely the improvement of the small sample behavior of cointegration tests thanks to additional restrictions. Several studies comment on the poor performance of the asymptotic Johansen test in small samples.\footnote{See Ho and Sorensen (1996), Gonzalo and Pitarakis (1999) Cheung and Lai (1993), Söderlind and Vredin.} Most of the existing Monte Carlo simulations point out that the dimension
of the data series for a given sample size may pose a problem since the number of parameters of
the VAR models grows very large as the dimension increases. It emerges that likelihood ratio
tests are often too liberal which leads to overestimate $r$ in empirical works. Likewise difficulties
often arise when the lag length is either over or under parametrized. Small sample corrections
(à la Reinsel and Ahn, 1992, for instance) may be helpful but they often lead in practice to
underestimate $r$. Indeed, the distortion is non monotone with the number of variables (Ho
and Sorensen, 1996). Johansen (2002) has recently proposed a correction factor to circumvent
this problem. An alternative solution is to use a bootstrap procedure such as in Fachin, 2000
or Harris and Judge, 1998 inter alia. Fachin and Omtzigt (2002) compare these two latter
methods and find that if they do better than asymptotic tests but they still suffer from size
distortions. Concerning common dynamics, only a few Monte Carlo results are available (see
Beine and Hecq, 1998; Hecq, 1998; Candelon, Hecq and Verschoor, 2005). From these studies
it emerges that common feature test statistics are strongly sensitive to misspecification such as
the presence of seasonality, conditional heteroskedasticity, outliers, aggregation.

Obviously, as far as both long-run and short-run co-movements are of interest for the
researcher, a natural practice consists in first estimating superconsistently the cointegrating
vectors. In a subsequent step, it will be easy to perform a test for common serial correlation
by considering the cointegrating relationships as given (see Vahid and Engle, 1993). This is
what we call the 2-step approach. The drawback of the latter procedure is that a misleading
inference about the cointegrating rank in the first step may damage the subsequent analysis
about common cycles (Hecq, Palm and Urbain, 2004). To try to improve the 2-step approach,
this article evaluates whether an iterative procedure would be helpful both for cointegration
and common feature test statistics.

The structure of this paper is as follows. Section 2 recalls the definitions. Section 3
describes the test statistics. We show how to implement the approach that consists in switch-
ing between long-run and short-run restrictions. Section 4 summarizes Monte Carlo results.
Section 5 demonstrates the use of these tests in a study of co-movements among the gross

(1996) and Jacobson et al. (1998 ) inter alia.

2 Model representation and definitions

We consider $\Pi(L)y_t = \Theta D_t + \varepsilon_t$ the $n$-dimensional vector autoregressive model of order $p$ for the $I(1)$ variables $y_t = (y_{1t}, \ldots y_{nt})'$, i.e.

$$y_t = \Theta D_t + \Pi_1 y_{t-1} + \ldots + \Pi_p y_{t-p} + \varepsilon_t, \quad t = 1, \ldots, T, \quad (1)$$

for fixed values of $y_{-p+1}, \ldots, y_0$ and where $\Pi(L) = I_n - \Pi_1 L - \ldots - \Pi_p L^p$; $D_t$ is a vector of deterministic terms and disturbances $\varepsilon_t$ are $NIID(0, \Omega)$. Let us assume that $\text{rank}(\Pi(1)) = r, 0 < r < n$, so that $\Pi(1)$ can be expressed as $\Pi(1) = -\alpha \beta'$, with $\alpha$ and $\beta$ both $(n \times r)$ matrices of full column rank $r$ and that the characteristic equation $|\Pi(z)| = 0$ has $n - r$ roots equal to 1 and all other roots outside the unit circle. The process $y_t$ is then cointegrated of order $(1,1)$. The columns of $\beta$ span the space of cointegrating vectors, and the elements of $\alpha$ are the corresponding adjustment coefficients or factor loadings. Decomposing the matrix lag polynomial $\Pi(L) = \Pi(1)L + \Gamma(L)(1-L)$, and defining $\Delta = (1-L)$, we obtain the vector error correction model

$$\Delta y_t = \Theta D_t + \alpha \beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \quad t = 1, \ldots, T, \quad (2)$$

where $\Gamma_0 = I_n, \Gamma_j = -\sum_{k=j+1}^p \Pi_k$ ($j = 1, \ldots, p - 1$).

Serial correlation common feature (SCCF hereafter, see Engle and Kozicki, 1993) holds for the VECM in (2), if there exists a $(n \times s)$ matrix $\delta$, whose columns span the cofeature space, such that $\delta' (\Delta y_t - \Theta D_t) = \delta' \varepsilon_t$ is a $s$-dimensional vector mean innovation process with respect to the information available at time $t$. Consequently, SCCF arises if there exists a matrix $\delta$ such that the conditions $\delta' \Gamma_j = 0_{(s \times n)}, j = 1 \ldots p - 1$ and $\delta' \Pi(1) = -\delta' \alpha \beta' = 0_{(s \times n)}$ are jointly satisfied. Let us denote a $(n(p - 1) + r) \times 1$ vector $W_{t-1} = [\Delta y'_t, \ldots, \Delta y'_{t-p+1}, y'_{t-1} \beta']$ and a
\[ n \times (n(p - 1) + r) \text{ matrix } \Phi = [\Gamma_1, \ldots, \Gamma_{p-1}, \alpha], \text{ so that (2) is written as} \]

\[ \Delta y_t = \Theta D_t + \Phi W_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T. \quad (3) \]

Under SCCF, \( \Phi \) is of reduced rank \( n - s \) and can be written as \( \Phi = A[F_1, \ldots, F_{p-1}, F_p] = AF \), where \( A \) is \( n \times (n - s) \) full column rank matrix and \( F \) is \( (n - s) \times (n(p - 1) + r) \). \( \delta' AF W_{t-1} = 0 \) means that \( \delta \in sp(A_{\perp}) \) where \( A_{\perp} \) is the orthogonal complement\(^3\) of \( A \). Consequently, as pointed out by Vahid and Engle (1993), in a \( n \)-dimensional \( I(1) \) vector process \( y_t \) with \( r < n \) cointegrating vectors, if the elements of \( y_t \) have common cyclical features (given by \( f_t = FW_{t-1} \)) there can be at most \( n - r \) linearly independent cofeature vectors that eliminate the common cyclical features since the cofeature matrix must lie in \( sp(\alpha_{\perp}) \). SCCF implies that \( s \leq n - r \) (or \( r + s \leq n \)) and that the common dynamic factors \( f_t \) consist of linear combinations of the elements of \( W_{t-1} \).

This set of long and short-run restrictions gives rise to a full description of the common trends and cycles in \( y_t \). Indeed, from the Wold representation of a stationary process \( \Delta y_t \) and focussing on the Beveridge-Nelson decomposition (ignoring deterministic terms for simplicity) we have

\[ \Delta y_t = C(L)\varepsilon_t, \quad = C(1)\varepsilon_t + \Delta C^*(L)\varepsilon_t, \]

with \( C(L) = I_n + \sum_{i=1}^{\infty} C_i L^i \) and \( \sum_{j=1}^{\infty} j|C_j| < \infty \) and \( C^*_i = -\sum_{j>1} C_j \) for all \( i \). If we integrate both sides we obtain

\[ y_t = C(1) \sum_{j=1}^{l} \varepsilon_j + C^*(L)\varepsilon_t, \quad = \text{Trends} + \text{Cycles}. \]

\( ^3 \text{Space will be denoted by } sp. \text{ We shall always denote the orthogonal complement of any } n \times s \)-dimensional matrix \( B \), with \( n > s \) and \( \text{rank}(B) = s \), by the \( n \times (n - s) \) matrix \( B_{\perp} \) such that \( B' B_{\perp} = 0 \) with \( \text{rank}(B_{\perp}) = n - s \) and \( \text{rank}(B : B_{\perp}) = n). \)
These trends and cycles will be common to the series depending on the rank of the matrices $C(1)$ and $C^*(L)$. Under cointegration $\text{rank}(C(1)) = n - r$ and we know that the $n - r$ common stochastic trends $\alpha_\perp \sum_{j=1}^{t} \varepsilon_j$ are annihilated by $\beta$ because $\beta' C(1) = 0$. Similarly, under SCCF, $C^*(L)$ is of reduced rank $n - s$ and these $n - s$ common cycles are such that $\delta' C^*(L) = 0$ (see inter alia Issler and Vahid, 2001).

When cycles are not perfectly synchronized, alternative specifications have been proposed by Vahid and Engle (1997) and Cubadda and Hecq (2001). These two latter models have the advantage to have a nice interpretation in terms of delays of adjustment to shocks in the multivariate Beveridge-Nelson representation. However, alike SCCF they are sensitive to misspecification made on the determination of the cointegrating rank. This is the reason why we consider the weak form common cycle specification (WF hereafter, see Hecq, Palm and Urbain, 2000, 2004) which is more robust to the choice of $r$.\footnote{Although less easy to interpret in terms of Beveridge-Nelson cycles (but see Hecq et al., 2000 on that).} Indeed, we have seen that the determination of the cointegrating space in the first step (possibly misspecified) bounds the number of SCCF cofeature vectors. Instead, in the WF we have under the null $\delta'(\Delta y_t - \Theta D_t) = \delta'' y_{t-1} + \delta' \varepsilon_t$ where $\delta'' = \delta' \alpha$. We analogously define a $n(p - 1) \times 1$ vector $Z_{t-1} = [\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}]'$ and the $n \times n(p - 1)$ matrix $\Phi^* = [\Gamma_1, \ldots, \Gamma_{p-1}]$, so that (2) becomes

$$
\Delta y_t = \Theta D_t + \alpha \beta' y_{t-1} + \Phi^* Z_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T, \quad (4)
$$

If $\Phi^*$ is of reduced rank $n - s$ it can be written as $\Phi^* = A^*[F^*_1, \ldots, F^*_p] = A^* F^*$, where $A^*$ is $n \times (n - s)$ full column rank matrix and $F^*$ is $(n - s) \times n(p - 1)$ such that $\delta' A^* F^* Z_{t-1} = 0$. The matrix $\delta$ must lie in $sp(A^*_{\perp})$ but not necessarily in $sp(\alpha_{\perp})$ and consequently we can have $s > n - r$ weak form cofeature vectors.
3 Test statistics

3.1 The two-step approach

As a shortcut, the expression \( \text{CanCor}\{\Delta y_t, y_{t-1} | (D'_t, \Delta y'_{t-1}, \ldots \Delta y'_{t-p+1})' \} \) will summarize the reduced rank regression procedure used in the Johansen approach. That means that one extracts the squared canonical correlations between \( \Delta y_t \) and \( y_{t-1} \), both sets concentrated out the effect of deterministic terms and lags of \( \Delta y_t \). In order to test for the significance of the largest eigenvalues, one can rely on Johansen’s trace statistic (5) or on one of the modified version to account for small samples like in (6).

\[
LR_r = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i), \tag{5}
\]

\[
LR^c_{r} = -(T - np) \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i), \tag{6}
\]

where the eigenvalues \( 1 > \hat{\lambda}_1 > \ldots > \hat{\lambda}_n > 0 \) are the solution of \( |\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0 \). \( S_{ij}, i, j = 0, 1 \) are the second moment matrices of the residual \( R_{0t} \) and \( R_{1t} \) obtained in the multivariate least squares regressions from respectively \( \Delta y_t \) and \( y_{t-1} \) on \( (D'_t, \Delta y'_{t-1}, \ldots \Delta y'_{t-p+1})' \). Under the null, statistics follow a functional of Brownian motions and asymptotic critical values can be found inter alia in Johansen (1995) or in Osterwald-Lenun (1992). Approximations of the asymptotic distribution has been proposed by Doornik (1998).

Once \( \beta \) has been found in a first step and superconsistently estimated by \( \hat{\beta} \), we can implement the common feature test statistics. We make again the distinction between the SCCF and the WF specifications. These are respectively based on the following reduced rank regressions:

\[
SCCF : \text{CanCor}\{\Delta y_t, (\Delta y'_{t-1}, \ldots \Delta y'_{t-p+1}, y'_{t-1}\hat{\beta})' | D_t \},
\]

\[
WF : \text{CanCor}\{\Delta y_t, (\Delta y'_{t-1}, \ldots \Delta y'_{t-p+1})' | (D'_t, y'_{t-1}\hat{\beta})' \}.
\]

Both procedures allow to obtain the squared canonical correlations, namely the eigenvalues
For the VECM of order $p - 1$, the significance of the $s$ smallest eigenvalues is evaluated through the following likelihood ratios:

$$LR_s^{SCCF} = -T \sum_{i=1}^{s} \ln(1 - \hat{\lambda}_i^{SCCF}) \sim \chi^2(v_1), \quad s = 1 \ldots n - r,$$

$$LR_s^{WF} = -T \sum_{i=1}^{s} \ln(1 - \hat{\lambda}_i^{WF}) \sim \chi^2(v_2), \quad s = 1 \ldots n,$$

with $v_2 = s \times n(p - 1) - s \times (n - s)$ for the WF and $v_1 = s \times (n(p - 1) + r) - s \times (n - s)$ for the SCCF. We also investigate the behavior of a correction for small sample sizes à la Reinsel and Ahn (1992), i.e.

$$LR_s^{SCCF\_cor} = -(T - n(p - 1) - r) \sum_{i=1}^{s} \ln(1 - \hat{\lambda}_i^{SCCF}),$$

$$LR_s^{WF\_cor} = -(T - n(p - 1)) \sum_{i=1}^{s} \ln(1 - \hat{\lambda}_i^{WF}),$$

for respectively the SCCF and the WF.\textsuperscript{6} Alternatively, information criteria can be used (see also Vahid and Issler, 2002). For $p$ fixed and $r$ given we can obtain these information criteria for different values of reduced rank $n - s$ using

$$AIC(\bar{p}, \bar{r}, s) = -\frac{2}{T} \log \text{lik} + \frac{2}{T} \times (\# \text{ parameters})$$

$$HQ(\bar{p}, \bar{r}, s) = -\frac{2}{T} \log \text{lik} + \frac{2 \ln \ln T}{T} \times (\# \text{ parameters})$$

$$SC(\bar{p}, \bar{r}, s) = -\frac{2}{T} \log \text{lik} + \frac{\ln T}{T} \times (\# \text{ parameters})$$

where $-\frac{2}{T}$ times the log likelihood is nothing else than the log of the determinant of the reduced rank residuals covariance matrix under common feature restrictions. The number of

\textsuperscript{5}Note that for the weak form common cycle analysis, an equivalent way to find the eigenvalues is through $\text{CanCor}\{\Delta y_i, y_{t-1}, \hat{\beta}\}, (\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}, y_{t-1}, \hat{\beta}) | D_t\}.$

\textsuperscript{6}The difference is due to the fact that in the WF, one concentrates out the cointegrating vectors that are considered as known. Also remark that $T$ is the real number of observations after the deduction of initial points in regressions containing lags.
parameters is obtained by subtracting the number of restrictions common dynamics impose from \( n^2 \times (p - 1) + nr \), that is to say the total number of parameters in the VECM for given \( r \) and \( p \).

### 3.2 Switching algorithms

Hansen and Johansen (1998, p.95) show how to jointly impose cointegration and SCCF restrictions. To illustrate the approach, let us consider a cointegrated VAR with an intercept, one lag in its VECM form (i.e. \( p = 2 \) in the VAR) and common factor restrictions similar to (3) such that

\[
\Delta y_t = \mu + \delta_\perp \Psi_1 \beta' y_{t-1} + \delta_\perp \Psi_2 \Delta y_{t-1} + \epsilon_t,
\]

where \( \delta_\perp \) is the orthogonal complement of the cofeature matrix, namely \( \delta_\perp' \delta_\perp = 0_{s \times n} \) and \( \text{rank}[\delta : \delta_\perp] = n \). Hansen and Johansen (1998) impose SCCF restrictions by premultiplying (11) by the partitioning matrix \( B \),

\[
B = \begin{pmatrix}
(\delta_\perp' \delta_\perp)^{-1} \delta_\perp' \\
(n-s) \times n
\end{pmatrix}
\]

\[
(\delta_\perp' \delta_\perp)^{-1} \delta_\perp' \Delta y_t = \mu^* + \Psi_1 \beta' y_{t-1} + \Psi_2 \Delta y_{t-1} + (\delta_\perp' \delta_\perp)^{-1} \delta_\perp' \epsilon_t,
\]

\[
\delta' \Delta y_t = \mu^{**} + \delta' \epsilon_t,
\]
where \( \mu^* = (\delta_\perp \delta_\perp)^{-1} \delta_\perp \mu \) and \( \mu^{**} = \delta' \mu \) are vector columns of size respectively \((n - s)\) and \(s\).

Solving (12) and (13) gives

\[
(\delta_\perp \delta_\perp)^{-1} \delta_\perp \Delta y_t = (\mu^* - \omega \mu^{**}) + \Psi_1' \beta' y_{t-1} + \Psi_2' \Delta y_{t-1} + (\delta_\perp \delta_\perp)^{-1} \delta_\perp \varepsilon_t - \omega \varepsilon_t,
\]

(14)

where \( \omega = \text{Cov}((\delta_\perp \delta_\perp)^{-1} \delta_\perp \varepsilon_t, \delta_\perp \varepsilon_t) \text{Var}(\delta_\perp \varepsilon_t)^{-1} \).

Now, one finds that to obtain the cointegrating vectors under common feature restrictions we consider \( \text{CanCor}\{((\delta_\perp \delta_\perp)^{-1} \delta_\perp \Delta y_t, y_{t-1}|(1, \Delta y_{t-1}, \Delta y'_0)(1)\}. \) The algorithm is as follows: (i) estimate \( \beta \) without constraints in a first step, i.e. the usual Johansen approach; (ii) fixing the matrix \( \beta \) to its estimated value, estimate \( s \) and \( \delta \); (iii) obtain the \( n - s \) common dynamic factors \( \Psi' = (\Psi_1', \Psi_2') \) using the duality principle of canonical correlations; (iv) estimate \( \delta_\perp \) in (11) by multivariate least squares\(^7\); (v) reestimate \( \beta \) and keep on iterating until convergence is reached.

As pointed out by Hansen and Johansen (1998, p.97), "there is no proof of convergence, but it will probably occur in practice, since each step maximized the likelihood function for fixed values of the other parameters." In the WF case there are not cross-equation restrictions similar to (11), so the constrained model is simply

\[
\Delta y_t = \mu + \alpha \beta' y_{t-1} + \delta_\perp \Psi_2' \Delta y_{t-1} + \varepsilon_t.
\]

(15)

Imposing WF restrictions is convenient because this allows to consider both cointegration and common feature test statistics without the constraint \( r + s \leq n \). To solve (15), we start by estimating \( \beta \) by ML and we fix it to find the number of common feature vectors \( s \). We estimate the \( n - s \) dynamic common factors forming \( \Psi_2' \) in (15) and we use this constraint to reestimate \( \beta \) using the program \( \text{CanCor}\{\Delta y_t, y_{t-1}|(1, \Delta y_{t-1}, \Delta y'_0)(1)\}. \) This sequence is iterated until convergence is reached.\(^8\) In both cases (SCCF and WF), tests for common features and

\(^7\)Alternatively the orthogonal complement of \( \delta \) can directly be obtained by standard routines such as the command NULL in Gauss. This would merge steps (ii) to (iv). However this latter approach is more sensitive to identifying restrictions on \( \delta \) (see Gonzalo and Ng, 2002).

\(^8\)In our simulations, convergence is defined when the difference in the value of log-likelihood between two iterations is less then \( 10^{-6} \).
cointegration can be obtained by evaluating log-likelihood functions at their maxima for each models with \( r = 0 \ldots n \) and \( s = 0 \ldots n \) and to form usual likelihood ratio tests by taking twice their differences.

4 The Monte Carlo design and the performance of test statistics

4.1 The data generating process

The underlying data generating process is a stylized second order VAR with four \( I(1) \) variables and two cointegrating vectors:

\[
\begin{pmatrix}
\Delta y_{1t} \\
\Delta y_{2t} \\
\Delta y_{3t} \\
\Delta y_{4t}
\end{pmatrix} = \begin{pmatrix}
.25 \\
-.15 \\
-.1 \\
.5
\end{pmatrix} + \begin{pmatrix}
-.2 & .2 \\
-.8 & -.4 \\
-1 & .8 \\
-.5 & 0
\end{pmatrix} \begin{pmatrix}
y_{1t-1} \\
y_{2t-1} \\
y_{3t-1} \\
y_{4t-1}
\end{pmatrix} \\
+ \begin{pmatrix}
-.1 \\
-.4 \\
-.2 \\
-.25
\end{pmatrix} \begin{pmatrix}
2 & 1 & 1 & 1
\end{pmatrix} \begin{pmatrix}
\Delta y_{1t-1} \\
\Delta y_{2t-1} \\
\Delta y_{3t-1} \\
\Delta y_{4t-1}
\end{pmatrix} + \begin{pmatrix}
\epsilon_{1t} \\
\epsilon_{2t} \\
\epsilon_{3t} \\
\epsilon_{4t}
\end{pmatrix}.
\]

There also exist in (16) three weak form common feature vectors, i.e. \( \text{rank}(\Gamma_1) = 1 \). These three normalized linearly independent cofeature vectors are \( \delta_1 = (1 - .25 0 0) \), \( \delta_2 = (1 0 -.5 0) \) and \( \delta_3 = (1 0 0 - .4) \).\(^9\) The roots of the determinant of the characteristic equation are 1, 1, -1.55, 1.02±.35i. Disturbances \( \epsilon_{it} \) follow a zero mean Gaussian distribution with variances 1 and all covariances equal 0.7. Three sample sizes are considered: \( T = 50, 100, 200 \). Computations have been carried out using Gauss; 10,000 replications are used; the first 50 observations are discarded to remove dependence on initial observations.

\(^9\)It is shown in Hecq et al. (2004) that \( s \) WF vectors imply \( s - r \) SCCF vectors. This is a direct extension of Vahid and Engle (1993)'s lemma which shows that in a cointegrated VAR(1), i.e. a model with \( n \) WF vectors, there exist \( n - r \) SCCF vectors. In this case there exists one SCCF vector. But this vector suffers from identification problems.
Next we present in two distinct subsections outcomes for cointegration and for common features.

4.2 Results on cointegration

Tables 1 and 2 report the empirical size and the size unadjusted power of Johansen’s trace statistics (model with a unconstrained constant) when imposing WF common feature restrictions. The nominal size has been fixed to 5% with asymptotic critical values given in Osterwald-Lenum (1992). Rows labelled $s = 0$ report the rejection frequencies when no common feature restrictions are imposed. For $s = 1, 2, 3, 4$ we report rejection frequencies obtained using an iterative procedure to reach the maxima. $s = 3$ is the correct number of restrictions. When one overestimates $s$, i.e. $s = 4$, the estimated system reduces to a VAR(1). We consider in Table 1 the case of a correct choice of the lag length, i.e. $p = 2$. Table 2 proposes the same output when we take $p = 4$ and thus we overestimate the lag order in the estimated model. Both tables give the rejection frequencies of the asymptotic trace test (5) and its small sample corrected version (6).

There are two cointegrating vectors in the DGP. Consequently, the rejection frequencies under the hypotheses $r = 0$ and $r \leq 1$ measure the unadjusted empirical power, that is to say the probability to reject respectively zero and one cointegrating vector for a higher number. A relatively small rejection frequency in these columns indicates that we might consider too few long-run relationships. For the size properties we focus on the column $H_0 : r \leq 2$ whose values should not be too far from 5%, the nominal size. A higher rejection frequency emphasizes the detection of two many cointegrating vectors. The last column $H_0 : r \leq 3$ gives the likelihood to obtain a stationary system.

Some comments are in order:

- For $s = 0$, namely the case without restrictions on the common dynamics, we observe, as expected, some size distortions in small samples (with $T = 50$ and to a lesser extent with
<table>
<thead>
<tr>
<th>$T$</th>
<th>$s$</th>
<th>$r = 0$</th>
<th>$r \leq 1$</th>
<th>$r \leq 2$</th>
<th>$r \leq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>$s = 0$</td>
<td>100</td>
<td>100</td>
<td>99.54</td>
<td>96.33</td>
</tr>
<tr>
<td></td>
<td>$s = 1$</td>
<td>100</td>
<td>100</td>
<td>99.58</td>
<td>96.52</td>
</tr>
<tr>
<td></td>
<td>$s = 2$</td>
<td>100</td>
<td>100</td>
<td>99.63</td>
<td>97.46</td>
</tr>
<tr>
<td></td>
<td>$s = 3$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 4$</td>
<td>100</td>
<td>100</td>
<td>79.59</td>
<td>67.62</td>
</tr>
<tr>
<td>100</td>
<td>$s = 0$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 1$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 2$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 3$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 4$</td>
<td>100</td>
<td>100</td>
<td>97.39</td>
<td>95.46</td>
</tr>
<tr>
<td>200</td>
<td>$s = 0$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 1$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 2$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 3$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$s = 4$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: Rejection frequencies of Johansen’s trace test - Switching procedure with WF restrictions and a lag length $p=2$ in the estimated model
Table 2: Rejection frequencies of Johansen’s trace test - Switching procedure with WF restrictions and a lag length p=4 in the estimated model

<table>
<thead>
<tr>
<th>T</th>
<th>s</th>
<th>$H_0$</th>
<th>$r = 0$</th>
<th>$r \leq 1$</th>
<th>$r \leq 2$</th>
<th>$r \leq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$LR_r$</td>
<td>$LR_{cor}^r$</td>
<td>$LR_r$</td>
<td>$LR_{cor}^r$</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td></td>
<td>98.45</td>
<td>53.03</td>
<td>68.16</td>
<td>10.07</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>100</td>
<td>98.99</td>
<td>60.14</td>
<td>71.74</td>
<td>12.52</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>100</td>
<td>99.84</td>
<td>82.98</td>
<td>81.13</td>
<td>23.63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>100</td>
<td>100</td>
<td>96.54</td>
<td>63.27</td>
<td>13.33</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100</td>
<td>100</td>
<td>79.59</td>
<td>50.32</td>
<td>44.75</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>97.01</td>
<td>84.79</td>
<td>9.66</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>97.94</td>
<td>87.72</td>
<td>9.66</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>100</td>
<td>100</td>
<td>98.80</td>
<td>93.16</td>
<td>8.87</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>100</td>
<td>100</td>
<td>100 99.98</td>
<td>7.89</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100</td>
<td>100</td>
<td>97.39</td>
<td>92.13</td>
<td>64.91</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100 100</td>
<td>6.67</td>
<td>4.10</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100 100</td>
<td>6.40</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>100</td>
<td>100</td>
<td>100 100</td>
<td>6.30</td>
<td>3.84</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>100</td>
<td>100</td>
<td>100 100</td>
<td>5.78</td>
<td>3.76</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>100</td>
<td>100</td>
<td>100 100</td>
<td>71.63</td>
<td>66.97</td>
</tr>
</tbody>
</table>
$T = 100$). Empirical sizes based on asymptotic tests are about 16% for $T=50$ and 9% for $T=100$ whatever the choice of $p$.

- Again reading the $s = 0$ line, the small sample correction behaves quite well when $p = 2$ but is not so nice when $p$ increases as the strong correction $T − np$ leads to underestimate the number of cointegrating vectors (empirical size less than 5%).

- $s = 4$ means we end up with a VAR(1) and consequently the dynamics is underestimated. This creates size distortions.

We have already mentioned that these issues are well documented in the literature. Next we consider the effect of also imposing common feature restrictions. In these cases (rows $s = 1, 2, 3$) we observe that:

- Taking into account the common feature restrictions always slightly reduces (mainly for $T=50$) size distortions, the reduction being maximum with the correct number of restrictions, i.e. $s = 3$. The difference is not so huge however. For instance with $p = 4$ and 50 observations, the asymptotic test statistics (5) has an empirical size of 16.98% while the size with $s = 3$ is 13.33%.

- The switching approach also help in not rejecting the null that $r \leq 3$ and thus to find a fourth cointegrating vector. This will be illustrated in the empirical section where, based on asymptotic distributions, we first find six cointegrating vectors among six countries.

- An interesting aspect when testing for cointegration under WF restrictions is also the increase of power when the lag length is large. With $T = 50$ and for $s = 0$, we determine in only 10% the presence of at least a second cointegrating vector for the small sample corrected statistics while this proportion increases to 63% with $s = 3$. Similar results are observed for $r = 0$. The power of the test being only 53% in the usual approach but reaches 100% if we use the iterative one.

As expected, we see that in VAR with longer lags, common feature restrictions are helpful to improve the performance of both the power and the size of cointegration test statistics. In
practice, the biggest empirical problem could be to know whether we face a power problem (asymptotic tests cannot detect the whole set of long-run relationships) or a size one (tests detect too many). Next we evaluate the impact of the switching procedure on common feature test statistics.

4.3 Results on common features

Tables 3 and 4 present for respectively $p = 2$ and $p = 4$ in the estimated VAR models, different testing strategy to find WF common features. On the one hand there are the 2-step and the iterative likelihood ratios, both using the asymptotic test and the modified version. We only stress the size properties, namely the frequency to reject the null hypothesis there exists three common feature vectors, i.e. $s = 3$. The results about the empirical power, i.e. the rejection of the existence of a fourth common feature vector is rejected in 100% of the cases and is not reported. On the other hand we also present the percentage of finding $s = 3$ using information criteria. We only report in the tables the frequencies to find the correct number of common feature vectors in a 2-step approach. We comment in the text some results for information criteria obtained in the switching framework.

Tables report the results when choosing different cointegrating vectors $r = 1$ to 3. With $r = 4$, it does not make sense to switch because the cointegrating space spans $\mathbb{R}^n$. In all these cases the cointegrating vectors are estimated. In order to have a benchmark, we also show the results with $r = 2$ but with the coefficients of $\beta$ fixed to their true values (entry $\hat{\beta}_{r=2}$ instead of $\hat{\beta}_{r=2}$).

The following results emerge:

- As it was noticed in Hecq et al. (2004) the performance of test statistics is dramatically altered when $r$ is underestimated, i.e. the rows with $r = 1$. This is why in the WF it is sensible to first fix the cointegrating rank to its upper plausible bound. For instance in a convergence analysis, to start with $r = n - 1$. 

16
Table 3: Empirical size of the 2-step and the iterative WF common feature tests and information criteria, p=2 in the estimated model.
\[ H_0 : s = 3 \]

<table>
<thead>
<tr>
<th></th>
<th>2-step</th>
<th>Iterative</th>
<th>2-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( LR_s )</td>
<td>( LR_{s}^{cor} )</td>
<td>( LR_s )</td>
</tr>
<tr>
<td>( T = 50 )</td>
<td>( \hat{\beta}_{r=1} )</td>
<td>98.75</td>
<td>92.49</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=2} )</td>
<td>57.85</td>
<td>20.82</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=3} )</td>
<td>51.71</td>
<td>12.43</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=4} )</td>
<td>49.16</td>
<td>11.27</td>
</tr>
<tr>
<td></td>
<td>( \beta_{r=2} )</td>
<td>32.48</td>
<td>5.19</td>
</tr>
<tr>
<td>( T = 100 )</td>
<td>( \hat{\beta}_{r=1} )</td>
<td>99.43</td>
<td>98.87</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=2} )</td>
<td>20.75</td>
<td>9.02</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=3} )</td>
<td>19.38</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=4} )</td>
<td>18.71</td>
<td>7.28</td>
</tr>
<tr>
<td></td>
<td>( \beta_{r=2} )</td>
<td>14.00</td>
<td>5.14</td>
</tr>
<tr>
<td>( T = 200 )</td>
<td>( \hat{\beta}_{r=1} )</td>
<td>99.98</td>
<td>99.97</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=2} )</td>
<td>10.58</td>
<td>6.32</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=3} )</td>
<td>10.09</td>
<td>6.15</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_{r=4} )</td>
<td>9.78</td>
<td>6.00</td>
</tr>
<tr>
<td></td>
<td>( \beta_{r=2} )</td>
<td>8.66</td>
<td>5.34</td>
</tr>
</tbody>
</table>

Table 4: Empirical size of the 2-step and the iterative WF common feature tests and information criteria, p=4 in the estimated model
• Comparing the columns where respectively the 2-step and the iterative procedure are reported we observe that when \( p \) is large, the switching procedure reduces the size distortions. This is less visible with \( p = 2 \) because there is only a small difference between imposing and not imposing the restrictions.

• Overall we should recognize the relative good performance of these test given the underlying large number of restrictions: respectively 9 and 33 for \( p = 2 \) and \( p = 4 \). In particular small sample corrections seem to work well. For instance, with \( r = 2, p = 4 \) and \( T = 50 \), the empirical size is 9.7% with the small sample correction and 45.82% without it. Notice that for the same specification, a two steps strategy would respectively yield empirical sizes of 20.82% and 57.85%. For 100 observations the sizes are quite close to the nominal ones.

• However, for small \( T \) these size distortions are larger than the ones obtained with known cointegrating vectors.

• Information criteria work remarkably well and especially the SC (frequency to find \( s = 3 \) is always higher than 95%) and to a less extent the HQ. What is generally not appreciated in practice with the SC, that is to say the fact that it is often too parsimonious when considering the lag length of a VAR for instance, seems to be an advantage here.

• The AIC has the tendency to pick up to few common feature vectors. This is a bit better when the cointegrating vectors are known. In order to point out the influence of the estimation of \( \beta \), we have also computed these information criteria in the iterative approach. For instance, with \( T = 50 \) the frequency to correctly find \( s = 3 \) are for the AIC, HQ and SC, 69.07%, 88.16% and 97.86% for \( p = 2 \) and 54.40%, 89.13%, 99.53% for \( p = 4 \). This is higher than the numbers from Tables 3 and 4, especially for AIC and HQ.

To conclude with the WF we might say that the iterative procedure may help if combined with an adjustment for small samples.\(^{10}\) We must be careful and avoid underestimating the

---

\(^{10}\)The issue of this paper was not to compare the merits of different small sample versions for common feature test statistics. At least we have illustrated that anything is better than to rely on asymptotic distributions. For
number of cointegrating vectors. Information criteria and in particular the SC are tools we should not forget. We do not report the results of the SCCF test statistics since we want to let $r$ and $s$ vary freely. The underlying DGP implies one SCCF vector but to detect it is better to rely on the mixed form framework developed in Hecq et al. (2004). To simplify the analysis, we consider the same DGP in (16) with $\Gamma_1 = 0_{n \times n}$. In this case, we have a VAR(1) and the space spanned by $\delta$ is actually the space spanned by the columns of $\alpha_\perp$. We consider the case with $T = 50$ and $p = 2$ in the estimated model and compare the SCCF empirical sizes ($H_0: s = 2$) when $\beta$ is known and estimated. In the latter case we can analyze the effect of an overestimation of the number of cointegrating vector $r$. The following pairs refer to the asymptotic and the modified tests respectively. When $\beta$ is fixed to its true value we obtain the rejection frequencies [9.65% 5.22%]. When $\beta$ is estimated, the two step approach gives respectively [14.63% 8.83%] and [47.43% 30.09%] for $r = 2$ and $r = 3$. A switching procedure reduces these percentages. But the important point here is the large size distortion when overestimating $r$. This explains why we have proposed to start with the WF to determine $r$ and $s$ and then to look at SCCF for some plausible number of common feature vectors. This approach is applied in the next section.

5 Common cycles and common trends in Latin America

This section investigates the presence of short and the long-run interactions between the output of six Latin American economies: Brazil, Venezuela, Mexico, Peru, Columbia and Chile. We consider two datasets, one for the real gross domestic product (RGDP hereafter) and another one with output per capita series (RGDP_K hereafter). The annual variables are extracted from the Total Economy Database and span the period 1950-2002, thus a sample of 53 observations instance, instead of using the small sample correction à la Reinsel and Ahn we could have considered a Bartlett correction (see Mardia, Kent and Bibby, 1979) by replacing respectively $(T - n(p - 1))$ and $(T - n(p - 1) - r)$ by $(T - \frac{1}{2}(n + n(p - 1) + 3))$ and $(T - \frac{1}{2}(n + n(p - 1) + r + 3))$. For our DGP (16), these Bartlett’s style modified statistics behave better for $p = 2$, are equivalent to the one proposed in this paper for $p = 3$ and the correction à la Reinsel and Ahn is far better when $p = 4$. 

20
for \( n = 6 \).\(^{11}\) A “pretest” check pleaded for excluding Argentina from the study. Indeed, adding Argentina made the analysis less robust for the identification of the VAR order and the number of cointegrating vectors. Moreover it was quite difficult to fit a system in which Argentinian variables were significant.

We consider the model with a restricted deterministic trend in the long-run and five lags were necessary to capture the dynamics of the multivariate process. Table 5 reports Johansen’s trace test. Familiar results emerge: the asymptotic test probably overestimates the number of cointegrating vectors as we reject the null \( H_0 : r \leq 5 \). As it is illustrated on Figure 1 for per capita series, it is indeed quite unlikely that all output variables are stationary around a deterministic trend. Similarly the short-run correction penalizes too much. Probably \( r \) should lie between two and five for these two samples. The determination of the number of long-run relationships is however crucial for the interpretation of the convergence among economies (see inter alia Bernard and Durlauf, 1995).

We leave the determination of the cointegrating rank unsolved for the moment and we go ahead with the analysis of short-run co-movements. We have seen that an overestimation of \( r \) only marginally affects the weak form common cycle test statistic while fixing \( r \) to a value below the one in the data generating process gives misleading answers. Consequently we choose \( r = 5 \) for a 2-step approach presented in Table 6. We report the value of the asymptotic likelihood ratio test (8), its associated degrees of freedom, the \( p \) values for both the asymptotic and the small sample corrected version (10) and the three information criteria. The observations made in the Monte Carlo experiment are again well illustrated here and are quite helpful to understand the results. Indeed, asymptotic tests would favor \( s = 0 \), namely no common feature vectors, while we would choose \( s = 2 \) for the real output series and maybe \( s = 3 \) for the per capita real output based on small sample modified versions. AIC would take \( s = 0 \) or \( s = 1 \), HQ \( s = 1 \) but SC would favor \( s = 3 \) for both sets of variables. Also note that the iterative

Figure 1: Log-levels of per capita gross domestic products
We continue the analysis and we purposely consider the three possibilities offered by cofeature tests in Table 6, namely $s = 1, 2$ or $3$. For each $s$ we then test for cointegration using the iterative approach. From Table 7 we would recommend $r = 3$ although it is a little bit intuitive here. The reason is that we believe that there exist three common feature vectors and in the column $s = 3$ the values of the asymptotic tests are quite different between rejecting the null hypothesis of less than two and less then three cointegrating vectors. For $r = 3$, Table 8 reports the results for the common feature switching approach. $P-values$ for asymptotic and small

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>$\lambda_i$</th>
<th>$LR_r$</th>
<th>$LR_r^{cor}$</th>
<th>95%cv</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rgdp _K$</td>
<td>$r = 0$</td>
<td>0.939</td>
<td>370.95 *</td>
<td>139.10 *</td>
</tr>
<tr>
<td></td>
<td>$r \leq 1$</td>
<td>0.837</td>
<td>238.63 *</td>
<td>89.48 *</td>
</tr>
<tr>
<td></td>
<td>$r \leq 2$</td>
<td>0.723</td>
<td>151.29 *</td>
<td>56.73</td>
</tr>
<tr>
<td></td>
<td>$r \leq 3$</td>
<td>0.581</td>
<td>89.62 *</td>
<td>33.60</td>
</tr>
<tr>
<td></td>
<td>$r \leq 4$</td>
<td>0.414</td>
<td>47.81 *</td>
<td>17.93</td>
</tr>
<tr>
<td></td>
<td>$r \leq 5$</td>
<td>0.369</td>
<td>22.12 *</td>
<td>8.29</td>
</tr>
</tbody>
</table>

| $rgdp$ | $r = 0$ | 0.939 | 371.72 * | 139.39 * | 114.96 |
|        | $r \leq 1$ | 0.824 | 237.39 * | 89.02 * | 86.96 |
|        | $r \leq 2$ | 0.725 | 153.82 * | 57.68 | 62.61 |
|        | $r \leq 3$ | 0.603 | 91.71 * | 34.39 | 42.20 |
|        | $r \leq 4$ | 0.437 | 47.35 * | 17.75 | 25.47 |
|        | $r \leq 5$ | 0.336 | 19.72 * | 7.39 | 12.39 |

Table 5: Cointegration trace statistics for the six Latin American economies over 1950-2002 - Model with a restricted trend in the long-run, $p=5$
<table>
<thead>
<tr>
<th></th>
<th>$H_0$</th>
<th>$LR_{WF}^{s}$</th>
<th>df</th>
<th>$p_{val}$</th>
<th>$p_{val}^{cor}$</th>
<th>AIC</th>
<th>HQ</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rgdp$</td>
<td>$s = 0$</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−47.220*</td>
<td>−44.657</td>
<td>−40.437</td>
</tr>
<tr>
<td></td>
<td>$s = 1$</td>
<td>39.91</td>
<td>19</td>
<td>0.003</td>
<td>0.397*</td>
<td>−47.180</td>
<td>−44.897*</td>
<td>−41.138</td>
</tr>
<tr>
<td></td>
<td>$s = 2$</td>
<td>103.58</td>
<td>40</td>
<td>&lt;0.001</td>
<td>0.100*</td>
<td>−46.729</td>
<td>−44.755</td>
<td>−41.505</td>
</tr>
<tr>
<td></td>
<td>$s = 3$</td>
<td>170.67</td>
<td>63</td>
<td>&lt;0.001</td>
<td>0.030</td>
<td>−46.290</td>
<td>−44.654</td>
<td>−41.962*</td>
</tr>
<tr>
<td></td>
<td>$s = 4$</td>
<td>290.28</td>
<td>88</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>−44.839</td>
<td>−43.572</td>
<td>−41.487</td>
</tr>
<tr>
<td></td>
<td>$s = 5$</td>
<td>423.44</td>
<td>115</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>−43.190</td>
<td>−42.321</td>
<td>−40.890</td>
</tr>
<tr>
<td></td>
<td>$s = 6$</td>
<td>584.79</td>
<td>144</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>−41.037</td>
<td>−40.596</td>
<td>−39.868</td>
</tr>
</tbody>
</table>

Table 6: 2-step WF common feature tests statistics with p=5 and r=5, 1950-2002

sample corrected tests are presented together with the values of iterated information criteria. Now there is a clear cut in favor of $s = 3$ if we look at the corrected test and either the HQ and the SC.

We have an interesting result here because there are as many long-run co-movements as short-run ones or say differently, there exist three permanent and three transitory shocks driven the output of the six countries. Moreover, the constraint $s + r \leq n$ allows us to go further and to analyze whether the three weak form cofeature relationships are also SCCF vectors (see Hecq et al. 2004). We use the algorithm exposed in Section 3.2., that is to say we fix $r = 3$ and we test for the number of SCCF vectors by iterating between the cointegrating and the common feature spaces. Table 9 reveals that we cannot reject the null of three serial correlation common feature vectors if we look at small sample corrected likelihood ratio tests, or more exactly the $p-values$ associated with these tests. SC also favors $s = 3$ while HQ takes $s = 3$ for per capita series and $s = 2$ for the other dataset. AIC and asymptotic likelihood ratio tests would keep $s = 1$ but we are aware from the Monte Carlo results that these two latter procedures
underestimate the number of common feature vectors in small samples. This result also means that by definition the three WF vectors were also SCCF ones. To formally study this issue we can compare the Schwarz criteria obtained in Tables 8 and 9 for $s = 3$. It emerges that their values are smaller for both datasets for SCCF than for WF.

This case with $r + s = n$ is ideal for extracting the long and the short-run co-movements because these components are exactly identified. Simple methods might be considered such as the one proposed in Vahid and Engle (1993) for the Beveridge-Nelson cycles. This latter decomposition would be identical to the Gonzalo-Granger (1995) ones were the three common cycles are given by $\beta^iy_t$ and the common trends by $\delta^iy_t$. Namely we only need estimated cointegrating and common feature vectors that maximize the likelihood after iterations. We do not graph these components because the scope was to point out that misleading results concerning the number of co-movements can be obtained when working with small samples. Moreover we do not want to give the false impression commonly shared in the literature that finding $r + s = n$ is the final goal of a common trend - common cycle study.

<table>
<thead>
<tr>
<th></th>
<th>$s = 1$</th>
<th>$s = 2$</th>
<th>$s = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>$L_{R_r}$</td>
<td>$L_{R_{r}^{cor}}$</td>
<td>$L_{R_r}$</td>
</tr>
<tr>
<td>$rgdp_K$</td>
<td>$r = 0$</td>
<td>353.91*</td>
<td>132.71*</td>
</tr>
<tr>
<td></td>
<td>$r \leq 1$</td>
<td>226.90*</td>
<td>85.08</td>
</tr>
<tr>
<td></td>
<td>$r \leq 2$</td>
<td>137.94*</td>
<td>51.72</td>
</tr>
<tr>
<td></td>
<td>$r \leq 3$</td>
<td>73.28*</td>
<td>27.48</td>
</tr>
<tr>
<td></td>
<td>$r \leq 4$</td>
<td>30.98*</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td>$r \leq 5$</td>
<td>8.26</td>
<td>3.09</td>
</tr>
<tr>
<td>$rgdp$</td>
<td>$r = 0$</td>
<td>350.12*</td>
<td>131.29*</td>
</tr>
<tr>
<td></td>
<td>$r \leq 1$</td>
<td>219.79*</td>
<td>82.42</td>
</tr>
<tr>
<td></td>
<td>$r \leq 2$</td>
<td>136.21*</td>
<td>51.08</td>
</tr>
<tr>
<td></td>
<td>$r \leq 3$</td>
<td>70.64*</td>
<td>26.49</td>
</tr>
<tr>
<td></td>
<td>$r \leq 4$</td>
<td>28.54*</td>
<td>10.70</td>
</tr>
<tr>
<td></td>
<td>$r \leq 5$</td>
<td>8.10</td>
<td>3.04</td>
</tr>
</tbody>
</table>

Table 7: Iterative cointegration trace statistics, p=5, 1950-2002
### Table 8: WF common feature tests statistics, p=5, r=3, 1950-2002

<table>
<thead>
<tr>
<th>H_0</th>
<th>LR_s</th>
<th>df</th>
<th>p_val</th>
<th>p_val_cor</th>
<th>AIC</th>
<th>HQ</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>rgdp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s = 1</td>
<td>19.03</td>
<td>19</td>
<td>0.454</td>
<td>0.963</td>
<td>-46.615*</td>
<td>-44.509</td>
<td>-41.041</td>
</tr>
<tr>
<td>s = 2</td>
<td>64.96</td>
<td>40</td>
<td>0.007</td>
<td>0.795</td>
<td>-46.534</td>
<td>-44.736</td>
<td>-41.778</td>
</tr>
<tr>
<td>s = 3</td>
<td>124.97</td>
<td>63</td>
<td>&lt;0.001</td>
<td>0.494</td>
<td>-46.242</td>
<td>-44.783*</td>
<td>-42.382*</td>
</tr>
<tr>
<td>s = 4</td>
<td>234.92</td>
<td>88</td>
<td>&lt;0.001</td>
<td>0.019</td>
<td>-44.993</td>
<td>-43.902</td>
<td>-42.108</td>
</tr>
<tr>
<td>s = 5</td>
<td>362.89</td>
<td>115</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-43.452</td>
<td>-42.759</td>
<td>-41.619</td>
</tr>
<tr>
<td>rgdp_K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s = 1</td>
<td>19.12</td>
<td>19</td>
<td>0.449</td>
<td>0.962</td>
<td>-46.531</td>
<td>-44.425</td>
<td>-40.957</td>
</tr>
<tr>
<td>s = 2</td>
<td>59.57</td>
<td>40</td>
<td>0.023</td>
<td>0.881</td>
<td>-46.563*</td>
<td>-44.766</td>
<td>-41.807</td>
</tr>
<tr>
<td>s = 3</td>
<td>121.69</td>
<td>63</td>
<td>&lt;0.001</td>
<td>0.553</td>
<td>-46.228</td>
<td>-44.769*</td>
<td>-42.368*</td>
</tr>
<tr>
<td>s = 4</td>
<td>237.23</td>
<td>88</td>
<td>&lt;0.001</td>
<td>0.016</td>
<td>-44.862</td>
<td>-43.772</td>
<td>-41.977</td>
</tr>
<tr>
<td>s = 5</td>
<td>367.00</td>
<td>115</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-43.268</td>
<td>-42.575</td>
<td>-41.436</td>
</tr>
</tbody>
</table>

### Table 9: SCCF tests statistics, p=5, r=3, 1950-2002

<table>
<thead>
<tr>
<th>H_0</th>
<th>df</th>
<th>p_value</th>
<th>p_value_cor</th>
<th>AIC</th>
<th>HQ</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>rgdp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s = 1</td>
<td>22</td>
<td>0.340</td>
<td>0.980</td>
<td>-46.634*</td>
<td>-44.572</td>
<td>-41.177</td>
</tr>
<tr>
<td>s = 2</td>
<td>46</td>
<td>&lt;0.001</td>
<td>0.885</td>
<td>-46.477</td>
<td>-44.768*</td>
<td>-41.955</td>
</tr>
<tr>
<td>s = 3</td>
<td>72</td>
<td>&lt;0.001</td>
<td>0.665</td>
<td>-46.060</td>
<td>-44.735</td>
<td>-42.552*</td>
</tr>
<tr>
<td>rgdp_K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s = 1</td>
<td>22</td>
<td>0.318</td>
<td>0.978</td>
<td>-46.543*</td>
<td>-44.481</td>
<td>-41.085</td>
</tr>
<tr>
<td>s = 2</td>
<td>46</td>
<td>0.005</td>
<td>0.937</td>
<td>-46.517</td>
<td>-44.808</td>
<td>-41.995</td>
</tr>
<tr>
<td>s = 3</td>
<td>72</td>
<td>&lt;0.001</td>
<td>0.774</td>
<td>-46.151</td>
<td>-44.825*</td>
<td>-42.642*</td>
</tr>
</tbody>
</table>

### 6 Conclusion

In this paper, we studied a linear Gaussian VAR model with nonstationary but cointegrated variables that have common cyclical features. Similarly to other papers we have pointed out through Monte Carlo simulations that cointegrating test statistics but also common feature procedures based on asymptotic distributions suffer from size distortions. Iterating between spaces helps to reduce the size distortion observed in Johansen’s trace test but without reaching the nominal size. However the power is improved. Common feature test statistics behave quite well
and especially if we combine a switching approach with a small sample correction. Information
criteria work remarkably well and in particular the Schwarz criterion that had already a good
behavior in a 2-step approach but can be improved by iterating between cointegrating and the
common feature spaces. Consequently we propose another strategy than the one advocated in
Vahid and Issler (2002). Indeed for forecasting purposes they show that the best strategy is to
let $p$ and $s$ (SCCF vectors) be ”freely” chosen by information criteria and in particular by the
HQ. In our case, for determining the number of long and short-run co-movements we propose
to choose the lag order in the VAR with AIC for instance (Gonzalo and Pitarakis, 2002) and
then to test for cointegration. One can choose a maximum $r$ and test for WF common features
using small sample likelihood ratio tests and information criteria (SC). In summary, different
information criteria are best used for different purposes. Having determined $r$ and $s$ WF vec-
tors, the presence of SCCF relationships can be tested for $r + s \leq n$. This strategy helps us
to determine the existence of three permanent and three transitory shocks within the six Latin
American economies whatever the dataset we use. If we had blindly trust the asymptotics, we
would have obtained that six transitory shocks drive the economies!

References

(John Wiley & Sons).


Applied Econometrics, 10, 97-108.

Features During Financial Turmoil, forthcoming in Journal of International Money and
Finance.


