

# Common Feature Analysis of Economic Time Series: An Overview and Recent Developments

Marco Centoni<sup>a</sup>, Gianluca Cubadda<sup>1,b</sup>

<sup>a</sup>LUMSA Università, Italy;

<sup>b</sup>Dipartimento di Economia e Finanza, Università degli Studi di Roma Tor Vergata, Italy

---

## Abstract

In this paper we overview the literature on common features analysis of economic time series. Starting from the seminal contributions by Engle and Kozicki (1993) and Vahid and Engle (1993), we present and discuss the various notions that have been proposed to detect and model common cyclical features in macroeconometrics. In particular, we analyze in details the link between common cyclical features and the reduced-rank regression model. We also illustrate similarities and differences between the common features methodology and other popular types of multivariate time series modelling. Finally, we discuss some recent developments in this area, such as the implications of common features for univariate time series models and the analysis of common autocorrelation in medium-large dimensional systems.

**Keywords:** common features, common cycles, reduced-rank regression, canonical correlation analysis, vector autoregressive models, dynamic factor models, business cycles

---

## 1. Introduction

Economic time series could be characterized by several features such as trends, cycles, seasonality, serial correlation, and so on. When a set of series possesses the same type of feature, it could be the case that a linear combination of them does not necessarily possess the feature: this is the most interesting case, for which Engle and Kozicki (1993) provided the following definition: “A feature, which is present in each of a set of series, is said to be common to those series when there exists a nonzero linear combination of these series that does not have the feature”. A well known example of common features is cointegration (Engle and Granger, 1987; Johansen, 1988): a group of series that possesses stochastic trends is cointegrated when there are some linear combinations of the variables that are stationary, i.e. do not have stochastic trends. Nowadays, there is a huge collection of special cases of common features. A comprehensive, although still partial, list includes: codependence (Gourieroux and Peaucelle, 1988; Vahid and Engle, 1997) and the scalar component model (Tiao and Tsay, 1989), when a linear combination of variables possesses shorter memory than individual series; common serial correlation (Engle and Kozicki, 1993; Vahid and Engle, 1993), when a linear combination of serially correlated series is an innovation w.r.t. the past; cotrending (Chapman and Ogaki, 1993), when a linear combination of trend-stationary time series no longer displays deterministic trend; common volatility (Engle and Kozicki, 1993; Engle and Susmel, 1993), when a linear combination of conditionally heteroskedastic time series eliminates conditional heteroskedasticity; seasonal

---

<sup>1</sup> Corresponding author: Gianluca Cubadda, Dipartimento di Economia e Finanza, Università degli Studi di Roma Tor Vergata, Via Columbia 2, 00133 Roma, Italy. E-mail: [gianluca.cubadda@uniroma2.it](mailto:gianluca.cubadda@uniroma2.it)

cointegration (Hylleberg *et al.*, 1990), when a linear combination of seasonally integrated series is nonseasonal; co-breaking (Hendry, 1999; Hendry and Massmann, 2007), when a set of series appears subject to structural breaks but a linear combination of them does not display the breaks; codependent cycles (Vahid and Engle, 1997), when a linear combination of a group of variables has shorter memory than the individual series; common nonlinearity (Anderson and Vahid, 1998), when the conditional expectation of each element of a vector time series is nonlinear w.r.t. the conditional vector but there exist a linear combination of them whose conditional expectation is linear w.r.t. the conditional vector; common seasonal cycles (Cubadda, 1999), when there exists a linear combination of seasonally differenced series which follows an MA process of low order; common panel structures (Hecq *et al.*, 2000), when there is a linear combination of the variables in a panel data which is white noise for all individuals of the panel; nonlinear cotrending (Bierens, 2000), when a linear combination of the components of a set of stationary time series around nonlinear deterministic time trends is stationary around a linear trend or a constant; polynomial common serial correlation (Cubadda and Hecq, 2001), when there exists a polynomial combination of serially correlated time series that is an innovation; long-run pure variance common feature (Engle and Marcucci, 2006), when the conditional variances of a collection of assets all depend upon a small number of variance factors; unpredictable polynomial combinations (Paruolo, 2006), when a polynomial linear combination of series integrated with different orders is an innovation; weak form of common serial correlation (Hecq *et al.*, 2006), when a linear combination of serially correlated series adjusted for the equilibrium errors is an innovation; common periodic correlation (Haldrup *et al.*, 2007), which extends the notion of common serial correlation to periodic autoregressive models. From a statistical point of view, common features imply a reduction to more parsimonious structures such as common factor representations (see, *i.a.*, Cubadda, 2007), which can often be estimated by reduced-rank regression techniques (Anderson, 1984, 1999). Imposing the common feature restrictions to the statistical model generally leads to considerable gains in both forecasting and structural analysis (Vahid and Issler, 2002).

Common features among economic time series are often predicted by economic theory. For example, in King *et al.* (1988) the solution of their macro model implies that output, consumption and investment have a common trend and a common cycle. The common stochastic trend is generated by an integrated productivity shock, while the deviation of capital stock from its steady state value determines the transitional dynamics of output, consumption and investment. Another example is Campbell (1987), where the saving path implies that disposable income and consumption cointegrate (Issler and Vahid, 2001). Further, Vahid and Engle (1993) and Issler and Vahid (2001) show that in models of aggregate consumption, either the existence of “myopic” individuals (Campbell and Mankiw, 1990) or the excess sensitivity of consumption to current income (Flavin, 1993) imply that the growth rates of consumption and income share a common cycle. Note that in the consumption model by Hall (1978) consumption and income share only a common stochastic trend. As another example, in the real business cycle model for sectoral output *à la* Long and Plosser (1983), as in Engle and Issler (1995), common cycles depend on the propagation mechanisms through the restrictions on the production function, *i.e.* technological constraints.

The rest of the paper is organized as follows. In Section 2, after introducing the general notion of common features and linking it to the reduced-rank regression model, we focus on the various forms of common cyclical features and their implication in terms of common short-run components. We also illustrate similarities and differences between these approaches and other popular types of multivariate time series modelling. Section 3 takes into account the consequences of the presence of common features for the univariate representation of multiple time series. Section 4 deals with the estimation methods of the models implied by the various form of common features, distinguishing between the

cases of small systems and medium-large systems. Finally, Section 5 draws some conclusions.

## 2. Common Features in Economic Time Series

In this section we first present and discuss the general notion of common features that was originally proposed by Engle and Kozicki (1993). We stress the link between the presence of common features and the multivariate Reduced-Rank Regression model (RRR) that was introduced by Anderson (1951). A detailed survey on this modelling may be found in Reinsel and Velu (1998). Then we focus on the common autocorrelation feature and its interplay with the notion of common cycles in the multivariate Beveridge and Nelson (1981) decomposition. Starting from the seminal work by Vahid and Engle (1993), we illustrate the various forms of common cyclical features that have been proposed in the literature and their implication in terms of common unobserved components. Finally, we illustrate similarities and differences between the common serial correlation approach and other types of multivariate time series modelling that are popular in statistics and econometrics, such the dynamic factor model (see, e.g., Stock and Watson (2011) and the references therein) and the multivariate autoregressive index model (Reinsel, 1983).

### 2.1. Common features and the reduced-rank regression model

Engle and Kozicki (1993) considered features that satisfy the following axioms:

1. The vector series  $X_t$  has (does not have) the feature if any non-singular linear transformation of  $X_t$  still has (does not have) such feature.
2. If two  $n$ -vector time series  $Y_{1t}$  and  $Y_{2t}$  do not have the feature then  $(Y_{1t} + Y_{2t})$  does not have the feature.
3. If  $Y_t$  does not have the feature but  $X_t$  has the feature then  $(Y_t + X_t)$  has the feature.

Then, any dynamic property of the data could be viewed as a special case of feature: for example, series with stochastic trends satisfies all axioms. As pointed out by Engle and Kozicki (1993) a linear combination of two series that both have the feature does not necessarily possess the feature. This is the most interesting case, and to this issue Engle and Kozicki gave particular attention by the following definition:

**Definition 1.** *A feature, which is present in each of a set of series, is said to be common to those series when there exists a nonzero linear combination of these series that does not have the feature.*

From a statistical point of view, the existence of such linear combinations can be linked to a common factor representation. Indeed, let  $Y_t$  be an  $n$ -vector time series such that

$$Y_t = Bf_t + e_t, \quad (2.1)$$

where  $(n - s)$  common factors  $f_t$  have the feature and  $e_t$  does not have the feature. Consider a  $s$ -vector  $\delta$  such that  $\delta' B = 0$ , then  $\delta' Y_t$  does not have the feature.

The main idea is that a small number of unobserved components possess a given feature and transmit it to a larger set of time series. It is then possible to combine such time series in order to cancel the influence of these unobserved components, thus removing the Common Feature (CF) from the data.

In order to illustrate the connection between CF's and RRR, let us assume that  $Y_t$  is a  $n$ -vector (weakly) stationary time series such that

$$Y_t = \beta' X_t + \vartheta' Z_t + \varepsilon_t, \quad (2.2)$$

where  $X_t$  and  $Z_t$  are, respectively, vectors of  $k$  and  $m$  stationary time series,  $\beta \neq 0$ , and  $\varepsilon_t$  are i.i.d. innovations, with  $E(\varepsilon_t) = 0$ ,  $E(\varepsilon_t \varepsilon_t') = \Sigma_{\varepsilon\varepsilon}$  (positive definite) and finite fourth moments, that are independent from both  $X_t$  and  $Z_t$ . Both  $X_t$  and  $Z_t$  may contain (linear functions of) lags of  $Y_t$  and, for the sake of simplicity, no deterministic terms are included.

Moreover, assume that:

1. Variables  $X_t$  possess the feature of interest whereas variables  $Z_t$  don't.
2. There exists a  $n \times s$  ( $s < \min\{n, k\}$ ) full-rank matrix  $\delta$  such that  $\delta' Y_t$  do not possess the feature. Then, it is said that variables  $Y_t$  have  $s$  CF's.

In view of the above assumptions we get

$$\delta' Y_t = \delta' \vartheta' Z_t + \delta' \varepsilon_t,$$

that is equivalent to

$$\delta' \beta' = 0 \Leftrightarrow \beta' = \delta_{\perp} \psi', \quad (2.3)$$

where  $\delta_{\perp}$  is a  $n \times (n - s)$  full-rank matrix such that  $\delta' \delta_{\perp} = 0$ , and  $\psi$  is a  $k \times (n - s)$  matrix. If  $n > k$ , the matrix  $\beta'$  has not full column-rank, then there exist  $(n - k)$  "trivial" CF's. Hence, we also assume that  $s < n \leq k$ , so the matrix  $\psi$  has full-rank as well.

In view of Equations (2.2) and (2.3), it is clear that the existence of  $s$  CF's is equivalent to postulate the following (partial) RRR model for series  $Y_t$

$$Y_t = \delta_{\perp} \psi' X_t + \vartheta' Z_t + \varepsilon_t. \quad (2.4)$$

The cofeature matrix  $\delta$  and the coefficient matrix  $\psi$  can be obtained as follows:

1. Obtain the partial regression model

$$y_t = \beta' x_t + \varepsilon_t,$$

where  $y_t = Y_t - \Sigma_{YZ} \Sigma_{ZZ}^{-1} Z_t$ ,  $x_t = X_t - \Sigma_{XZ} \Sigma_{ZZ}^{-1} Z_t$ ,  $\beta = \Sigma_{xx}^{-1} \Sigma_{xy}$ , and

$$\Sigma_{Ab} = E[A_t - E(A_t)]b_t',$$

for two generic stationary time series  $A_t$  and  $b_t$ .

2. Solve the following maximization problem

$$v_1 = \arg \max_{v \in \mathbb{R}^n} \left\{ \frac{\overbrace{v' \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} v}^{\beta' \Sigma_{xx} \beta}}{v' \Sigma_{yy} v} \right\} \quad \text{s.t. } v' \Sigma_{yy} v = 1.$$

Then we get the solution  $v_1 = \Sigma_{yy}^{-1/2} v_1$ , where  $v_j$  ( $j = 1, \dots, n$ ) is the eigenvector associated to the  $j^{\text{th}}$  largest eigenvalue  $\lambda_j$  ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ ) of the symmetric, semi-positive definite matrix

$$\Sigma_{yy}^{-\frac{1}{2}} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-\frac{1}{2}}.$$

Note that  $R^2(v'_i y_t | \eta'_i x_t)$ , where  $\eta = \beta v$ , is maximized for  $v = v_1$  and  $\eta = \eta_1 \equiv \Sigma_{xx}^{-1} \Sigma_{xy} v_1$  since

$$\begin{aligned} v'_1 \Sigma_{yy} v_1 &= 1, \\ \eta'_1 \Sigma_{xx} \eta_1 &= v'_1 \Sigma_{yy}^{-\frac{1}{2}} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-\frac{1}{2}} v_1 = \lambda_1. \end{aligned}$$

3. Solve the following maximization problem

$$v_j = \arg \max_{v \in \mathbb{R}^n} \left\{ \frac{v' \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} v}{v' \Sigma_{yy} v} \right\} \quad \text{s.t.} \quad \begin{cases} v'_j \Sigma_{yy} v_j = 1, \\ v'_i \Sigma_{yy} v_j = 0, \end{cases}$$

for  $i \leq j = 2, \dots, n$ .

We get the solution  $v_j = \Sigma_{yy}^{-1/2} v_j$ . Note that  $R^2(v'_j y_t | \eta'_j x_t) = \lambda_j$ , where  $\eta_j = \Sigma_{xx}^{-1} \Sigma_{xy} v_j$ . Moreover,  $R^2(v'_i y_t | \eta'_i x_t) = 0$  since

$$\eta'_j \Sigma_{xy} v_i = v'_j \Sigma_{yy}^{-\frac{1}{2}} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-\frac{1}{2}} v_i,$$

and  $v_i$  and  $v_j$  are eigenvectors corresponding to different eigenvalues.

4. Since  $\delta' \beta' = 0$ , then  $\lambda_{n-s+1} = \dots = \lambda_n = 0$ , and the cofeatures matrix  $\delta$  is obtained (up to an identification matrix) as follows

$$\delta = (v_{n-s+1}, \dots, v_n).$$

5. Since  $\lambda_1 \geq \dots \geq \lambda_{n-s} > 0$ , the coefficient matrix  $\psi$  is obtained (up to an identification matrix) as follows

$$\psi = (\eta_1, \dots, \eta_{n-s}).$$

6. Finally, the loading matrix  $\delta_{\perp}$  is given by the regression coefficients of  $y_t$  on  $\psi' x_t$ .

It is worth remarking that the eigenvalue problem

$$\Sigma_{yy}^{-\frac{1}{2}} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-\frac{1}{2}} v = v \lambda$$

is equivalent to finding the roots of the equation:

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} v = v \lambda \quad (2.5)$$

with the normalization  $v' \Sigma_{yy} v = 1$ .

One easily recognizes that the solution of the problem (2.5) coincides with the partial Canonical Correlation Analysis (CCA) between  $Y_t$  and  $X_t$  conditional to  $Z_t$  (see, e.g., Anderson, 1984).

## 2.2. Common trends and common cyclical features

A very well known example of CF is cointegration (see, e.g., Johansen (1996) and the references therein), where linear combinations of series having nonstationary stochastic trends feature are stationary.

Let us assume that the elements of a  $n$ -vector of time series  $U_t$  are integrated of order 1, denoted as  $U_t \sim I(1)$ , and that they admit the following VAR( $p$ ) representation:

$$A(L)U_t = \varepsilon_t, \quad (2.6)$$

where  $A(L) = I_n - \sum_{i=1}^p A_i L^i$ .

When variables  $U_t$  are cointegrated of order (1, 1), denoted as  $U_t \sim CI(1, 1)$ , we can rewrite the model (2.6) in the Vector Error Correction Model (VECM) representation:

$$\Gamma(L)\Delta U_t = \alpha\gamma'U_{t-1} + \varepsilon_t, \quad (2.7)$$

where  $\Gamma(L) = I_n - \sum_{i=1}^{p-1} \Gamma_i L^i$ , and  $\Gamma_i = -\sum_{j=i+1}^p A_j$ ,  $\alpha$  and  $\gamma$  are full-rank  $n \times r$ -matrices ( $r = 1, \dots, n-1$ ) and  $\gamma'U_t \sim I(0)$ .

Since  $\Delta U_t$  is a stationary stochastic process, it admits the following Wold representation:

$$\Delta U_t = C(L)\varepsilon_t,$$

where  $C(L) = I_n + \sum_{i=1}^{\infty} C_i L^i$  is such that  $\sum_{j=1}^{\infty} j|C_j| < \infty$ .

By expanding  $C(L)$  on 1 and integrating both sides of the above equation, we get the multivariate Beveridge-Nelson representation (BN; Beveridge and Nelson, 1981):

$$U_t = U_0 + C(1) \sum_{j=0}^{t-1} \varepsilon_{t-j} + \sum_{j=0}^{t-1} \tilde{C}_j \varepsilon_{t-j}, \quad (2.8)$$

where  $\tilde{C}_i = -\sum_{j>i} C_j$  for all  $i$ .

Since we know from the Engle-Granger representation theorem (Engle and Granger, 1987; Johansen, 1996) that  $C(1) = \gamma_{\perp}(\alpha'_{\perp}\Gamma\gamma_{\perp})^{-1}\alpha'_{\perp}$ , series  $U_t$  share the following  $(n-r)$  common stochastic trends:

$$\tau_t = \sum_{j=1}^t \alpha'_{\perp} \varepsilon_j.$$

Note the analogy between (2.8) and the common factor representation (2.1); the last term in (2.8) does not have the feature, while  $\tau_t$  has the feature, and the  $(n-r)$  common stochastic trends are removed by the cointegration vectors, since  $\gamma'C(1) = 0$ .

However, detrended economic time series often display clear evidence of comovements (Lucas, 1977), which cannot be due to cointegration, thus suggesting the presence of common cycles. If this is the case, we expect that there exist linear combinations of cyclical series that are not cyclical. This Common Cyclical Feature (CCF) (Engle and Kozicki, 1993; Vahid and Engle, 1993) can be interpreted as short-run equilibrium relationships, similarly to the interpretation of the cointegration relations as long-run equilibrium.

From the multivariate BN decomposition (2.8), the stochastic cycles of series  $U_t$  are:

$$\kappa_t = \tilde{C}(L)\varepsilon_t,$$

where  $\tilde{C}(L) = \sum_{i=0}^{\infty} \tilde{C}_i L^i$ .

While the presence of cointegration implies reduced-rank restrictions on the VECM parameters that are responsible for the long-run behavior of series  $U_t$ , since  $\gamma' \tau_t = 0$ , the analysis of CCF's is instead concerned with reduced-rank restrictions on the short-run VECM parameters (the polynomial matrix  $\Gamma(L)$  and the adjustment matrix  $\alpha$ ) that have interesting implications on the cycles  $\kappa_t$ .

However, differently from cointegration, there is not a unique notion of common short-run components. Indeed, also the degree of synchronicity of the common cycle plays a role in the definitions. Alternative notions of CCF impose differing reduced-rank structures to the VAR. Let us briefly review various form of CCF starting with the seminal notion proposed by Engle and Kozicki (1993).

Series  $\Delta U_t$  have  $s$  ( $s < n$ ) serial correlation common features (SCCF) iff there exists an  $n \times s$  matrix  $\delta_S$  with full column rank such that the VECM (2.7) can be rewritten into the following RRR model

$$\Delta U_t = \delta_{S\perp} \psi'_S \left( U'_{t-1} \gamma, \Delta U'_{t-1}, \dots, \Delta U'_{t-p+1} \right)' + \varepsilon_t,$$

where  $\psi_S$  is an  $(np - n + r) \times (n - s)$  matrix with full column rank (Engle and Kozicki, 1993). Since  $\delta'_S \Delta U_t = \delta'_S \varepsilon_t$ , there exist  $s$  linear combinations of series  $\Delta U_t$  that are unpredictable from the past, i.e. they are innovation processes. Since  $\delta'_S \kappa_t = 0$ , there exist  $(n - s)$  common cycles in series  $U_t$  (Vahid and Engle, 1993) that are perfectly synchronized, as a results the impulse responses functions are exactly collinear. However, one observes that due to technical reasons (e.g., seasonal adjustment) as well as economic reasons (e.g., adjustment costs, labor market rigidities), the hypothesis underlying SCCF is sometimes too strong, since SCCF is not able to detect the existence of non-contemporaneous cyclical comovements (Ericsson, 1993). Then, some less restrictive variants of the SCCF have been introduced in the literature, which we will review in turn.

Cubadda and Hecq (2001) propose the notion of polynomial serial correlation common features (PSCCF) as a measure of non-contemporaneous cyclical comovements. Non-synchronous common cycles arises, for example, in economic model of consumption with several types of consumer goods as, *i.a.*, in Vahid and Engle (1997) and Schleicher (2007); it is shown that the maximization problem of the representative agent implies a unsynchronized common cycle (codependent cycle) in the consumer goods vector.

By definition, series  $\Delta U_t$  have  $s$  PSCCF's iff there exists an  $n \times s$  matrix  $\delta_P$  with full column rank such that  $\delta'_P \Gamma_1 \neq 0$ , and the VECM (2.7) can be rewritten into the following partial RRR model

$$\Delta U_t = \Gamma_1 \Delta U_{t-1} + \delta_{P\perp} \psi'_P \left( \Delta U'_{t-2}, \dots, \Delta U'_{t-p+1}, U'_{t-1} \gamma \right)' + \varepsilon_t,$$

where  $\psi_P$  is an  $(np - 2n + r) \times (n - s)$  matrix with full column rank (Cubadda and Hecq, 2001).

In order to interpret the notion of PSCCF, Cubadda and Hecq (2001) show that there exists a first-order polynomial matrix  $\delta(L) = \delta_P - \Gamma'_1 \delta_P L$  such that

$$\delta(L)' \Delta U_t = \delta'_P \varepsilon_t.$$

Hence, PSCCF requires that there exists a first-order polynomial matrix  $\delta(L)$  such that  $\delta(L)' \Delta U_t$  is white noise. The presence of PSCCF has an interesting implication for the BN cycles of series  $U_t$ : indeed, since  $\delta(L)' \kappa_t = \delta'_P C(1) \varepsilon_t$ , the same PSCCF relationships cancel the dependence from the past of both the first differences and cycles of series  $U_t$ .

In the above definitions of CCF, the number of SCCF's or PSCCF's,  $s$ , cannot exceed the number of common trends  $(n - r)$ . In order to remove this restrictions, Hecq *et al.* (2006) proposed the notion

of weak form of SCCF (WF): series  $\Delta U_t$  have  $s$  WF's iff there exists an  $n \times s$  matrix  $\delta_W$  with full column rank such that  $\delta'_W \alpha \neq 0$ , and the VECM in (2.7) can be rewritten into the following partial RRR model

$$\Delta U_t = \alpha \gamma' U_{t-1} + \delta_{W\perp} \psi'_W \left( \Delta U'_{t-1}, \dots, \Delta U'_{t-p+1} \right)' + \varepsilon_t,$$

where  $\psi_W$  is an  $(np - n) \times (n - s)$  matrix with full column rank (Hecq *et al.*, 2006).

In order to uncover interesting implications of the WF for the BN cycle, Cubadda (2007) shows that there exists a first-order polynomial matrix  $\delta_W(L) \equiv \delta_W - (\gamma \alpha' + I_n) \delta_W L$  such that

$$\delta_W(L)' U_t = \delta'_W \varepsilon_t.$$

As consequence, since  $\delta_W(L)' \kappa_t = \delta'_W (I_n - C(1)) \varepsilon_t$ , the same WF relationships cancel the dependence from the past of both the cycles and linearly detrended levels of series  $U_t$ .

A limitations of the above methods for cyclical features analysis is that they cannot handle the possible coexistence of differing types of reduced-rank restrictions in the same vector. In order to overcome this limitation, Cubadda (2007) introduced the notion of weak form of PSCCF (WFP), which encompasses most of the existing formulations: series  $\Delta U_t$  have  $s$  WFP's iff there exists an  $n \times s$  matrix  $\delta_F$  with full column rank such that  $\delta'_F \alpha \neq 0$ ,  $\delta'_F \Gamma_1 \neq 0$ , and the VECM in (2.7) can be rewritten into the following partial RRR model

$$\Delta U_t = \alpha \gamma' U_{t-1} + \Gamma_1 \Delta U_{t-1} + \delta_{F\perp} \psi'_F \left( \Delta U'_{t-2}, \dots, \Delta U'_{t-p+1} \right)' + \varepsilon_t,$$

where  $\psi_F$  is an  $(np - 2n) \times (n - s)$  matrix with full column rank.

The WFP requires the existence of a second-order polynomial matrix  $\delta_F(L) \equiv \delta_F - (\gamma \alpha' + I_n + \Gamma'_1) \delta_F L + \Gamma'_1 \delta_F L^2$  such that

$$\delta_F(L)' U_t = \delta'_F \varepsilon_t.$$

An important implication of the WFP is that the polynomial matrix  $\delta_F(L)$  transforms the BN cycles  $\kappa_t$  into a process with shorter memory, since  $\delta_F(L)' \kappa_t \sim \text{VMA}(1)$ .

The CCF analysis was extended even to the case of series having different forms of stationarity than I(1)-ness. In particular, Cubadda (1999, 2001) explored the presence of common cycles in seasonal time series that are also integrated at (a subset of) the seasonal frequencies, whereas Paruolo (2006) focused on the case of I(2) systems.

Franchi and Paruolo (2011) offered a comprehensive theoretical analysis of the conditions of existence of the various form of CCF's and of the characterization of the CCF relations in I(0), I(1) and I(2) systems.

### 2.3. Relations with alternative multivariate time series models

It is interesting to analyze similarities and differences of the CF approach with other popular multivariate time models. For the sake of simplicity, we will refer within this subsection to the basic SCCF model, which can be formulated as

$$Y_t = \underbrace{\delta_{\perp} \psi'}_{\beta'} \underbrace{\left( Y'_{t-1}, \dots, Y'_{t-p+1} \right)'}_{X_t} + \varepsilon_t, \quad (2.9)$$

where series  $Y_t$  are assumed to be I(0).

Ahn and Reinsel (1988) proposed a variant of the basic RRR model that is called Nested Reduced-Rank AR model (NRRAR). The main assumption is that the VAR coefficient matrices have reduced ranks, which are nested each other:  $\text{Rank}(A_j) \supset \text{Rank}(A_{j+1})$  for  $j = 1, \dots, p-1$ . Then the NRRAR reads

$$Y_t = \sum_{j=1}^p \varrho_j \psi'_j Y_{t-j} + \varepsilon_t. \quad (2.10)$$

The NRRAR is a very general statistical model since it is easy to see that both the SCCF and its polynomial extensions are particular cases of Equation (2.10). However, the interpretation of the NRRAR in terms of common short-run components is more involved.

A different modelling, which is also endowed with a reduced-rank structure, is the Multivariate Autoregressive Index model (MAI) as originally proposed by Reinsel (1983). The basic version of the MAI reads

$$Y_t = \sum_{j=1}^p \phi_j \xi' Y_{t-j} + \varepsilon_t, \quad (2.11)$$

where  $\xi$  is  $n \times q$  ( $q < n$ ) full-rank matrix, and  $\phi_j$  is  $n \times q$  matrix for  $j = 1, \dots, p$ . The linear combinations  $I_t = \xi' Y_t$  are called the indexes.

Notice that the regression coefficient matrix implied by the MAI has the following structure

$$\beta' = [\phi_1 \xi', \dots, \phi_p \xi'],$$

which implies that  $\beta' \xi_{\perp} = 0$ . Hence, whereas SCCF imposes a common left null space to the VAR coefficient matrices, MAI imposes a common right null space to those matrices.

Notwithstanding both SCCF and MAI have a reduced-rank structure, the mathematical properties of these two modelling approaches are only partially similar. Indeed, although the interpretation of the canonical variates  $\psi' X_t$  and the index lags  $[\xi' Y_{t-1}, \dots, \xi' Y_{t-p}]$  is analogous, since both of them represent the relevant predictors of series  $Y_t$ , the same statement cannot be applied to the linear combinations  $\delta' Y_t$  and  $\xi'_{\perp} Y_t$ . The former are in fact innovations w.r.t. the past, whereas the latter are not, as one can easily observe by premultiplying both sides of Equation (2.11) by the matrix  $\xi'_{\perp}$ .

A specific property of the MAI is that the indexes themselves follow a VAR( $p$ ) process. Indeed, if we premultiply both sides of Equation (2.11) by the matrix  $\xi'$  we get

$$I_t = \sum_{j=1}^p \xi' \phi_j I_{t-j} + \xi' \varepsilon_t, \quad (2.12)$$

whereas linear combinations of series generated by an unrestricted VAR( $p$ ) model generally follow a VARMA process, see e.g. Cubadda *et al.* (2009) and the references therein. It is easy to see that a similar implication does not hold for the canonical variates  $\delta'_{\perp} Y_t$ .

A different approach that gained large popularity is the Dynamic Factor Model (DFM), see e.g. Stock and Watson (2011) and the references therein. As shown by Stock and Watson (2005), the DFM

can be represented in the VAR form as follows

$$D(L)Y_t = \Lambda \sum_{j=1}^p \Phi_j F_{t-j} + \varepsilon_{Yt}, \quad (2.13)$$

$$F_t = \sum_{j=1}^p \Phi_j F_{t-j} + \varepsilon_{Ft}, \quad (2.14)$$

where  $D(L)$  is a diagonal finite-order  $n \times n$  polynomial matrix,  $\Lambda$  is a  $n \times q$  matrix,  $\Phi_j$  is a  $n \times q$  matrix for  $j = 1, \dots, p$ ,  $F_t$  is a  $q$ -vector of unobserved factors, and  $(\varepsilon'_{Yt}, \varepsilon'_{Ft})'$  are innovations w.r.t. the past of both  $Y_t$  and  $F_t$ .

A key difference between the DFM and the previously considered approaches is that the former requires for inferential purposes that the number of series  $n$  is large compared to the sample size  $T$ . Indeed, as shown by Bai and Ng (2006), if the factors  $F_t$  are obtained through principal component methods, the estimated factors can be treated as observed for statistical inference on Equation (2.13) provided that  $n \rightarrow \infty$  and  $T^{1/2}/n \rightarrow 0$  as  $T \rightarrow \infty$ .

Apart from the different asymptotic frameworks, the DFM and the MAI have some degree of similarity in their mathematical formulations. Indeed, Equation (2.12) is entirely analogous to Equation (2.14), and Equation (2.13) can be seen as generalization of Equation (2.11) such that each series is endowed with an individual AR structure.

It is less obvious how to relate the DFM to the SCCF. One may notice that in the particular case that  $D(L) = I_n$ , series  $Y_t$  in Equation (2.13) have indeed the SCCF with a matrix  $\delta$  that is equal to  $\Lambda_{\perp}$ . However, lags of the individual series are typically included in empirical applications of the DFM.

### 3. Common Features and Univariate Time Series Models

It is well known that each series generated by a VAR process admits a univariate ARIMA representation, see e.g. Zellner and Palm (1974). However, the VAR models that are typically used in macroeconomic analysis would imply highly non parsimonious ARIMA models for individual time-series, whereas low order ARIMA models are empirically appropriate. This is the so-called ‘‘autoregressivity paradox’’. Cubadda *et al.* (2008, 2009) argued that the presence of common cyclical features can provide a solution to this paradox.

Indeed, let us assume that the  $n$  series of interest are generated by the following VAR( $p$ ) model

$$A(L)Y_t = \varepsilon_t.$$

The so-called Final Equations (FEs) of series  $Y_t$  can be obtained by premultiplying both sides of the VAR equation by  $A(L)^{adj}$ , the adjoint matrix associated with  $A(L)$ :

$$\det[A(L)]Y_t = A(L)^{adj} \varepsilon_t,$$

where  $\det[A(L)]$  is the determinant of the polynomial matrix  $A(L)$ .

Since  $\det[A(L)]$  is a polynomial of order  $np$  and  $A(L)^{adj}$  is matrix polynomial of order  $(n-1)p$ , it follows that elements of  $Y_t$  should admit a univariate ARMA( $np, (n-1)p$ ) representation. Hence, a typical VAR model with  $n = 5$  and  $p = 4$  would imply that each individual variable follows an ARMA(20, 16) model, which is at odds with the empirical evidence.

Table 1: Maximum ARIMA orders of univariate series generated by an  $n$ -dimensional VAR( $p$ ) with  $s$  cofeature restrictions

Model	AR order	I( $d$ )	MA order
I(0)VAR	$np$	0	$(n-1)p$
SCCF	$(n-s)p$	0	$(n-s)p$
PSCCF	$(n-s)p + s$	0	$(n-s)p + (s-1)$
CI(1, 1) VAR	$n(p-1) + r$	1	$(n-1)(p-1) + r$
SCCF	$(n-s)(p-1) + r$	1	$(n-s)(p-1) + r$
PSCCF	$(n-s)(p-1) + r + s$	1	$(n-s)(p-1) + r + s - 1$
WF	$(n-s)(p-1) + r$	1	$(n-s)(p-1) + r$

SCCF = serial correlation common feature; PSCCF = polynomial SCCF; WF = weak form.

Following Cubadda *et al.* (2009), let us assume that  $n = 3$  series are generated by the following VAR(1):

$$Y_t = \begin{bmatrix} 0.5 & -0.5 & 0.5 \\ 0.25 & -0.25 + \omega & 0.25 \\ 0.5 & -0.5 & 0.5 + \omega \end{bmatrix} Y_{t-1} + \varepsilon_t,$$

where  $Y_t = (y_{1t}, y_{2t}, y_{3t})'$ .

For the above VAR, the FEs are:

$$\det[A(L)] = 1 - 0.5L^3\omega^2 - (0.75 + 2\omega)L + (1.25\omega + \omega^2)L^2,$$

such that individual series follow ARMA(3, 2) models.

However, if  $\omega = 0$ , the VAR has reduced-rank structure

$$Y_t = \begin{bmatrix} 0.5 \\ 0.25 \\ 0.5 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \end{bmatrix} Y_{t-1} + \varepsilon_t,$$

which produces the FEs:

$$(1 - 0.75L)Y_t = \varepsilon_t + \begin{bmatrix} -0.25 & -0.5 \\ 0.25 & -1 \\ 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -0.5 \end{bmatrix} \varepsilon_{t-1}.$$

This implies that the univariate representations are parsimonious ARMA(1, 1) models with the same autoregressive parameter and cross-correlated VMA errors having a factor structure.

More generally, Table 1 summaries the reduction of the individual ARMA orders due to common features restrictions.

As one can see, the existence of CCF's provides a possible, economically meaningful, solution of the autoregressivity paradox. Notice that Table 1 provides the *maxima* ARIMA orders under CCF's restrictions. However, the orders can be even smaller due to additional restrictions on the VAR parameters, such as block-diagonal or block-triangular structures.

The presence of short-run comovements has also consequences for the VMA part of FEs. Cubadda *et al.* (2009) show that in a stationary VAR( $p$ ), the existence of  $s$  SCCF relationships implies that in the FEs the VMA coefficient matrices associated with degrees strictly larger than  $(n-s-1)p$  have a common right null space that is spanned by  $\delta_{\perp}$ . Hence, it is possible to reduce the order of the VMA component to a degree of at most  $(n-s-1)p$  instead of  $(n-s)p$ . In particular, when  $n-1 = s$ , the

FEs follow a model that is popular in the macro-panel literature: an homogeneous AR component and cross-correlated VMA errors having a factor structure.

Cubadda *et al.* (2009) illustrate this point with an empirical example. They consider the industrial production indexes of Canada and the US. Since no cointegration in log levels is found, a VAR(1) in first differences seems appropriate. The estimation by OLS delivers (standard errors in brackets)

$$\begin{bmatrix} \widehat{\Delta \ln US}_t \\ \widehat{\Delta \ln CA}_t \end{bmatrix} = \begin{bmatrix} 0.333 & 0.273 \\ (0.088) & (0.079) \\ 0.265 & 0.360 \\ (0.102) & (0.092) \end{bmatrix} \begin{bmatrix} \Delta \ln US_{t-1} \\ \Delta \ln CA_{t-1} \end{bmatrix}.$$

Theoretical FEs orders imply an ARMA(2, 1) processes. However, SCCF test statistics is in favor of  $s = 1$  ( $p$ -value = 0.31), with the estimated SCCF relationship ( $\Delta \ln US_t - 1.05\Delta \ln CA_t$ ). Statistical identification of the univariate models provides the following ARIMA(1,1,0) structures

$$\begin{aligned} \widehat{\Delta \ln US}_t &= 0.003 + 0.554\Delta \ln US_{t-1}, \\ &\quad (0.001) \quad (0.062) \\ \widehat{\Delta \ln CA}_t &= 0.004 + 0.533\Delta \ln CA_{t-1}. \\ &\quad (0.001) \quad (0.064) \end{aligned}$$

As expected, the AR coefficients are very similar. Moreover, since the estimated cofeature vector  $\hat{\delta}$  is close to  $(1 : -1)$ , and the VAR residuals have similar variances and a correlation around 0.65, the factor structure of the VMA part of the FEs may explain why the MA(1) components are empirically negligible.

#### 4. Statistical Analysis

In this section we first illustrate how to estimate the RRR models implied by the various form of CCF's when the system dimension is small. In this case, a Maximum Likelihood (ML) under the Gaussianity assumption is generally employed. However, ML methods break down when the number of regressors becomes large compared to the typical sample size in macroeconomic datasets ( $100 \leq T \leq 200$ ). The problem is due to the need of inverting large covariance matrices. Hence, we will review some recent proposals for estimating VAR model with reduced-rank restrictions when the number of series  $n$  exceeds 20.

##### 4.1. Maximum likelihood inference of small-scale multivariate models

Under the assumptions that series  $Y_t$ ,  $X_t$  and  $Z_t$  are  $I(0)$  and that the innovations  $\varepsilon_t$  are Gaussian, ML inference on model (2.2) is provided by partial CCA of series  $Y_t$  and  $X_t$  conditional on  $Z_t$ , denoted by  $\text{CanCor}\{Y_t, X_t | Z_t\}$ , see *i.a.* Anderson (1999, 2002).

In particular, the LR test statistic on the existence of  $s$  CF's is:

$$\text{LR}(s) = -T \sum_{i=n-s+1}^n \ln(1 - \hat{\lambda}_i), \quad (4.1)$$

where  $\hat{\lambda}_i$  is the  $i^{\text{th}}$  largest eigenvalue of the sample matrix  $\widehat{\Sigma}_{yy}^{-1/2} \widehat{\Sigma}_{yx} \widehat{\Sigma}_{xx}^{-1} \widehat{\Sigma}_{xy} \widehat{\Sigma}_{yy}^{-1/2}$ . Under the null hypothesis that  $s$  CF's are present, the test statistic (4.1) is asymptotically distributed as

$$\text{LR}(s) \xrightarrow{d} \chi^2[s(k - n + s)]. \quad (4.2)$$

Table 2: Canonical correlations and CCF's

Model	$X'_t$	$Z'_t$
SCCF	$(\Delta U'_{t-1}, \dots, \Delta U'_{t-p+1}, U'_{t-1}\gamma)$	
WF	$(\Delta U'_{t-1}, \dots, \Delta U'_{t-p+1})$	$U'_{t-1}\gamma$
PSCCF	$(\Delta U'_{t-2}, \dots, \Delta U'_{t-p+1}, U'_{t-1}\gamma)'$	$\Delta U'_{t-1}$
WFP	$(\Delta U'_{t-2}, \dots, \Delta U'_{t-p+1})$	$(D'_t, \Delta U'_{t-1}, U'_{t-1}\gamma)$

SCCF = serial correlation common feature; WF = weak form; PSCCF = polynomial SCCF; WFP = weak form of PSCCF.

The ML estimator of the cofeatures matrix  $\delta$  is obtained (up to an identification matrix) as follows

$$\widehat{\delta} = (\widehat{v}_{n-s+1}, \dots, \widehat{v}_n),$$

where  $\widehat{v}_j = \widehat{\Sigma}_{yy}^{-1/2}\widehat{v}_j$  and  $\widehat{v}_j$  ( $j = 1, \dots, n$ ) is the eigenvector associated to the  $j^{th}$  largest sample eigenvalue  $\widehat{\lambda}_j$ .

The ML estimator of the coefficient matrix  $\psi$  is obtained (up to an identification matrix) as follows

$$\widehat{\psi} = (\widehat{\eta}_1, \dots, \widehat{\eta}_{n-s}),$$

where  $\widehat{\eta}_j \equiv \widehat{\Sigma}_{xx}^{-1}\widehat{\Sigma}_{xy}\widehat{v}_j$ .

The ML estimator of the loading matrix  $\delta_{\perp}$  is finally obtained by the regression coefficients of  $y_t$  on  $\widehat{\psi}'x_t$ .

The degrees of freedom of the test statistic (4.1) are obtained as follows. Under the alternative hypothesis, the matrix  $\beta$  is composed by  $n \times k$  elements. Under the null hypothesis, we have that

$$\beta = \underbrace{\psi}_{k \times (n-s)} \left( \underbrace{\delta_{\perp,1}}_{(n-s) \times (n-s)} : \underbrace{\delta'_{\perp,2}}_{(n-s) \times s} \right).$$

Without loss of generality, we assume that  $\delta_{\perp,1}$  has full-rank. Then we can write

$$\beta = \psi\delta_{\perp,1}\delta_{\perp,1}^{-1}\delta'_{\perp} = \widetilde{\psi}\delta'_{\perp},$$

where  $\widetilde{\delta}'_{\perp} = (I_s, \widetilde{\delta}'_{\perp,2})$ . Hence, the matrix  $\beta$  is composed by  $(n-s) \times (s+k)$  elements. It follows that under the null hypothesis  $s(k-n+s)$  restrictions are imposed.

In order to conduct inference on the various forms of CCF's, a two-step procedure is usually employed (see, *i.a.*, Vahid and Engle (1993), Ahn (1997), Cubadda and Hecq (2001), Hecq *et al.* (2006), and Cubadda (2007)). First, the cointegrating vectors are estimated by ML on the VECM without imposing any CCF restrictions. This can be done by using the procedures suggested in Johansen (1988) or Ahn and Reinsel (1990). Second, the cointegration matrix  $\gamma$  is fixed to its ML estimate, and inference on CCF's is carried on. Since the ML estimator of the cointegration matrix  $\gamma$  converges at rate  $T$  irrespective of the existence of constraints in the short-run polynomial matrix  $\Gamma(L)$  (see, e.g., Johansen (1996)), the asymptotic distribution of the test statistics (4.1) remains the same as in (4.2).

Hence, ML inference on the various forms of common features is obtained by solving  $\text{CanCor}\{\Delta U_t, X_t | Z_t\}$  for proper choices of the variables  $X_t$  and  $Z_t$  as detailed in Table 2.

For each of the models in Table 2, under the null that  $s$  common features of a given form exist, the test statistic (4.1) is asymptotically distributed as a  $\chi^2(d)$  as detailed in Table 3 (see, *i.a.*, Velu *et al.* (1986), Anderson (2002)).

Table 3: Tests for common features

Model	Degress of freedom $d$
SCCF	$s \times (n(p-2) + r + s)$
WF	$s \times (n(p-2) + s)$
PSCCF	$s \times (n(p-3) + r + s)$
WFP	$s \times (n(p-3) + s)$

SCCF = serial correlation common feature; WF = weak form; PSCCF = polynomial SCCF; WFP = weak form of PSCCF.

Table 4: Estimators of the common features vectors and RRR coefficients

Model	$(\widehat{\eta}_1, \dots, \widehat{\eta}_s)$	$(\widehat{v}_{s+1}, \dots, \widehat{v}_n)$
SCCF	$\widehat{\delta}_S$	$\widehat{\psi}_S$
WF	$\widehat{\delta}_W$	$\widehat{\psi}_W$
PSCCF	$\widehat{\delta}_P$	$\widehat{\psi}_P$
WFP	$\widehat{\delta}_F$	$\widehat{\psi}_F$

SCCF = serial correlation common feature; WF = weak form; PSCCF = polynomial SCCF; WFP = weak form of PSCCF.

Moreover, optimal estimates of both the common features vectors and (partial) RRR coefficients are then obtained as described in Table 4.

Finally, the remaining parameters of the various RRR models are estimated by OLS after fixing the matrices  $\psi$ 's to their estimated values.

Notice that the two step procedures previously illustrated do not maximize the Gaussian likelihood, although the estimated parameters have the same distribution as the optimal ones. Centoni *et al.* (2007) suggested an iterative algorithm for computing the ML estimates. The main idea is to switch between the estimation of the cointegration matrix  $\gamma$  having fixed the short-run parameters and the estimation of the short-run parameters having fixed the cointegration matrix.

The merits of imposing CCF restrictions in VAR models are investigated in several Monte Carlo studies. Vahid and Issler (2002) documented the advantages of simultaneously choosing the order  $p$  of the VAR model and the number  $s$  of SCCF restrictions for both forecasting and structural analysis. Athanopoulos *et al.* (2011) extended the analysis to cointegrated systems. They proposed to simultaneously choose the cointegration rank  $r$  with  $p$  and  $s$ . Using the switching algorithm in Centoni *et al.* (2007) for estimation and a novel criterion for model identification, significant gains in forecasting accuracy are found both in simulations and empirical applications.

#### 4.2. Methods for medium-large dimensional time series

The ML approach discussed so far is typically applied to small scale multivariate models, i.e. in situations where the number of series  $n$  is not larger than five and the sample size  $T$  rarely exceeds 200. However, it is of obvious interest to apply the common features framework to medium-large VAR models, i.e. when the  $n$  ranges from 10 to 40. Indeed, Koop (2013) *i.a.* shows that medium-large VAR's outperform small-scales models in forecasting the key US macroeconomic variables. However, it is clear that when  $n$  is large ML inference suffers of the curse of dimensionality problem. Indeed, Cubadda and Hecq (2011) show by simulations that in a VAR(1) with  $n = 25$  variables the LR test exhibits severe size distortions even with a large sample size as  $T = 600$ .

The main reason why ML inference performs poorly in high-dimensional systems is that the inversion of large variance-covariance matrices is required. Hence, most of methods that have been proposed to use RRR with large  $n$  try to "regularize" these covariance matrices in different ways.

The first attempt in this direction can be dated back to Vinod (1976), who proposed the canonical ridge model. When applied to the SCCF modelling, this approach would require to substitute the

eigenvalues and eigenvectors that are used for CCA with those of the matrix

$$\left(\widehat{\Sigma}_{yy} + \kappa_y I_n\right)^{-1} \widehat{\Sigma}_{yx} \left(\widehat{\Sigma}_{xx} + \kappa_x I_{np}\right)^{-1} \widehat{\Sigma}_{xy}, \quad (4.3)$$

where  $\kappa_y$  and  $\kappa_x$  are two positive scalars,  $\widehat{\Sigma}_{yy}$  and  $\widehat{\Sigma}_{xx}$  are respectively the sample variance-covariance of series  $Y_t$  and  $X_t$  in Equation (2.9), and  $\widehat{\Sigma}_{xy}$  is the matrix of the covariances between the elements of  $Y_t$  and  $X_t$ . It is clear that the matrices that appear in parentheses in Equation (4.3) are invertible even when  $T > np$  because the positive quantities  $\kappa_y$  and  $\kappa_x$  are respectively added to all the eigenvalues of  $\widehat{\Sigma}_{yy}$  and  $\widehat{\Sigma}_{xx}$ . Vinod (1976) proposed various criteria to choose the values for  $\kappa_y$  and  $\kappa_x$  in empirical applications. Notwithstanding its potential interest for high-dimensional time series analysis, to the best of our knowledge the canonical ridge approach has never been applied to either SCCF or other multivariate dynamic models.

Cubadda and Hecq (2011) suggested a Partial Least Squares (PLS) approach to test and impose the SCCF restriction to a VAR even when CCA is not feasible due a lack of degrees of freedom. PLS 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 detailed survey. The idea is to consistently estimate the SCCF matrix  $\delta$  as the eigenvectors associated with the  $s$  smallest eigenvalues of the matrix

$$\widehat{D}_{yy}^{-1} \widehat{\Sigma}_{yx} \widehat{D}_{xx}^{-1} \widehat{\Sigma}_{xy},$$

where  $D_{yy}$  and  $D_{xx}$  are diagonal matrices having the diagonal elements of, respectively,  $\Sigma_{yy}$  and  $\Sigma_{xx}$ . Since PLS require to invert diagonal matrices only, this method can provide estimates of  $\delta$  (up to an identification matrix) that are less disperse and more numerically stable than those coming from CCA when the dimension of  $X_t$  approaches the sample size  $T$ .

In order to consistently estimate the factor weights  $\psi$ , it is necessary to take the additional assumption that the columns of the matrix  $\psi_S$  are equal to  $(n - s)$  distinct eigenvectors of the matrix  $D_{xx}^{-1} \Sigma_{xx}$ . Then we get

$$D_{xx}^{-1} \Sigma_{xy} D_{yy}^{-1} \Sigma_{yx} = \psi_S V_\psi \delta'_S \perp D_{yy}^{-1} \Sigma_{yx},$$

where  $V_\psi$  is the diagonal matrix of the  $(n - s)$  eigenvalues of the matrix  $D_{xx}^{-1} \Sigma_{xx}$  that are associated with the eigenvectors  $\psi_S$ , which implies that the matrix  $\psi_S$  lies in the space generated by the eigenvectors associated with the positive eigenvalues of the matrix  $D_{xx}^{-1} \Sigma_{xy} D_{yy}^{-1} \Sigma_{yx}$ .

In order to detect the presence of SCCF in large dimensional systems, Cubadda and Hecq (2011) proposed to replace the condition that a linear combination of variables must be orthogonal to the past, namely

$$E(\delta' Y_t | X_{t-1}) = 0$$

with the one of absence of autocorrelation, namely

$$E\left(\delta'_i Y_t \mid \left[ Y'_{t-1} \delta_i, \dots, Y'_{t-p} \delta_i \right]'\right) = 0, \quad i = 1, 2, \dots, s, \quad (4.4)$$

where  $\delta = [\delta_1, \dots, \delta_s]$ . This drastically decreases the number of restrictions to be imposed under the null hypothesis, thus making a test for (an implication of) SCCF feasible even when CCA is not. Condition (4.4) can be verified by means of univariate tests for no serial correlation of each  $\delta'_i \Delta Y_t$  having fixed  $\delta$  to its PLS estimate.

Having controlled for the overall size of the test when different values of  $s$  are involved, Cubadda and Hecq (2011) show by simulations that the proposed testing procedure has negligible size distortions and remarkable power with  $n = 9, 25$  and small sample size as  $T = 50$ .

Carriero *et al.* (2011) proposed a Bayesian approach to estimate large VAR's with reduced-rank restrictions. In particular, the Reduced Rank Posterior (RRP) method goes as follows. First, estimate an unrestricted VAR by implementing the Kadiyala and Karlsson (1997) version of the Minnesota prior (see, *i.a.*, Litterman, 1986). where the estimates of elements of  $\beta$  in (2.9) are shrunk towards zeros. This approach is computationally convenient since the posterior mean of  $\beta$  can be obtained by OLS on a system augmented with proper sets of dummy variables. Second, compute the singular value decomposition of the posterior mean of  $\beta$  and obtain a reduced-rank approximation of it by retaining the  $(n - s)$  largest eigenvalues and the associated vectors only.

Carriero *et al.* (2011) compare the RRP with several alternatives, including a more elaborated Bayesian Reduced-Rank VAR (BRR) method, in a forecasting exercise where the reference model is a VAR(1) for  $n = 52$  variables and a rolling window of 120 monthly observations is used for estimation. The empirical results show that BRR and RRP outperform the competitors for intermediate and long forecasting horizons, whereas univariate AR models are hard to beat for shorter horizons.

Bernardini and Cubadda (2015) proposed to regularize the estimate of the autocorrelation matrix prior on performing CCA. In particular, they suggest to use, in place of the natural estimator, a proper shrinkage estimator of the covariance matrix of  $w_t = [y_t', x_t']'$ :

$$\tilde{\Sigma}_{ww} = \rho \hat{D}_{ww} + (1 - \rho) \hat{\Sigma}_{ww},$$

where  $\rho \in [0, 1]$ . Then a Regularized version of CCA (RCCA) can be obtained by solving the eigenvector equation

$$\tilde{\Sigma}_{xx}^{-1} \tilde{\Sigma}_{xy} \tilde{\Sigma}_{yy}^{-1} \tilde{\Sigma}_{yx} \tilde{\eta} = \tilde{\lambda} \tilde{\eta}.$$

Notice that when  $\rho = 1$  the full-rank regression case coincides with  $n$  univariate white noises, whereas when  $\rho = 0$  one gets the usual CCA solution. Hence, RCCA can be seen as a frequentist analogous of the Kadiyala-Karlsson version of the Minnesota prior, as it shrinks elements of  $\beta$  towards zero as well. In order to choose the value of  $\rho$ , Bernardini and Cubadda (2015) follow the approach of Ledoit and Wolf (2003), which requires to minimize a risk function based on the Frobenious norm of the difference between the shrinkage estimator  $\tilde{\Sigma}_{ww}$  and the covariance matrix  $\Sigma_{ww}$ . In particular, they provide a feasible estimator of the optimal  $\rho$  to the case that data are generated by a vector stationary process. Remarkably, such estimator converges to zero at rate  $T$ , hence RCCA is asymptotically equivalent to CCA.

Bernardini and Cubadda (2015) document, both by simulations and empirical applications, that RCCA improves both forecasting and estimation of structural parameters over traditional medium-size ( $n = 20$  and  $p = 2$ ) macroeconomic methods.

## 5. Conclusions

The main goal of this survey was to create a “common thread” between various topics, related to each other by the idea of modelling various forms of comovements that are typically observed in economic time series and often predicted by economic theory. From a statistical point of view, common features imply a reduction to more parsimonious structure such as common factor representation: a small number of unobserved components possesses a given feature and transmits it to a larger set of

economic time series. As we have seen, RRR is often the solution to the inferential problem. Due to the large amount of literature on common features, we focused the discussion on common cyclical features. However, we have tried to take into account recent developments as the implications of common features for univariate time series models and the statistical issues that arise when the number of the variables is fairly large.

The main drawback of the methods discussed in this paper is that different features are evaluated separately; one important exception is the unifying framework for analyzing common cyclical features that we discussed in Subsection 2.2. Then, a major goal ahead is to develop estimation and testing procedures that allow for the joint identification of several forms of common features using an integrated statistical approach.

Another challenge for future research is the extension of common cycle analysis to nonstationary medium-large dimensional systems. Indeed, all the methods that are presented in Subsection 4.2 require that series are individually stationarized prior to the multivariate analysis. This is clearly a methodological limitation since possible cointegration relations are ignored. Hence, it is still missing a statistical modelling that allows for the simultaneous analysis of common trends and cycles when the number of time series is not small.

### Acknowledgement

We wish to thank an associate editor as well as two anonymous referees for helpful comments. The usual disclaimers apply.

### References

- Ahn, S. K. (1997). Inference of vector autoregressive models with cointegration and scalar components, *Journal of the American Statistical Association*, **92**, 350–356.
- Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time-series, *Journal of the American Statistical Association*, **83**, 849–856.
- Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association*, **85**, 813–823.
- Anderson, H. M. and Vahid, F. (1998). Testing multiple equation systems for common nonlinear components, *Journal of Econometrics*, **84**, 1–36.
- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics*, **22**, 327–351.
- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, (2nd ed.), John Wiley & Sons, New York.
- Anderson, T. W. (1999). Asymptotic theory for canonical correlation analysis, *Journal of Multivariate Analysis*, **70**, 1–29.
- Anderson, T. W. (2002). Canonical correlation analysis and reduced rank regression in autoregressive models, *Annals of Statistics*, **30**, 1134–1154.
- Athanasopoulos, G., de Carvalho Guillén, O. T., Issler, J. V. and Vahid, F. (2011). Model selection, estimation and forecasting in VAR models with short-run and long-run restrictions, *Journal of Econometrics*, **164**, 116–129.
- Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions, *Econometrica*, **74**, 1133–1150.
- Bernardini, E. and Cubadda, G. (2015). Macroeconomic forecasting and structural analysis through regularized reduced-rank regression, *International Journal of Forecasting*, **31**, 682–691.

- Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory component with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics*, **7**, 151–174.
- Bierens, H. J. (2000). Nonparametric nonlinear cotrending analysis, with an application to interest and inflation in the United States, *Journal of Business & Economic Statistics*, **18**, 323–337.
- Campbell, J. Y. (1987). Does saving anticipate declining labor income - An alternative test of the permanent income hypothesis, *Econometrica*, **55**, 1249–1273.
- Campbell, J. Y. and Mankiw, N. G. (1990). Permanent income, current income, and consumption, *Journal of Business & Economic Statistics*, **8**, 265–279.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2011). Forecasting large datasets with Bayesian reduced rank multivariate models, *Journal of Applied Econometrics*, **26**, 735–761.
- Centoni, M., Cubadda, G. and Hecq, A. (2007). Common shocks, common dynamics, and the international business cycle, *Economic Modelling*, **24**, 149–166.
- Chapman, D. A. and Ogaki, M. (1993). Cotrending and the stationarity of the real interest rate, *Economics Letters*, **42**, 133–138.
- Cubadda, G. (1999). Common cycles in seasonal non-stationary time series, *Journal of Applied Econometrics*, **14**, 273–291.
- Cubadda, G. (2001). Common features in time series with both deterministic and stochastic seasonality, *Econometric Reviews*, **20**, 201–216.
- Cubadda, G. (2007). A unifying framework for analyzing common cyclical features in cointegrated time series, *Computational Statistics & Data Analysis*, **52**, 896–906.
- Cubadda, G. and Hecq, A. (2001). On non-contemporaneous short-run co-movements, *Economics Letters*, **73**, 389–397.
- Cubadda, G. and Hecq, A. (2011). Testing for common autocorrelation in data-rich environments, *Journal of Forecasting*, **30**, 325–335.
- Cubadda, G., Hecq, A. and Palm, F. C. (2008). Macro-panels and reality, *Economics Letters*, **99**, 537–540.
- Cubadda, G., Hecq, A. and Palm, F. C. (2009). Studying co-movements in large multivariate data prior to multivariate modelling, *Journal of Econometrics*, **148**, 25–35.
- Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica*, **55**, 251–276.
- Engle, R. F. and Issler, J. V. (1995). Estimating common sectoral cycles, *Journal of Monetary Economics*, **35**, 83–113.
- Engle, R. F. and Kozicki, S. (1993). Testing for common features, *Journal of Business & Economic Statistics*, **11**, 369–380.
- Engle, R. F. and Marcucci, J. (2006). A long-run pure variance common features model for the common volatilities of the Dow Jones, *Journal of Econometrics*, **132**, 7–42.
- Engle, R. F. and Susmel, R. (1993). Common volatility in international equity markets, *Journal of Business & Economic Statistics*, **11**, 167–176.
- Ericsson, N. R. (1993). Comment: Testing for common features, *Journal of Business & Economic Statistics*, **11**, 380–383.
- Flavin, M. (1993). The excess smoothness of consumption - identification and interpretation, *Review of Economic Studies*, **60**, 651–666.
- Franchi, M. and Paruolo, P. (2011). A characterization of vector autoregressive processes with common cyclical features, *Journal of Econometrics*, **163**, 105–117.
- Gourieroux, C. S. and Peaucelle, I. (1988). *Detecting a long run relationship (with an application to*

- the p.p.p. hypothesis*), Centre d'études prospectives d'économie mathématique appliquées à la planification, Paris.
- Haldrup, N., Hylleberg, S., Pons, G. and Sansó, A. (2007). Common periodic correlation features and the interaction of stocks and flows in daily airport data, *Journal of Business & Economic Statistics*, **25**, 21–32.
- Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence, *Journal of Political Economy*, **86**, 971–987.
- Hecq, A., Palm, F. C. and Urbain, J. P. (2000). Testing for common cyclical features in nonstationary panel data models, *Advances in Econometrics*, **15**, 131–160.
- Hecq, A., Palm, F. C. and Urbain, J. P. (2006). Common cyclical features analysis in VAR models with cointegration, *Journal of Econometrics*, **132**, 117–141.
- Hendry, D. F. (1999). Co-breaking. In M. P. Clements and D. F. Hendry (Eds.), *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge, MA.
- Hendry, D. F. and Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature, *Journal of Business & Economic Statistics*, **25**, 33–51.
- Hylleberg, S., Engle, R. F., Granger, C. W. and Yoo, B. S. (1990). Seasonal integration and cointegration, *Journal of Econometrics*, **44**, 215–238.
- Issler, J. V. and Vahid, F. (2001). Common cycles and the importance of transitory shocks to macroeconomic aggregates, *Journal of Monetary Economics*, **47**, 449–475.
- Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**, 231–254.
- Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford, UK.
- Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models, *Journal of Applied Econometrics*, **12**, 99–132.
- King, R. G., Plosser, C. I. and Rebelo, S. T. (1988). Production, growth and business cycles: II. New Directions, *Journal of Monetary Economics*, **21**, 309–341.
- Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs, *Journal of Applied Econometrics*, **28**, 177–203.
- Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance*, **10**, 603–621.
- Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: Five years of experience, *Journal of Business & Economic Statistics*, **4**, 25–38.
- Long, J. B. and Plosser, C. I. (1983). Real business cycles, *The Journal of Political Economy*, **91**, 39–69.
- Lucas, R. E. (1977). Understanding business cycles, *Carnegie-Rochester Conference Series on Public Policy*, **5**, 7–29.
- Paruolo, P. (2006). Common trends and cycles in I(2) VAR systems, *Journal of Econometrics*, **132**, 143–168.
- Reinsel, G. (1983). Some results on multivariate autoregressive index models, *Biometrika*, **70**, 145–156.
- Reinsel, G. C. and Velu, R. P. (1998). *Multivariate Reduced Rank Regression*, Springer, New York.
- Rosipal, R. and Krämer, N. (2006). Overview and recent advances in partial least squares, In C. Saunders et al. (Eds.), *Subspace, Latent Structure and Feature Selection*, Springer, Berlin, 34–51.
- Schleicher, C. (2007). Codependence in cointegrated autoregressive models, *Journal of Applied*

- Econometrics*, **22**, 137–159.
- Stock, J. H. and Watson, M. W. (2005). *Implications of dynamic factor models for VAR analysis (No. w11467)*, Cambridge, National Bureau of Economic Research, MA.
- Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook of Economic Forecasting*, Oxford University Press, New York, 35–59.
- Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series, *Journal of the Royal Statistical Society. Series B (Methodological)*, **51**, 157–213.
- Vahid, F. and Engle, R. F. (1993). Common trends and common cycles, *Journal of Applied Econometrics*, **8**, 341–360.
- Vahid, F. and Engle, R. F. (1997). Codependent cycles, *Journal of Econometrics*, **80**, 199–221.
- Vahid, F. and Issler, J. V. (2002). The importance of common cyclical features in VAR analysis: A Monte-Carlo study, *Journal of Econometrics*, **109**, 341–363.
- Velu, R. P., Reinsel, G. C. and Wichern, D. W. (1986). Reduced rank models for multiple time series, *Biometrika*, **73**, 105–118.
- Vinod, H. D. (1976). Canonical ridge and econometrics of joint production, *Journal of Econometrics*, **4**, 147–166.
- Zellner, A. and Palm, F. (1974). Time series analysis and simultaneous equation econometric models, *Journal of Econometrics*, **2**, 17–54.

Received September 9, 2015; Revised September 12, 2015; Accepted September 12, 2015