Testing for Common Autocorrelation in Data Rich Environments

Gianluca Cubadda, Alain Hecq

University of Rome "Tor Vergata", Maastricht University

March 2009

What's the matter?

1. This paper proposes a strategy to detect the presence of common cyclical features in large dimensional systems.
2. Indeed we are interested in co-movements observed in business cycles.
3. In this setting, the dimensionality issue can be even more important than for cointegration because we have less economic theory for selecting series (e.g 27 European GDP's).
4. Hence, there exist problems of size and power in usual tests when n and p are large relative to T.

What we do

1. We propose to replace the orthogonality to the past condition by the condition of absence of autocorrelation (see Lucke).
2. This drastically decreases the number of restrictions and makes a test feasible when canonical correlation is not.
3. If we are ready to rely on this more restrictive way to see this issue, remains the problem of estimating common feature vectors.
4. We use a PLS approach instead of cancor and...this works (in MC and for empirical investigations).

In summary the two main contributions of this pape arer: (1) to weaken the definition of common features and (2) to propose an alternative way to estimate cofeature space which is more robust in the case of high dimensional systems.

Outline of the talk

1. I say something on common cyclical features, i.e. on different factor structures for the short-run dynamics.
2. I show the two different orthogonality conditions.
3. Canonical correlations versus PLS (different algorithms)
4. Simple Box-Pierce type Q test.
5. Some Monte Carlo results.
6. Examples.

Common cyclical feature:
General principle

- A Common Feature setting (Engle and Kozicki (1993)) is nothing else than a factor model for instance with $Z_t\tilde{}I(1)$ and $\Delta Z_t\tilde{}I(0)$

$$\Delta Z_t = \Lambda \underbrace{F_t}_{feature\ x} + \underbrace{\varepsilon_t}_{no-feature\ x}$$

- Where F_t has the feature and ε_t doesn't have that feature.
- F_t can have almost any feature you have in mind (seasonality, volatility,...)
- In business cycle analysis, the feature could be the cycle.
- And a common cyclical feature analysis evaluates whether a cycle is common to several variables.
- Different approaches have been proposed: SCCF, PSCCF, Codependence.

Different forms of common cyclical features

For the VAR(p) in first differences

$$\Delta Z_t = \Phi_1 \Delta Z_{t-1} + ... + \Phi_p \Delta Z_{t-p+1} + \varepsilon_t$$

- SCCF (Vahid and Engle, 1993):

$$\delta' \Delta Z_t = \delta' \varepsilon_t$$

- PSCCF (Cubadda and Hecq, 2001)

$$\delta' \Delta Z_t = \delta' \Phi_1 \Delta Z_{t-1} + \delta' \varepsilon_t$$

- Codependence

$$\delta' \Delta Z_t = \delta' \varepsilon_t + \delta' \Theta_1 \varepsilon_{t-1}$$

- These models imply different types of reduced rank restrictions in the VAR.
- These modes can be extended to the VECM but we focus on the VAR in this paper.

Different orthogonality conditions (SCCF case only here)

SCCF allows for rewriting the VAR into the following Reduced Rank Regression model

$$\Delta Z_t = \delta_{\perp} \psi' W_{t-1} + \varepsilon_t,$$

where $W_{t-1} = [\Delta Z'_{t-1}, \cdots, \Delta Z'_{t-p}]'$ and ψ is a full-rank $np \times (n-s)$ matrix and $\delta_{\perp}$ is a full-rank $n \times (n-s)$ matrix such that $\delta'_{\perp} \delta = 0$.

It is worth noting that the $(n-s)$ common factors $F_t = \psi' W_t$ are responsible for all the predictable dynamics of the system. Indeed, we have that

$$\mathrm{E}(\Delta Z_t | W_{t-1}) = \delta_\perp F_{t-1}$$

which in turn implies

$$\mathrm{E}(\delta' \Delta Z_t | W_{t-1}) = 0.$$

The above orthogonality condition reveals that either an instrumental variable approach or a canonical correlation analysis are appropriate statistical tools for conducting statistical inference on SCCF.

However, when n or p are large relatively to the sample size T, $W_{t-1} = [\Delta Z'_{t-1}, \cdots, \Delta Z'_{t-p}]'$ becomes quite large too.
It is consequently convenient to resort to a weaker orthogonality condition such that

$$\mathrm{E}[\delta_i' \Delta Z_t | (\Delta Z_{t-1}{}' \delta_i, \cdots \Delta Z'_{t-p} \delta_i)'] = 0, \quad i = 1, 2, ..., s,$$

where $\delta = [\delta_1, \delta_2, \cdots, \delta_s]$ (see, e.g., Lucke, 1994).
In particular, we propose to test for this condition by means of univariate tests for no autocorrelation of each $\delta_i' y_t$ having fixed δ to a consistent estimate of a base for the SCCF space.

It should be noticed that if the first condition implies the second one, the reverse does not necessarily hold.
For instance, in the following bivariate VAR(1)

$$\begin{pmatrix} \Delta Z_{1t} \\ \Delta Z_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \phi_{21} & 0 \end{pmatrix} \begin{pmatrix} \Delta Z_{1t-1} \\ \Delta Z_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

there is a single SCCF vector such as $\delta = [1, 0]'$ that also implies $\mathrm{E}(\delta' \Delta Z_t | \delta' \Delta Z_{t-1}) = 0$.
However we also have that $\mathrm{E}(\delta'_{\perp} \Delta Z_t | \delta'_{\perp} \Delta Z_{t-1}) = 0$ when $\mathrm{E}(\varepsilon_{1t} \varepsilon_{2t}) = 0$ and consequently there exists a second linear combination satisfying the absence of autocorrelation condition.

- Hence, our testing procedure may lack power against some alternatives.
- However, we think that cases where linear combinations of autocorrelated time series satisfy the second condition but not the first condition are of theoretical interest rather than of practical relevance.
- This is the first point of the paper.
- Still remains the point concerning the estimation of the cofeature space.

- Indeed a crucial role in our testing procedure is played by the estimator of the SCCF matrix.
- Although a Canonical Correlation Analysis provides the maximum likelihood estimator of δ, it may perform poorly with high-dimensional systems because inversions of large variance matrices are required.
- Hence, we explore the possibility of resorting to Partial Least Squares as an alternative to CCA.

PLS, introduced by Wold (1958), are a family of multivariate techniques with the aim of maximizing the covariance between linear combinations of two variable sets, see, e.g., Rosipal and Krämer (2006) for a recent survey. Groen and Kapetanios (2008) have recently documented that PLS have superior forecasting performance than better known data-rich prediction methods as principal component and ridge regressions.

Partial Least Squares, and Canonical Correlation Analysis

In order to discuss statistical inference on δ, let us determine from the VAR,, under SCCF the covariance matrix $\mathrm{E}(y_t w'_{t-1})$

$$\Sigma_{yw} = \delta_{\perp}\psi'\Sigma_{ww},$$

where $\Sigma_{ww} = \mathrm{E}(W_{t-1}W'_{t-1})$ and denoting $y_t = \Delta Z_t$.
Since it follows that

$$\delta'\Sigma_{yw} = 0,$$

the matrix δ lies in the space generated by the eigenvector associated with the null eigenvalues of the symmetric semi-positive definite matrix $\Sigma_{yw}\Sigma_{wy}$.

In order to make the solution of this eigenvalue problem invariant to scale changes of individual elements of both y_t and W_t, we rather compute a base for the SCCF space as $[\nu_1^{PLS}, ..., \nu_s^{PLS}]$, where ν_i^{PLS} $(i = 1, 2, ..., s)$ is the eigenvector associated with the i-th smallest eigenvalue of the matrix

$$D_{yy}^{-1}\Sigma_{yw}D_{ww}^{-1}\Sigma_{wy},$$

with $\Sigma_{yy} = \mathrm{E}(y_t y_t')$, and D_{yy} and D_{ww} are diagonal matrices having the diagonal elements of, respectively, Σ_{yy} and Σ_{ww}.
The solution of this problem is known in multivariate statistics as a form of PLS.

A better known method to obtain the matrix δ is CCA, which can be seen as PLS after standardizing both the time-series vectors y_t and w_t. Indeed, since

$$\mathrm{E}(\Sigma_{yy}^{-1/2} y_t w_t' \Sigma_{ww}^{-1/2}) = \Sigma_{yy}^{-1/2} \Sigma_{yw} \Sigma_{ww}^{-1/2} = \Sigma_{yy}^{-1/2} \delta_{\perp} \psi' \Sigma_{ww}^{1/2},$$

the matrix δ lies in the space generated by $[\nu_1^{CCA}, ..., \nu_s^{CCA}]$, where ν_i^{CCA} $(i = 1, 2, ..., s)$ is the eigenvector associated with the i-th smallest eigenvalue of the matrix

$$\Sigma_{yy}^{-1} \Sigma_{yw} \Sigma_{ww}^{-1} \Sigma_{wy}.$$

Let $\widehat{\Omega}$ indicate the maximum likelihood estimator of a moment matrix Ω. Since the eigenvalues and eigenvector of a positive semi-definite matrix are continuous functions of that matrix (see, e.g., Magnus and Neudecker (1999)), by Slutsky's theorem the eigenvectors associated with the s smallest eigenvalues of both

$$\widehat{D}_{yy}^{-1}\widehat{\Sigma}_{yw}\widehat{D}_{ww}^{-1}\widehat{\Sigma}_{wy},$$

and

$$\widehat{\Sigma}_{yy}^{-1}\widehat{\Sigma}_{yw}\widehat{\Sigma}_{ww}^{-1}\widehat{\Sigma}_{wy}$$

are consistent estimators of a base of the SCCF space.

- Moreover, we know that CCA provides the maximum likelihood estimator of δ under Gaussianity assumption (Anderson, 1984).
- However, since PLS require to invert diagonal matrices only, this method can provide estimates of δ (up to an identification matrix) that are less disperse and more numerically stable when the dimension of W_t approaches the sample size T.

Test Statistics for Common Serial Correlation

In order to test for the presence of common cyclical features using the CCA framework, the likelihood ratio test for the null hypothesis that there exist at least s SCCF vectors is based on the statistic

$$LR_s = -T \sum_{i=1}^{s} \ln(1 - \hat{\lambda}_i^{CCA}), \qquad s = 1, \ldots, n \tag{1}$$

where $\hat{\lambda}_i^{CCA}$ is the i-th smallest eigenvalue.
The test statistic follows asymptotically a $\chi^2_{(v)}$ distribution under the null where $v = s \times np - s(n - s)$. The maximum likelihood estimator of the SCCF matrix is given by

$$\widehat{\delta}^{CCA} = [\widehat{\nu}_1^{CCA}, \cdots, \widehat{\nu}_s^{CCA}],$$

where $\widehat{\nu}_i^{CCA}$ is the eigenvector associated with $\hat{\lambda}_i^{CCA}$ for $i = 1, 2, ..., s$.

- It is clear that CCA may encounter numerical problems when n and/or p are large compared to the sample size T.
- Indeed, the matrices $\widehat{\Sigma}_{yy}$ and $\widehat{\Sigma}_{ww}$ could be (almost) numerically singular and estimates of small eigenvalues could be heavily biased.
- This is the reason why we also consider the PLS estimator of the SCCF matrix, denoted as

$$\hat{\delta}^{PLS} = [\widehat{\upsilon}_1^{PLS}, \cdots, \widehat{\upsilon}_s^{PLS}]$$

where $\widehat{\upsilon}_i^{PLS}$ is the eigenvector associated with the i-th smallest eigenvalue of the matrix $\widehat{D}_{yy}^{-1}\widehat{\Sigma}_{yw}\widehat{D}_{ww}^{-1}\widehat{\Sigma}_{wy}$ for $i = 1, 2, ..., s$.

In order to test for the presence of SCCF in systems where n and/or p are large, we consider tests based on the non autocorrelation condition.
In particular, we test for the null hypothesis that $\delta_i' y_t$ is a white-noise for $i = 1, 2, ..., s$. In order to do that, we look at the Box-Pierce test statistics and their Ljung-Box refinements

$$Q_i^j = T \sum_{k=1}^{k} \hat{r}_{i,j}^2, \qquad Q_i^{j\prime} = T(T+2) \sum_{l=1}^{k} \frac{\hat{r}_{i,j}^2}{T-l}$$

where

$$\hat{r}_{i,j} = \left(\sum_{t=l+1}^{T} \frac{e_{t,i}^j e_{t-l,i}^j}{T-l} \right) \Bigg/ \left(\frac{1}{T} \sum_{t=1}^{T} e_{t,i}^{j^2} \right),$$

and $e_{t,i}^j = \hat{\delta}_i^{j\prime} y_t$, for $j = CCA, PLS$ and $i = 1, 2, ..., s$.

- Both Q_i^j and $Q_i^{j\prime}$ follows asymptotically a $\chi^2_{(k)}$ distribution under the null.
- However, we should be careful about the size of the test when under the null hypothesis we assume the existence of more than one common feature vector, i.e. when $s > 1$.
- In this paper, we propose two different strategies to solve this problem.

1. In the first one, we use a simple Bonferroni bound approach and we apply a correction to keep the overall significance test at its nominal level. This amounts to compute

$$B_s^j = \max(Q_i^j) \text{ or } B_s^{j\prime} = \max(Q_i^{j\prime}), \qquad j = CCA, PLS$$

and to confront it with the critical level at $\frac{\alpha}{s}\%$ in the $\chi^2_{(k)}$.

2. In the second approach, based on Cubadda *et al.* (2008, 2009), we test for the null of no autocorrelation on the aggregate $a_{t,s}^j = \sum_{i=1}^s e_{t,i}^j$ for $j = CCA, PLS$.

Notice that under the null hypothesis $a_{t,s}^j$ is a white noise in large sample whereas $a_{t,s}^j$ does not converge to an innovation process when the SCCF rank is less than s. We denote the Box-Pierce and Ljung-Box test on $a_{t,s}^j$ as A_s^j and $A_s^{j\prime}$ respectively.

Monte Carlo investigation

In order to investigate the small sample behavior of LR_s, B_s^j and $A_{s,t}^j$, we simulate a stationary VAR(1) with the following reduced rank structure

$$y_t = \alpha + \Phi_1 y_{t-1} + \varepsilon_t = \alpha + \delta_\perp C_1^{'} y_{t-1} + \varepsilon_t,$$

where $C_1^{'} = [0.5, -0.5, 0.5, -0.5, \cdots - 0.5, 0.75]$, $\delta_\perp' = [1, 1, \cdots, 1]$, α is a n dimensional vector of constant terms we generate from an uniform distribution on (0,1), and ε_t are i.i.d. $N_n(0, I_n)$.

This system has $s = n - 1$ SCCF relationships of the shape

$$\delta = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix}.$$

We first consider $n = 9$ and 25 individuals for successively $T = 50, 200$ and 600 data points corresponding, say, to 50 annual, 200 quarterly and 600 monthly observations. $M = 10.000$.

- Table 1 compares the rejection frequencies from $s = 22$ to $s = n = 25$ of the the univariate tests B_s^0, $B_s^{0\prime}$, A_s^0 and $A_s^{0\prime}$, where the superscript "0" indicates that linear combinations with known coefficients of $[\delta, \delta_\perp]$ are used in the testing procedures.
- This allows to compare the behavior of the different tests proposed in the previous section when there is not estimation uncertainty.
- The frequency for $s = 25$ is the power (size unadjusted) of the test, $s = 22, 23, 24$ is the empirical size, i.e. the frequency with which 24 vectors are not rejected.
- From Table 1 we see that the behavior of univariate tests for the null of no autocorrelation with the known combinations is very good as it might be expected with the exception in some circumstances of the Ljung-Box test for small sample sizes and large k.

Table: Size and power of common features tests statistics

		$k=2$				$k=5$			
	$T\backslash s =$	22	23	24	25	22	23	24	25
B_s^0	50	4.14	4.18	4.29	96.1	9.98	10.1	10.3	93.2
	200	4.09	4.07	4.08	100	6.15	6.19	6.25	100
	600	4.34	4.3	4.24	100	5.25	5.26	5.22	100
$B_s^{0\prime}$	50	6.13	6.07	6.19	96.8	16.7	17	17.3	94.9
	200	4.6	4.6	4.52	100	7.17	7.25	7.28	100
	600	4.48	4.43	4.4	100	5.48	5.51	5.58	100
A_s^0	50	4.45	4.60	4.45	96.40	5.92	6	5.96	93.7
	200	4.98	4.95	5	100	5.14	5.22	5.2	100
	600	4.6	4.55	4.5	100	4.8	4.82	4.77	100
$A_s^{0\prime}$	50	5.75	5.69	5.60	95.80	8.3	8.48	8.57	94.9
	200	5.26	5.21	5.24	100	5.69	5.76	5.91	100
	600	4.73	4.64	4.6	100	4.99	4.98	5.04	100

Results from Table 1 refer to an unfeasible testing procedure in the sense that we use the true common feature vectors to construct the combination that are used for univariate autocorrelation tests.

Let us now compare the results of test LR_s with those of tests based the estimated combinations obtained from either CCA or PLS.

To save space, we only report the rejection frequencies for $n = 9$ series for $s = 8$ and $s = 9$, namely the empirical size and the empirical power of these tests.

We choose a smaller dimension than the one from Table 1 because we wish to illustrate that, even with this relatively small number of series, the CCA framework already shows some large size distortions. Given the findings of Table 1 we also focus on the Box-Pierce test instead of its modified Ljung-Box version.

- From Table 2 we see that it is only with a very large sample size such as $T = 600$ and $k = 2$ that one observes roughly correct empirical size with the LR_s test. In other cases the frequency with which the null $s = 8$ is rejected is very high, e.g. 93% with $T = 50$ and $k = 2$.
- Turning to autocorrelation tests on estimated linear combinations, it clearly emerges that tests based on PLS give better results in terms of size distortion and power than tests based on CCA.
- The best strategy is A_s^{PLS}. For this procedure, there are negligible size distortions and the power if high even in small samples.
- The results of B_s^{PLS} are a bit disappointing, especially when one compares them with the case of known common feature vectors.
- Interestingly enough however, alternative identifications of the SCCF vectors give different results but we leave this issue for further investigations. For instance, normalizing the estimated vectors such that $\widehat{\delta}' = [I_s, \widetilde{\delta}_{s\times(n-s)}]$ provide tests with rejections frequencies that are very close to those obtained with known SCCF vectors.

Table: Size and power of common features tests statistics

		$k=2$		$k=5$	
	$T \backslash s =$	8	9	8	9
LR_s	50	93.2	100	100	100
	200	17.3	100	68.8	100
	600	7.97	100	16.2	100
B_s^{CCA}	50	10.2	97.3	38.4	54.7
	200	12.7	100	12.2	100
	600	14	100	13.4	100
B_s^{PLS}	50	10.2	97.9	16.4	95.7
	200	14.6	100	13.40	100
	600	15.7	100	11.7	100
A_s^{CCA}	50	3.75	7.54	8.82	10.9
	200	3.48	35.3	4.49	31.30
	600	3.25	62.2	4.20	60.9
A_s^{PLS}	50	5.17	95.5	6.19	93.1
	200	5.28	100	5.40	100
	600	5.04	100	5.24	100

- We finally simulate series from the same DGP but with $n = 25$ series.
- Only A_s^{PLS} and B_s^{PLS} are reported in Table 3 for the size $s = 24$ and the empirical power $s = 25$.
- The results confirm what we observed so far because we correctly detect with A_s^{PLS} the presence of 24 common feature vectors even with this large set of series.
- There is obviously a small size distortion for $T = 50$ and $k = 8$ (the empirical size is 7% in this worse case) but this case represents a situation in which CCA would not even be feasible in a system of 125 regressors for each variable.
- We would consequently recommend to add the A_s^{PLS} statistics, namely a Box-Pierce test on the aggregate of the PLS factors, to the traditional toolkit for common features analysis.

Table: Size and power of common features tests statistics

		$k=2$		$k=5$		$k=8$	
	$T \backslash s =$	24	25	24	25	24	25
B_s^{PLS}	50	12.0	96.40	42.2	94.9	64.3	95.6
	200	22.60	100	30	100	34.9	100
	600	27.10	100	23.9	100	22.3	100
A_s^{PLS}	50	5.20	98.6	5.72	96.70	7.06	95.50
	200	5.03	100	5.31	100	5.36	100
	600	4.75	100	5.05	100	5.22	100

.

This section presents a couple of applications, in which we apply both LR_s and A_s^{PLS} tests to some interesting empirical issues.

1. First, we examine the serial correlation properties of revision errors based on the "preliminary" data releases of the EU12 industrial production index.
2. Second, we analyze if the dynamics of the US economy become less predictable during the "Great Moderation".

Data revision:

- Data currently produced by statistical offices typically undergo a recurrent revision process resulting in different releases of the same phenomenon.
- Consequently a database consists in general of vintages of the major macroeconomic data available in real time.
- Let us denote $x_{t,t+v}$ the point estimate for x_t published in $t+v$ with $v \geq 0$. Collecting the whole series for $t = 1 \ldots T$ we have what is called a diagonal of the real time data matrix.
- We can look at the different diagonals denoted $X_{t,t+v}$ and for instance at the next vintage diagonal $X_{t,t+v+1}$. The problem to know that the data revision process brings news or noise in the data depends on whether $X_{t,t+v+1} - X_{t,t+v}$ is serially correlated or a white noise (see, e.g., Croushore 2008).

- We have considered vintage series for the EU12 industrial production index. The 16 series are in months from $X_{t,t+3}$ to $X_{t,t+18}$ and $T = 58$.
- The results show that LR_s and A_s^{PLS} provide different conclusions. The LR_s test is feasible only for $k = 2$ and it provides no evidence of SCCF, whereas the A_s^{PLS} tests suggests that approximately the half of the PLS linear combinations are white noises.

Table: Number of SCCF's (with a significance level of 0.05) on 16 diagonal vintages

	$k = 2$	$k = 4$	$k = 8$
LR_s	0	–	–
A_s^{PLS}	8	7	9

Great Moderation

- We consider a system with 18 variables, which are listed in our paper along with their transformations.
- The data are observed at the monthly frequency for the period 1959.01-2003.12.
- We report the results of the application of both LR_s and A_s^{PLS} tests for the samples 1959–1983 (pre–Great Moderation) and 1984–2003 (Great Moderation). The two tests provide opposite conclusions.

- According to the LR_s tests, we even find less evidence of SCCF's in the Great Moderation period, whereas the A_s^{PLS} tests point out that the number of PLS factors with no autocorrelation has almost doubled in the second sub-sample.
- In view of the results of the Monte Carlo experiment, we conclude that the evidence based on the LR_s statistics is likely to be spurious, since these tests suffer of sever size distortion in large dimensional systems. Instead, the A_s^{PLS} tests confirm previous findings in the literature.

Table: Number of SCCF's on 18 US economic indicators

		$k = 2$	$k = 4$	$k = 8$
1959:1983	LR_s	3	1	0
	A_s^{PLS}	2	2	1
1984:2003	LR_s	1	0	0
	A_s^{PLS}	3	4	2

Conclusion

We present new tools for detecting common cycles among a large number of variables (but not enough to rely on factor models):

- A weaker condition based on absence of autocorrelation is used. Simple Box-Pierce test on aggregates residuals works very good in simulations.
- A PLS approach makes the estimation of the common feature space feasible.
- This method can easily be generalized to accomodate for less restrictive models: PSCCF or codependence.
- Extension to cointegration under investigation.