Università di Roma Tor Vergata

CEIS Tor Vergata Research Papers

ISSN 2610-931X

CEIS Tor Vergata

RESEARCH PAPER SERIES

Vol. 23, Issue 7, No. 611 – September 2025

VAR Models With An Index Structure: A Survey With New Results

Gianluca Cubadda

VAR models with an index structure: A survey with new results

Gianluca Cubadda*
Università di Roma "Tor Vergata"

September 19, 2025†

Abstract

The main aim of this paper is to review recent advances in the multivariate autoregressive index model [MAI], originally proposed by Reinsel (1983), and their applications to economic and financial time series. MAI has recently gained momentum because it can be seen as a link between two popular but distinct multivariate time series approaches: vector autoregressive modeling [VAR] and the dynamic factor model [DFM]. Indeed, on the one hand, the MAI is a VAR model with a peculiar reduced-rank structure; on the other hand, it allows for identification of common components and common shocks in a similar way as the DFM. The focus is on recent developments of the MAI, which include extending the original model with individual autoregressive structures, stochastic volatility, time-varying parameters, high-dimensionality, and cointegration. In addition, new insights on previous contributions and a novel model are also provided.

Keywords: Multivariate autoregressive index models, vector autoregressive models, dynamic factor models, reduced-rank regression.

*Universita' di Roma "Tor Vergata", Dipartimento di Economia e Finanza, Via Columbia 2, 00133 Roma, Italy. Email: gianluca.cubadda@uniroma2.it.

†Previous versions of this paper were presented at the intermediate workshop on methodological and computational issues in large-scale time series models for economics and finance in Messina, the Villa Mondragone time series symposium in honour of Marco Lippi in Monte Porzio Catone (Rome), the 11^{th} ICEEE in Palermo, and the final workshop on methodological and computational issues in large-scale time series models for economics and finance in Monte Porzio Catone (Rome). The author thanks the participants for their helpful comments and suggestions. The usual disclaimers apply. The financial support of MUR under the 20223725WE (PRIN 2022) grant is gratefully acknowledged.

1 Introduction

The Vector Auto-Regressive model [VAR] and the Dynamic Factor Model [DFM] are arguably among the most popular tools for analyzing economic and financial variables over time. Since the seminal contribution of Sims (1980), VARs have been theoretically extended and practically implemented to forecast, perform structural analysis, and detect comovements in multivariate time series. DFMs were introduced more recently (Forni et al. (2000), Forni and Lippi (2001), Stock and Watson (2002a), Stock and Watson (2002b), Bai and Ng (2002), and Bai (2003)), but rapidly contested the role of the workhorse in empirical macroeconomics.

The main reason for the success of the DFM is two-fold. First, it allows handling a much larger number of variables than those that are generally employed in traditional small-scale VARs, thus potentially boosting forecasting accuracy and solving the informational deficiency problems that arise in structural analyses when the agent's information set is richer than the econometrician's information set (see, e.g., Forni and Gambetti (2014)). Second, DFM allows one to disentagle the shocks that drive the common components of several time series and recover the structural shocks from those common shocks. Hence, in structural DFMs, the number of shocks is smaller than the number of variables (Forni et al. (2009)), which is in line with dynamic stochastic general equilibrium models (see Fernández-Villaverde et al. (2016) and the references therein) and, more generally, with the standard macroeconomic view that a small number of shocks drives aggregate fluctuations.

Efforts have recently been made to endow the VAR with the above-mentioned features of the DFM. On the one hand, shrinkage estimators have been proposed for medium-large VARs, both from a Bayesian perspective (e.g., Bańbura et al. (2010), Koop (2013), and Carriero et al. (2015)) and from a classical standpoint (e.g., Hsu et al. (2008), Kock and Callot (2015), and Hecq et al. (2023)). On the other hand, the Multivariate Autoregressive Index (MAI) model – originally proposed by Reinsel (1983) as a convenient approach to dimension reduction in stationary VARs – has recently gained renewed attention.[1].

Late advances have shown that MAI and its variants allow for both forecasting variables and identifying shocks analogously to the DFM but without encountering some issues in model identification and statistical inference that characterize the latter, such as the requirement that the number of variables diverges at a given rate and the need of specific assumptions on both the correlation structure of the idiosyncratic components and the factor loadings (see, *i.a.*, Bai (2003), and Bai and Ng (2006)). Moreover, VARs with index structures have been shown to be able to accommodate features such as stochastic volatility (Carriero et al. (2022)) and time-varying parameters (Cubadda et al. (2025)), which are not easy to handle within the DFM framework.

The MAI falls within reduced-rank VARs, a general class of models that include as special cases both the cointegrated VAR (see Johansen (1995) and the references therin) and the Common Serial Correlation [CSC] models (see Cubadda and Hecq (2022b) and the references therin). Although CSC models and MAI have similar mathematical formulations, their respective goals and properties are rather different; whereas the former are based on the existence of (possibly dynamic) linear combinations of autocorrelated time series that are white noise (Engle and Kozicki (1993), Vahid and Engle (1993), Cubadda (2007), Carriero et al. (2011), Cubadda and Hecq (2011), Bernardini and Cubadda (2015)), VARs with an index structure assume that there is a limited number of channels through which information from the past is transmitted to the variables of interest.

The main aim of this paper is twofold. First, recent developments in MAI are reviewed, such as the structural MAI (Carriero et al. (2016)), the vector heterogeneous index model for realized volatilities (Cubadda

[1] At the end of 2024, the annual citation rate of Reinsel (1983) in Scopus has increased by about 54% in the last 9 years, with the majority of recent citations coming from econometric journals.

et al. (2017)), as well as augmentations of the original model with individual autoregressive structures (Cubadda and Guardabascio (2019)), stochastic volatility (Carriero et al. (2022)), time-varying parameters (Cubadda et al. (2025)), high-dimensionality (Cubadda and Hecq (2022a)), and cointegration (Cubadda and Mazzali (2024)). Second, new results are provided in terms of representation theory for the various models, and a novel modeling is proposed, namely the cointegrated index-augmented autoregressive model, which combines and extends the results in Cubadda and Guardabascio (2019) and Cubadda and Mazzali (2024).

This paper is organized as follows. Focusing on representation theory, Section 2 reviews previous contributions and provides new insights into some of them. Section 3 presents the new model and deals with its estimation, whereas some details of the estimation procedure are relegated to the Appendix. Finally, Section 4 provides some conclusions.

2 VAR Models with an Index Structure

In this Section we review the models that are rooted from the original MAI formulation and provide new results on representation theory of some of them. Analogies and differences with the DFM are discussed in detail. Estimation and identification issues are also covered.

2.1 The Stuctural Multivariate Autoregressive Index Model

Let us assume that the $n-$vector time series $Y_t = (y_{1t}, ..., y_{nt})'$ is generated by the following stationary VAR(p) model:

$$\Phi(L)Y_t = \varepsilon_t, \quad t = 1...T, \tag{1}$$

where L is the lag operator, $\Phi(L) = I_n - \sum_{j=1}^{p} \Phi_j L^j$, and ε_t is a vector or n errors with $\mathrm{E}(\varepsilon_t \varepsilon_t') = \Sigma$ (positive definite) and finite fourth moments, $\mathrm{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, and $\mathcal{F}_t$ is the natural filtration of the process Y_t. For simplicity, deterministic elements are ignored.

The key assumption of MAI (Reinsel (1983)) is the following:

Assumption 1 *It holds*

$$[\Phi_1', \ldots, \Phi_p']' = [\alpha_1', \ldots, \alpha_p']'\omega',$$

where ω is a full-rank $n \times q-$matrix with $q < n$, and α_j is a $n \times q-$matrix for $j = 1, \ldots, p$.

Under Assumption 1, Model (1) can be rewritten as

$$Y_t = \sum_{j=1}^{p} \alpha_j \underbrace{\omega' Y_{t-i}}_{f_{t-i}} + \varepsilon_t \tag{2}$$

where linear combinations $f_t = \omega' Y_t$ are called the indexes. The MAI has at most $nq(p + 1) - q^2$ mean parameters, which implies a significant dimension reduction when p is small w.r.t. n.[2]

By premultiplying with ω' both sides of Equation (2) we get

$$f_t = \sum_{j=1}^{p} \omega' \alpha_j f_{t-j} + \omega' \varepsilon_t, \tag{3}$$

which shows that the indexes follow a VAR(p) process and not a VARMA, as is generally the case for linear combinations of elements of a VAR (see Cubadda et al. (2009) and the references therein).

[2] Indeed, the matrix ω, once identified through normalizing restrictions, has $q(n - q)$ free parameters.

Remark 1 *In view of Equations (2) and (3), the MAI resembles the exact DFM [EDFM] (see Lippi (2019) and the references therein), but there are also some relevant differences. First, in the EDFM series Y_t load the factors even contemporaneously and not only with lags. Second, the factors and the idiosyncratic terms in the EDFM are uncorrelated at any lag-lead, whereas in the MAI we have* $\mathrm{E}(f_t\varepsilon_{t+j}) = 0$ *only for* $j > 0$*. Third, the contemporaneous variance matrix of the idiosyncratic terms in the EDFM is diagonal, whereas* Σ *is generally not.*

Putting emphasis on the analogies between MAI and EDFM, Carriero et al. (2016) propose identifying structural shocks as linear transformations of the index shocks only. Starting from the Wold representation of series Y_t

$$Y_t = \Psi(L)\varepsilon_t,$$

and inserting between $\Psi(L)$ and ε_t the decomposition of the identity matrix as in Centoni and Cubadda (2003)

$$I_n = \Sigma\omega(\omega'\Sigma\omega)^{-1}\omega' + \omega_\perp(\omega_\perp'\Sigma^{-1}\omega_\perp)^{-1}\omega_\perp'\Sigma^{-1}, \tag{4}$$

one gets the following decomposition of series Y_t :

$$Y_t = \chi_t + \iota_t \tag{5}$$

where

$$\chi_t = \Psi(L)\Sigma\omega\underline{\Sigma}^{-1}\varepsilon_t^\chi, \tag{6}$$

$$\iota_t = \Psi(L)\omega_\perp(\omega_\perp'\Sigma^{-1}\omega_\perp)^{-1}\varepsilon_t^\iota, \tag{7}$$

$\underline{\Sigma} = \omega'\Sigma\omega$, $\varepsilon_t^\chi = \omega'\varepsilon_t$, $\varepsilon_t^\iota = \omega_\perp'\Sigma^{-1}\varepsilon_t$, $\mathrm{E}(\varepsilon_t^\chi\varepsilon_t^{\iota\prime}) = 0$, and $\mathrm{E}(\chi_t\iota_{t-j}') = 0$ for $\forall j$.

Since the shocks ε_t^χ are those of the indexes, ε_t^χ may be interpreted as the common shocks and χ_t as the common components of the series Y_t. Similarly, ε_t^ι and ι_t can be labeled, respectively, as uncommon shocks and a uncommon component.

Interestingly, post-multiplying with $\omega_\perp$ both sides of the relation

$$\Psi(L)(I_n - \sum_{j=1}^{p-1}\alpha_j\omega' L^j) = I_n$$

we get $\Psi(L)\omega_\perp = \omega_\perp$, which in turn implies that the Wold polynomial matrix of the MAI has the form

$$\Psi(L) = I_n + \sum_{j=1}^{\infty}\theta_j\omega' L^j \tag{8}$$

where θ_j is an $n \times q-$matrix for $j > 0$.

Having substituted $\Psi(L)$ in Equations (6) and (7) with the RHS of Equation (8), we can finally prove the following proposition:

Proposition 1 *In the MAI, the components of* Y_t *in (5) read*

$$\chi_t = (\Sigma\omega\underline{\Sigma}^{-1} + \sum_{j=1}^{\infty}\theta_j L^j)\varepsilon_t^\chi,$$

$$\iota_t = \omega_\perp(\omega_\perp'\Sigma^{-1}\omega_\perp)^{-1}\varepsilon_t^\iota,$$

where the uncommon component ι_t *is a* $n-$*dimensional white noise such that* $\mathrm{Rank}\,(\mathrm{E}(\iota_t\iota_t')) = n - q$.

Corollary 1 *The indexes and the common component are linked through the relation* $f_t = \omega' \chi_t$*, which trivially follows from Proposition 1.*

Remark 2 *In view of Proposition 1, the decomposition (5) has clear analogies with the analogous decomposition in the EDFM. However, differently from the idiosyncratic terms in the EDFM, the uncommon component* ι_t *is obviously cross-sectionally dependent.*[3]

Carriero et al. (2016) suggest to recover the structural shocks as linear transformations of the common shocks ε_t^{χ} only. Hence, at most $q < n$ structural shocks can be recovered, as occurs in DFMs and in dynamic stochastic general equilibrium models. In principle, all the identification strategies that are available for structural VARs or structural DFMs (see Stock and Watson (2016) and the references therein) can be adopted.

On the estimation side, Carriero et al. (2016) prove that the iterative maximum likelihood procedure proposed by Reinsel (1983) is consistent when $n = o(\sqrt{T})$. Moreover, they provide an MCMC algorithm for Bayesian estimation and show by simulations that the Bayesian approach outperforms the classical one when $n = 15, 20$. Finally, they document the practical value of the structural MAI by two empirical applications, on the transmission mechanism of monetary policy and on the propagation of demand and supply shocks.

2.2 The Vector Heterogeneous Autoregressive Index Model

The univariate Heterogeneous AR model [HAR], originally proposed by Corsi (2009), is a popular tool to analyze and forecast daily realized volatility [RV] measures without resorting to more involved long-memory models. Technically speaking, the HAR is a constrained AR(22) model where the predictors are the first lags of: (i) the daily RV; (ii) the weekly (5 days) average of the daily RV; (iii) the monthly (22 days) average of the daily RV.

Cubadda et al. (2017) propose a multivariate HAR for a set of n daily RV measures $Y_t^{(d)} \equiv \left(Y_{1,t}^{(d)}, \ldots, Y_{n,t}^{(d)}\right)'$ that is endowed with an index structure. In particular, the Vector Heterogeneous Autoregressive Index model [VHARI] reads as follows

$$Y_t^{(d)} = \alpha^{(d)} \omega' Y_{t-1d}^{(d)} + \alpha^{(w)} \omega' Y_{t-1d}^{(w)} + \alpha^{(m)} \omega' Y_{t-1d}^{(m)} + \varepsilon_t,$$

where (d), (w), and (m) denote, respectively, time horizons of one day, one week, and one month such that

$$Y_t^{(w)} = \frac{1}{5} \sum_{j=0}^{4} Y_{t-jd}^{(d)}, \quad Y_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} Y_{t-jd}^{(d)}$$

The VHARI enjoys two important properties that are not shared by alternative approaches to inducing dimension reduction in the Vector HAR[4]: First, the indexes $f_t^{(d)} = \omega' Y_{t-1d}^{(d)}$ preserve the temporal cascade structure of the HAR model since

$$f_t^{(w)} = \frac{1}{5} \sum_{j=0}^{4} f_{t-jd}^{(d)}, \quad f_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} f_{t-jd}^{(d)},$$

[3] Remarkably, when the factors in the EDFM are estimated by some principal components of series Y_t, the sample variance matrix of the estimated idiosyncratic component has reduced-rank as well.

[4] The most obvious alternatives to the VHARI likely are multivariate principal component regression and reduced-rank regression.

Second, pre-multiplying both sides of the VHARI by ω' yields the following:

$$f_t^{(d)} = \omega'\alpha + \omega'\alpha^{(d)} f_{t-1d}^{(d)} + \omega'\alpha^{(w)} f_{t-1d}^{(w)} + \omega'\alpha^{(m)} f_{t-1d}^{(m)} + \omega'\varepsilon_t,$$

which shows that the indexes follow a multivariate HAR model. In particular, when $q = 1$ a univariate HAR model generated all the dynamics of the n RVs.

On the estimation side, Cubadda et al. (2017) suggest using a switching algorithm [SA], an iterative method for numerical maximization of the log-likelihood of complex models that has a long tradition in time-series analysis (see Boswijk and Doornik (2004) and the references therein). In particular, the proposed SA requires the following steps:

1. Given an (initial) estimate of ω, maximize the conditional Gaussian likelihood $\ell(A, \Sigma|\omega)$ where $A = [\alpha^{(d)'}, \alpha^{(w)'}, \alpha^{(m)'}]'$.
2. Given the previously obtained estimates of A and Σ, maximize the conditional likelihood $\ell(\omega|A, \Sigma)$.
3. Repeat Steps 1 and 2 until numerical convergence occurs.[5]

A key point of the above SA is that both Steps 1 and 2 require running OLS regressions only. This feature provides the SA with several advantages over Newton-type optimization methods, such as computational simplicity, no need for normalization conditions in ω, explicit optimization at each step, and ease of application of regularization schemes or linear restrictions on parameters (see Cubadda and Guardabascio (2019) for additional discussion). Furthermore, when the SA is initialized with consistent estimates and is iterated sufficiently often, the resulting estimator is asymptotically equivalent to the ML one (Hautsch et al. (2023)). Cubadda et al. (2017) show by simulation that the suggested SA performs well even when elements of ε_t have a log-normal error distribution with GARCH variances.

Following Patton and Sheppard (2009), Cubadda et al. (2017) use the VHARI to build the optimal linear combination of ten different estimators of the volatility of the same market and to evaluate its merits through an out-of-sample forecasting exercise. The VHARI model proves to work well, often outperforming previously existing methods.

2.3 The Index-Augmented Auto-Regressive Model

A possible limitation of MAI as a forecasting tool is that the only predictors of the series $y_{i,t}$, for $i = 1, \ldots, n$, are the lagged indexes, whereas the forecasts obtained through the DFM exploit information coming from the past of both factors and the series $y_{i,t}$ itself (see the seminal contributions by Stock and Watson (2002a) and Stock and Watson (2002b)). Although the indexes may be interpreted as 'supervised' factors that are constructed for emphasizing the comovements between the present and the past of the system, it may occur that some variables are better predicted by their own lags rather than by any linear combination of all variables only.

In order to overcome such limitation, Cubadda and Guardabascio (2019) extended the basic MAI model by allowing individual AR structures for each element of Y_t. Their key assumption is the following.

Assumption 2 *It holds*

$$\phi_{ik}^{(j)} = \sum_{m=1}^{q} \alpha_{im}^{(j)} \omega_{km},$$

[5] A general proof of the convergence of this family of iterative procedures is given by Oberhofer and Kmenta (1974).

where $\phi_{ik}^{(j)}$ is the generic element of the polynomial matrix Φ_j, ω_{km} is the generic element of ω, and $\alpha_{im}^{(j)}$ is the generic element of α_j for $j = 1, \ldots, p$, $i = 1, \ldots, n$, $k = 1, \ldots, i-1, i+1, \ldots, n$.

In words, Assumption 2 states that there is a reduced number of channels p through which each variable is influenced by the past of other variables in the system, which is consistent with the common view that few shocks are responsible for most macroeconomic fluctuations.

Under Assumption 2 and using the reparametrization $\delta_{ii}^{(j)} = \phi_{ii}^{(j)} - \sum_{m=1}^{q} \alpha_{im}^{(j)} \omega_{im}$, Model (1) can be rewritten into the following Index-Augmented Auto-Regressive model [IAAR]:

$$Y_t = \sum_{j=1}^{p} D_j Y_{t-j} + \sum_{j=1}^{s} \alpha_j f_{t-j} + \varepsilon_t, \tag{9}$$

where D_j is a $n \times n$ diagonal matrix with $\delta_{ii}^{(j)}$ as generic diagonal element, and, for greater generality, $s \leq p$.

Remark 3 *Since the number of parameters of Model (9) is equal to $n(qs + q + p) - q^2$, it is necessary to impose proper upper bounds to either q or s to ensure that the MIAAR is more parsimonious than the VAR. To this end, it is easy to see that sufficient conditions are $q < n - 1$ for $s = p \geq 2$ or $s < p - 1$ for any p and $q < n$. However, in empirical applications, the estimated values of q are typically much smaller than n (see Cubadda and Guardabascio (2019) and Carriero et al. (2022)).*

Remark 4 *The individual forecasting equation of the IAAR reads*

$$y_{it+1} = \sum_{j=0}^{p-1} \delta_{ii}^{(j)} y_{it-j} + \sum_{j=0}^{s-1} \alpha_{i\cdot}^{(j)\prime} f_{t-j} + \varepsilon_{it+1}, \tag{10}$$

where $\alpha_{i\cdot}^{(j)\prime}$ is the $i-$th row of matrix α_j. Equation (10) is entirely analogous to the individual forecasting equation of the DFM, with one important difference. Whereas factors are typically estimated using principal component methods, which aim to maximize the contemporaneous variability of series Y_t, the indexes in (10) are constructed explicitly taking into account the covariability between each series y_{it} and the lags of other elements of Y_t conditionally on the lags of the series y_{it}.

Remark 5 *Interestingly, by the same argument underlying Proposition 1, we see that, differently from the MAI, $\Psi(L)\omega_\perp \neq \omega_\perp$, which implies, in view of Equation (7), that the uncommon component ι_t is generally autocorrelated in the case of the IAAR. Hence, the decomposition (5) for the IAAR closely resembles the analogous decomposition in the approximate DFM (see Lippi (2019) and the references therein). However, estimation of the indexes f_t does not require that $n \to \infty$ and any condition on the autocorrelations and cross-correlations of elements of ι_t or on the loadings α_j as in the approximate DFM.*

Cubadda and Guardabascio (2019) proposed a two-step SA for the estimation of the IAAR, along with a variant where a ℓ_2 regularization scheme is applied in both steps. They show by simulations that the regularized version of the SA outperforms the standard one with $n = 20$. Regarding model specification, they opt for the use of Information Criteria [IC], in line with previous contributions showing that IC outperform likelihood ratio tests in selection of reduced-rank VAR models (see, e.g., Gonzalo and Pitarakis (1999), Cavaliere et al. (2015), Cavaliere et al. (2018)). Finally, the IAAR proves to outperform well-known macroeconomic forecasting methods when applied to systems with n ranging from 4 to 40.

Carriero et al. (2022) endowed the IAAR with Stochastic Volatility [IAAR-SV] in the errors ε_t and offered Bayesian estimation using Markov Chain Monte Carlo [MCMC] techniques. Furthermore, they use (4) to decompose the time-varying volatility $\mathrm{E}(\varepsilon_t\varepsilon_t') = \Sigma_t$ as follows:

$$\Sigma_t = \underbrace{\Sigma_t\omega(\omega'\Sigma_t\omega)^{-1}\omega'\Sigma_t}_{\text{common}} + \underbrace{\omega_\perp(\omega_\perp'\Sigma_t^{-1}\omega_\perp)^{-1}\omega_\perp'}_{\text{uncommon}}$$

Carriero et al. (2022) apply the IAAR-SV to analyze the commonality in both levels and volatilities of inflation rates in several countries, and their main finding is that a substantial fraction of inflation volatility can be attributed to a global factor that also drives inflation levels and their persistence.

2.4 The Time-Varying Multivariate Autoregressive Index Model

A further step towards taking into account parameter instabilities over time was taken by Cubadda et al. (2025), who proposed the following MAI with Time Varying Parameters and Time-Varying Volatility [MAI-TVP-TVV]:

$$Y_t = \sum_{j=1}^{p} \alpha_{j,t}\omega' Y_{t-i} + \varepsilon_t,$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_t + \kappa_t$$

where $\boldsymbol{\alpha}_t = \mathrm{Vec}(\alpha_{1,t}', \dots, \alpha_{n,t}')'$, $\varepsilon_t \sim N(0, \Sigma_t)$, $\kappa_t \sim N(0, Q_t)$, ε_t and κ_t are independent at any lag and lead. Notice that it is assumed that the index loadings evolve over time as random walks, while the index weights ω remain stable.

In order to overcome the computational limitation related to MCMC procedures, Cubadda et al. (2025) offers a hybrid estimation method that combines the SA, Kalman filter with forgetting factors (Koop and Korobilis (2014)), and exponentially weighted moving average techniques (Johansson et al. (2023)) for the time-varying volatility.

An empirical application, where 25 US quarterly time series are used to forecast three key macroeconomic variables, shows that the MAI-TVP-TVV is one of the best models in a large set of competitors for all targets, improving upon its counterparts especially at short horizons. Other interesting findings are that, once the MAI is endowed with time-varying volatility [MAI-TVV], there are no clear improvements in adding time-varying parameters for point forecasting, but the MAI-TVP-TVV always outperforms the MAI-TVV in density forecasting.

2.5 The Dimension-Reducible VAR

Cubadda and Hecq (2022a) studied the conditions under which the dynamics in a large-dimensional VAR are entirely generated by a small-scale VAR. They show that such conditions are met when the coefficient matrices of the large VAR have the same common right space and a common left null space. This entails combing Assumption 1 with the following:

Assumption 3 *It holds*

$$\omega_\perp'[\Phi_1, \dots, \Phi_p] = 0$$

Assumption 3 is popularly known in time series econometrics as the CSC (see Cubadda and Hecq (2022b) and the reference therein) given that

$$\omega_\perp' Y_t = \omega_\perp' \varepsilon_t,$$

that is, there exist $(n-q)$ linear combinations of variables Y_t that are white noise and, as such, cannot exhibit cyclical behavior.

Taking Assumptions 1 and 3 together leads to the Dimension-Reducible VAR model [DRVAR]:

$$Y_t = \sum_{j=1}^{p} \omega\phi_j f_{t-j} + \varepsilon_t, \tag{11}$$

where ϕ_j is a $q \times q$ matrix for $j = 1, ..., p$.

Assuming, without loss of generality, that $\omega'\omega = I_q$ and $\omega'_\perp \omega_\perp = I_{n-q}$, we can decompose series Y_t as follows

$$Y_t = \omega f_t + \omega_\perp \eta_t, \tag{12}$$

where f_t is the dynamic component and $\eta_t = \omega'_\perp \varepsilon_t$ is the static one. Premultiplying both sides of DRVAR by ω' one gets

$$f_t = \sum_{j=1}^{p} \phi_j f_{t-j} + \varepsilon_t^\chi,$$

where $\varepsilon_t^\chi = \omega'\varepsilon_t$, which shows that f_t is generated by a $q-$dimensional VAR(p) process.

By inserting the Wold representation of the dynamic components f_t in Equation (12) it follows:

$$Y_t = \omega\gamma(L)\varepsilon_t^\chi + \omega_\perp \eta_t, \tag{13}$$

where $\gamma(L)^{-1} = I_n - \sum_{j=1}^{p} \phi_j L^j$. Finally, by linearly projecting $\omega_\perp \eta_t$ on ε_t^χ, we obtain $\omega_\perp \eta_t = \rho\varepsilon_t^\chi + \nu_t$ with $\mathrm{E}(\varepsilon_t^\chi v'_t) = 0$, which can be inserted into Equation (13) to get

$$Y_t = C(L)\varepsilon_t^\chi + \nu_t, \tag{14}$$

where $C_0 = \omega + \rho$ and $C_j = \omega\gamma_j$ for $j > 0$.

Representation (14) highlights that system dynamics are completely generated by the common reduced form errors ε_t^χ. Consequently, Cubadda and Hecq (2022a) label ν_t as the ignorable errors, as they are noise without structural interpretation. Since the errors ε_t^χ and ν_t are uncorrelated at any lead and lag, it is then possible to recover the structural shocks solely from the reduced form errors ε_t^χ of the common component χ_t using any of the procedures that are commonly employed in structural VARs or structural DFMs (see Stock and Watson (2016) and the references therein).

In order to estimate the matrix ω, one may rely on a nonparametric estimator proposed by Lam et al. (2011). The underlying intuition is that the matrix ω lies in the space generated by the eigenvectors associated with the q nonzero eigenvalues of the symmetric and semipositive definite matrix.

$$M = \sum_{j=1}^{p_0} \Sigma_y(j)\Sigma_y(j)',$$

where $p_0 \geq p$ and $\Sigma_y(j)$ is the autocovariance matrix of series Y_t in lag j. Under some regularity conditions, the matrix formed by the eigenvectors associated with the q largest eigenvalues of the sample estimate of M is a $\sqrt{T}$-consistent estimator of ω (up to an orthonormal transformation) when q is fixed, $n, T \to \infty$, and $\omega'_i\omega_i = O(n)$ for $i = 1, ..., q$, where $\omega = [\omega_1, ..., \omega_q]$. Remarkably, the speed of convergence of the estimator, namely $\sqrt{T}$, is the same as when the dimension n is finite.

Moreover, Cubadda and Hecq (2022a) provide both the OLS and the GLS estimators of the coefficients ϕ' s in equation (11) and consistent information criteria for the selection of q, and show by simulations that the proposed methodology works well with the temporal and cross-sectional sizes that are typical in macroeconomics. Finally, the approach is applied to analyze a large set of US economic time series and to identify the shock that is responsible for most of the common volatility in the business cycle frequency band.

2.6 The Vector-Error Correction Index Model

The models considered so far do not explicitly deal with the possible presence of unit roots. Given that most macroeconomic and financial time series are characterized by stochastic trends, it is important to understand how a cointegrated VAR model can be augmented with an index structure.

Let us assume that series Y_t follow the Vector Error-Correction Model [VECM]

$$\Delta Y_t = \alpha_0 \beta' Y_{t-1} + \sum_{j=1}^{p-1} \Pi_j \Delta Y_{t-j} + \varepsilon_t, \tag{15}$$

where α_0 and β are full-rank $n \times r$ ($r < n$) matrices such that $\alpha_0\beta' = -\Phi(1)$, $\Pi_j = -\sum_{i>j} \Phi_i$ for $j = 1, \ldots, p-1$, ${\alpha_0}'_{\perp} \bar{\Pi} \beta_{\perp}$ is non-singular, and $\bar{\Pi} = I_n - \sum_{j=1}^{p-1} \Pi_j$. Under such assumptions, it is well known that the elements of Y_t are individually, at most, $I(1)$ and that they are jointly cointegrated of order 1, in the sense that $\beta' Y_{t-1}$ is $I(0)$ (see Johansen (1995) and the references therein).

To possibly reduce the number of parameters in the VECM, Cubadda and Mazzali (2024) take the following assumptions:

Assumption 4 *For* $\Pi = [\Pi_1', \ldots, \Pi_{p-1}']'$ *it holds*

$$\Pi = A\omega',$$

where ω *is a full-rank* $n \times q$ *matrix with* $q < n$ *and* A *is a full-rank* $n(p-1) \times q$ *matrix.*

Assumption 5 *It holds*

$$\beta = \omega\gamma,$$

where γ *is a full-rank* $q \times r$ *matrix with* $q \geq r$*.*

Under Assumptions 4 and 5, Model (15) can be rewritten in the following Vector-Error Correction Index Model [VECIM]:

$$\Delta Y_t = \alpha_0 \gamma' f_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta f_{t-j} + \varepsilon_t,$$

where γ is a full-rank $q \times r$ matrix ($q \geq r$), and α_j is an $n \times q$ matrix for $j = 1, ..., p-1$ such that $\text{rank}([\alpha_1', ..., \alpha_{p-1}']') = q$. Notice that the cointegration matrix is given by $\beta = \omega\gamma$.

Interestingly, the indexes f_t themselves are generated by a $q-$dimensional VECM:

$$\Delta f_t = \underline{\alpha}_0 \gamma' f_{t-1} + \sum_{j=1}^{p-1} \underline{\alpha}_j \Delta f_{t-j} + \varepsilon_t^{\chi},$$

where $\underline{\alpha}_j = \omega' \alpha_j$, for $j = 0, 1 \ldots, p-1$.

By first inserting the decomposition (5) between $\Psi(L)$ and ε_t into the Wold representation of the first differences ΔY_t:

$$\Delta Y_t = \Psi(L)\varepsilon_t,$$

and then further decomposing the common component χ_t into permanent and transitory subcomponents as in Centoni and Cubadda (2003), we get the following:

$$Y_t = \chi_t + \iota_t = \pi_t + \tau_t + \iota_t, \tag{16}$$

where

$$\Delta\pi_t = \Psi(L)\Sigma\omega\underline{\Sigma}^{-1}\Sigma\underline{\alpha}_{0\perp}(\underline{\alpha}_{0\perp}'\Sigma\underline{\alpha}_{0\perp})^{-1}\underbrace{\underline{\alpha}_{0\perp}'\varepsilon_t^{\chi}}_{\varepsilon_t^{\pi}}, \tag{17}$$

$$\Delta\tau_t = \Psi(L)\Sigma\omega\underline{\Sigma}^{-1}\underline{\alpha}_0(\underline{\alpha}_0'\underline{\Sigma}^{-1}\underline{\alpha}_0)^{-1}\underbrace{\underline{\alpha}_0'\underline{\Sigma}^{-1}\varepsilon_t^{\chi}}_{\varepsilon_t^{\tau}}, \tag{18}$$

$$\Delta\iota_t = \Psi(L)\omega_\perp(\omega_\perp'\Sigma^{-1}\omega_\perp)^{-1}\underbrace{\omega_\perp'\Sigma^{-1}\varepsilon_t}_{\varepsilon_t^{\iota}} \tag{19}$$

Since errors ε_t^{π} are the innovations of the common trends of the indexes f_t (see e.g. Johansen (1995)) and errors ε_t^{τ} are such that $\mathrm{E}(\varepsilon_t^{\pi}\varepsilon_t^{\tau\prime}) = 0$, Cubadda and Mazzali (2024) label π_t as the common permanent component, τ_t as the common transitory component, whereas ι_t is the uncommon component given that $\mathrm{E}(\varepsilon_t^{\iota}\varepsilon_t^{\pi\prime}) = 0$ and $\mathrm{E}(\varepsilon_t^{\iota}\varepsilon_t^{\tau\prime}) = 0$.

Following a similar reasoning as the one leading to Proposition 1, post-multiplying with $\omega_\perp$ both sides of the relation

$$\Psi(L)(\Delta I_n - \sum_{j=1}^{p-1}\alpha_j\omega'\Delta L^j - \alpha_0\gamma'\omega' L) = \Delta I_n$$

we again get $\Psi(L)\omega_\perp = \omega_\perp$, which in turn implies that the Wold polynomial matrix of the VECIM has the same form as (8). Finally, inserting (8) in Equations (17), (18), and (19) we can prove the following proposition:

Proposition 2 *In the VECIM, the first differences of the components of Y_t in (16) read*

$$\Delta\pi_t = (\Sigma\omega\underline{\Sigma}^{-1} + \sum_{j=1}^{\infty}\theta_j L^j)\Sigma\underline{\alpha}_{0\perp}(\underline{\alpha}_{0\perp}'\Sigma\underline{\alpha}_{0\perp})^{-1}\varepsilon_t^{\pi} \equiv P(L)\varepsilon_t^{\pi},$$

$$\Delta\tau_t = (\Sigma\omega\underline{\Sigma}^{-1} + \sum_{j=1}^{\infty}\theta_j L^j)\underline{\alpha}_0(\underline{\alpha}_0'\underline{\Sigma}^{-1}\underline{\alpha}_0)^{-1}\varepsilon_t^{\tau} \equiv T(L)\varepsilon_t^{\tau}$$

$$\Delta\iota_t = \omega_\perp(\omega_\perp'\Sigma^{-1}\omega_\perp)^{-1}\varepsilon_t^{\iota},$$

where the uncommon component ι_t is a $n-$dimensional random walk such that $\mathrm{Rank}\,(\mathrm{E}(\Delta\iota_t\Delta\iota_t')) = n - q$.

Notice that Proposition 2 implies that Corollary 1 applies to the VECIM as well.

Remark 6 *Given that the components in (16) are not correlated with each other at any lag and lead, the VECIM allows one to perform a structural analysis taking advantage of the features of both the DFM, namely isolating shocks that are common among variables, and the VECM, namely disentangling shocks having transitory or permanent effects. For instance, one may identify the structural transitory shocks as $u_t = C^{-1}D\varepsilon_t^{\tau}$ and the impulse response functions as $\Theta(L) = T(L)D^{-1}C$, where D is the matrix formed by the first r rows of $T(0)$ and C is a lower triangular matrix such that*

$$CC' = D\underline{\alpha}_0'\Sigma^{-1}\omega'\Sigma\omega\Sigma^{-1}\underline{\alpha}_0 D'$$

Since the first r rows of $\Theta(0)$, being equal to C, form a lower triangular matrix, the usual interpretation of structural shocks obtained through a Cholesky factorization applies to u_t.

Cubadda and Mazzali (2024) offer a three-step SA for the estimation of the VECIM and propose to select the triple (p, q, r) in a unique search by IC. An extensive Monte Carlo study shows that the proposed methodology works reasonably well for n ranging from 6 to 18 when the model is identified by the Hannan-Quinn IC. Moreover, in an empirical application, they identify a shock that maximizes the variability of the common transitory component of unemployment at the business cycle frequencies, and another one that does the same, but for the common permanent component of unemployment. These two shocks are endowed with a neater economic interpretation than a unique main business cycle shock identified according to Angeletos et al. (2020).

3 A New Proposal: The Cointegrated Index-Augmented Autoregressive Model

A possible limitation of the VECIM is that the uncommon component ι_t is necessarily a random walk, which may be considered restrictive for some applications. For example, Barigozzi et al. (2021) propose a DFM where the idiosyncratic components may be I(0) or I(1).

In order to overcome this issue, one can combine the VECIM with the IAAR. Formally, this involves taking Assumption 5 along with the following one:

Assumption 6 *For the VECM (15) it holds*

$$\pi_{ik}^{(j)} = \sum_{m=1}^{q} \alpha_{im}^{(j)} \omega_{km},$$

where $\pi_{ik}^{(j)}$ is the generic element of the polynomial matrix Π_j for $j = 1, \ldots, p-1$, $i = 1, \ldots, n$, $k = 1, \ldots, i-1, i+1, \ldots, n$.

Taking Assumptions 5 and 6, the model (15) can be rewritten into the following Cointegrated Index-Augmented Auto-Regressive model [CIAAR]

$$\Delta Y_t = \sum_{j=1}^{p-1} D_j \Delta Y_{t-j} + \alpha_0 \gamma' \omega' Y_{t-1} + \sum_{j=1}^{s-1} \alpha_j \omega' \Delta Y_{t-j} + \varepsilon_t, \tag{20}$$

where D_j is a $n \times n$ diagonal matrix with $\delta_{ii}^{(j)} = \pi_{ii}^{(j)} - \sum_{m=1}^{q} \alpha_{im}^{(j)} \omega_{im}$ as generic diagonal element.

When the elements of the series Y_t are I(1), Model (20) includes several earlier models as special cases. In fact, we have a MAI for series ΔY_t if $p, r = 0$, an IAAR for series ΔY_t is obtained if $r = 0$, and a VECIM if $p = 0$.

Remark 7 *Interestingly, by the same argument underlying Proposition 2, we see that, differently from the VECIM, $\Psi(L)\omega_\perp \neq \omega_\perp$, which implies, in view of Equation (19), that the first differences of the uncommon component $\Delta\iota_t$ are generally autocorrelated in the case of the CIAAR. The uncommon component ι_t is still stochastically singular with rank $n - q$. Since system (20) has overall $n - r$ unit roots, while the common component χ_t has $q - r$ unit roots, the uncommon component ι_t has $n - q$ unit roots (see Deistler and Wagner (2017) and Barigozzi et al. (2020) on the properties of singular I(1) stochastic processes).*

Following Cubadda and Guardabascio (2019) and Cubadda and Mazzali (2024), the estimation procedure is based on an SA where each step is designed to increase the Gaussian likelihood of Model (20). In detail, when $0 < r < q$, the procedure goes as follows:

1. Given (initial) estimates of γ, ω, and $D = [D_1, \ldots, D_{p-1}]'$, maximize the conditional Gaussian likelihood $\mathcal{L}(A^\dagger, \Sigma|\gamma, \omega, D)$ by estimating $A^\dagger = [\alpha_0', A']'$, where $A = [\alpha_1', ..., \alpha_{s-1}']'$, and Σ with OLS on the following equation
$$\Delta Y_t - \sum_{j=1}^{p-1} D_j \Delta Y_{t-j} = \alpha_0 \gamma' \omega' Y_{t-1} + \sum_{j=1}^{s-1} \alpha_j \omega' \Delta Y_{t-j} + \varepsilon_t$$

2. Premultiply by $\Sigma^{-1/2}$ and apply the Vec operator to both the sides of Equation (20), then use the property $\text{Vec}(ABC) = (C' \otimes A)\text{Vec}(B)$ to get
$$\Sigma^{-1/2}\Delta Y_t = \sum_{j=1}^{p-1}(Y_{t-j}' \otimes \Sigma^{-1/2})\text{Vec}(D_j) + (Y_{t-1}' \otimes \Sigma^{-1/2}\alpha_0\gamma' + \sum_{j=1}^{s-1}\Delta Y_{t-j}' \otimes \Sigma^{-1/2}\alpha_j)\text{Vec}(\omega') + \Sigma^{-1/2}\varepsilon_t,$$
and reparametrize the above model as
$$\Sigma^{-1/2}\Delta Y_t = \sum_{h=1}^{p-1}[(Y_{t-j}' \otimes \Sigma^{-1/2})M]\delta_j + (Y_{t-1}' \otimes \Sigma^{-1/2}\alpha_0\gamma' + \sum_{j=1}^{s-1}\Delta Y_{t-j}' \otimes \Sigma^{-1/2}\alpha_j)\text{Vec}(\omega') + \Sigma^{-1/2}\varepsilon_t, \tag{21}$$
where δ_j is a $n-$vector such that $D_j = \text{diag}(\delta_j)$, and M is a binary $n^2 \times n-$matrix whose generic element m_{ik} is such that
$$m_{ik} = \begin{cases} 1 & \text{if } i = 1 + (k-1)(n+1), \quad k = 1, ..., N \\ 0 & \text{otherwise} \end{cases}$$
Given the previously obtained estimates of $A^\dagger$, γ, and Σ, maximize $\mathcal{L}(\omega, D|A^\dagger, \gamma, \Sigma)$ by estimating $\text{Vec}(\omega')$ and $\delta = [\delta_1', ..., \delta_{p-1}']'$ with OLS on Equation (21).

3. Given the previously obtained estimates of ω and D, maximize $\mathcal{L}(\gamma|\omega, D)$ by estimating γ as the eigenvectors that correspond to the r largest eigenvalues of the matrix
$$S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$$
where $S_{ij} = \sum_{t=p+1}^T R_{i,t}R_{j,t}'$ for $i, j = 0, 1$, $R_{0,t}$ and $R_{1,t}$ are, respectively, the residuals of an OLS regression of $\Delta Y_t - \sum_{j=1}^{p-1} D_j \Delta Y_{t-j}$ and $\omega' Y_{t-1}$ on $[\Delta Y_{t-1}'\omega, \ldots, \Delta Y_{t-s+1}'\omega]'$.

4. Repeat steps 1 to 3 until numerical convergence occurs.

When $r = 0$, step 3 is clearly not needed, and steps 1 and 2 must be modified as follows:

1.1 Given (initial) estimates of ω and D, maximize $\mathcal{L}(A, \Sigma|\omega, D)$ by estimating A and Σ with OLS on the following model
$$\Delta Y_t - \sum_{j=1}^{p-1} D_j \Delta Y_{t-j} = \sum_{j=1}^{s-1} \alpha_j \omega' \Delta Y_{t-j} + \varepsilon_t$$

2.1 Given the previously obtained estimates of A and Σ, maximize $\mathcal{L}(\omega|A, \Sigma)$ by estimating $\text{Vec}(\omega')$ and δ with OLS on the following model
$$\Sigma^{-1/2}\Delta Y_t = \sum_{h=1}^{p-1}[(Y_{t-j}' \otimes \Sigma^{-1/2})M]\delta_j + (\sum_{j=1}^{s-1}\Delta Y_{t-j}' \otimes \Sigma^{-1/2}\alpha_j)\text{Vec}(\omega') + \Sigma^{-1/2}\varepsilon_t,$$

Finally, when $r = q$, we can assume without loss of generality that $\gamma = I_q$. Then Step 3 is again not needed, whereas Steps 1 and 2 must be modified as follows.

1.3 Given (initial) estimates of ω and D, maximize $\mathcal{L}(A^{\dagger}, \Sigma|\omega, D)$ by estimating $A^{\dagger}$ and Σ with OLS on the following model

$$\Delta Y_t - \sum_{j=1}^{p-1} D_j \Delta Y_{t-j} = \alpha_0 \omega' Y_{t-1} + \sum_{j=1}^{s-1} \alpha_j \omega' \Delta Y_{t-j} + \varepsilon_t$$

2.3 Given the previously obtained estimates of $A^{\dagger}$ and Σ, maximize $\mathcal{L}(\omega, D|A^{\dagger}, \Sigma)$ by estimating $\mathrm{Vec}(\omega')$ and δ with OLS on the following model

$$\Sigma^{-1/2}\Delta Y_t = \sum_{h=1}^{p-1}[(Y'_{t-j} \otimes \Sigma^{-1/2})M]\delta_j + (Y'_{t-1} \otimes \Sigma^{-1/2}\alpha_0 + \sum_{j=1}^{s-1} \Delta Y'_{t-j} \otimes \Sigma^{-1/2}\alpha_j)\mathrm{Vec}(\omega') + \Sigma^{-1/2}\varepsilon_t$$

The choice of initial values for the above procedures is discussed in the Appendix, whereas the selection of the quadruple (p, s, q, r) can be done by IC, sequentially or in a unique search as suggested by Cubadda and Mazzali (2024).

4 Conclusions

The DFM and the VAR are, arguably, among the most popular tools in macroeconometrics and financial econometrics. The two approaches should be considered complementary rather than substitutive, since each of them has its own merits. The MAI represents a link between these two methodologies: On the one hand, it is a VAR with a specific reduced-rank structure that alleviates the dimensionality problem; on the other hand, the MAI and its variants have several analogies with the DFM; in particular, they allow for identifying a small number of common reduced-form errors and for recovering structural shocks from those errors only.

However, the MAI is not affected by some theoretical limitations of the DFM such as the requirement that the cross-sectional dimension diverges to infinity and the need for specific assumptions on the dynamic correlation structure of the idiosyncratic component and on the factor loadings. In a more practical perspective, VARs with an index structure can also handle features such as stochastic volatility (Carriero et al. (2022)) and time-varying parameters (Cubadda et al. (2025)), which are not easily accommodated in DFMs.

Recent developments in VAR models with index structures have considerably extended the original MAI formulation, endowing the model with individual autoregressive structures, stochastic volatility, time-varying parameters, high dimensionality, and cointegration. These extensions have proven to be useful tools for detecting common components, obtaining efficiency gains through the imposition of parameter restrictions, performing structural analysis, and boosting forecast accuracy.

Having reviewed most of the recent advances on the MAI and provided new insights on the representation theory underlying the various formulations, a new model, namely the CIAAR, has been proposed along with an estimation procedure. The hope is that this paper will contribute to providing room for future research on VAR models with index structures.

5 Appendix

The choice of the initial values for a SA is important. Not only is an accurate initialization necessary to boost numerical convergence, but the SA is asymptotically equivalent to the ML one when the parameters to be initialized are consistently estimated (Hautsch et al. (2023)).

With reference to the SA in Section 3, the initial values for γ, ω, and D can be obtained as follows:

1. Use the usual Johansen procedure on the model (15) and obtain estimates $\hat{\alpha}_0$, $\hat{\beta}$, and $\hat{\Pi}_j$ for $j = 1, \ldots, m$, where $m = \max\{p, s\} - 1$.

2. Construct matrices $\tilde{\Pi}_j = \hat{\Pi}_j - \text{diag}[\hat{\pi}_{11}^{(j)}, \ldots, \hat{\pi}_{nn}^{(j)}]'$ for $j = 1, \ldots, m$.

3. Construct the matrix $\tilde{\Phi} = [\tilde{\Pi}_1', \ldots, \tilde{\Pi}_m', \hat{\beta}\hat{\alpha}_0',]'$.

4. Compute the singular value decomposition $\tilde{\Phi} = U\Lambda V'$, where the singular values are not increasingly ordered, and obtain $\hat{\omega}$ as the matrix formed by the first q columns of V.

5. Compute the $q-$rank approximation of $\tilde{\Phi}$ as $\bar{\Phi} = U\bar{\Lambda}V'$, where $\bar{\Lambda}$ is obtained from Λ by setting to 0 the smallest singular values of $n - q$.

6. Construct $\bar{\Pi} = [\bar{\Pi}_1', \ldots, \bar{\Pi}_{p-1}']'$ as the matrix formed by the first $n(p-1)$ rows of $\bar{\Phi}$.

7. Construct $\hat{D}_j$ as a diagonal matrix with the diagonal equal to $\text{diag}[\hat{\pi}_{11}^{(j)} - \bar{\pi}_{11}^{(j)}, \ldots, \hat{\pi}_{nn}^{(j)} - \bar{\pi}_{nn}^{(j)}]'$ for $j = 1, \ldots, s - 1$.

The motivation for the above choices is twofold. First, the asymptotic distribution of the Johansen estimator of β is not affected by restrictions on the short-run parameters (Johansen (1995)), which implies that $\hat{\alpha}_0$, $\hat{\Pi}_j$ and $\tilde{\Pi}_j$ are consistent, although inefficient, estimators of the associated parameters. Second, the right-singular vectors that correspond to the q largest singular values of the matrix $\tilde{\Phi}$ consistently estimate ω (see, e.g., Reinsel et al. (2022)). By the same argument, $\bar{\Pi}$ provides a consistent estimator of Π. Finally, the consistency of $\hat{D} = [\hat{D}_1, \ldots, \hat{D}_{p-1}]'$ trivially follows from the ones of $\hat{\Pi}$ and $\bar{\Pi}$.

References

Angeletos, G. M., Collard, F., and Dellas, H. (2020). Business-cycle anatomy. *American Economic Review*, 110(10):3030–70.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Barigozzi, M., Lippi, M., and Luciani, M. (2020). Cointegration and error correction mechanisms for singular stochastic vectors. *Econometrics*, 8(1):3.

Barigozzi, M., Lippi, M., and Luciani, M. (2021). Large-dimensional dynamic factor models: Estimation of impulse–response functions with I(1) cointegrated factors. *Journal of Econometrics*, 221(2):455–482.

Bernardini, E. and Cubadda, G. (2015). Macroeconomic forecasting and structural analysis through regularized reduced-rank regression. *International Journal of Forecasting*, 31(3):682–691.

Boswijk, H. and Doornik, J. (2004). Identifying, estimating and testing restricted cointegrated systems: An overview. *Statistica Neerlandica*, 58(4):440–465.

Carriero, A., Clark, T. E., and Marcellino, M. (2015). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73.

Carriero, A., Corsello, F., and Marcellino, M. G. (2022). The global component of inflation volatility. *Journal of Applied Econometrics*, 37,(4):700–721.

Carriero, A., Kapetanios, G., and Marcellino, M. (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5):735–761.

Carriero, A., Kapetanios, G., and Marcellino, M. (2016). Structural analysis with multivariate autoregressive index models. *Journal of Econometrics*, 192(2):332–348.

Cavaliere, G., De Angelis, L., Rahbek, A., and Taylor, A. (2015). A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic VAR models. *Oxford Bulletin of Economics and Statistics*, 77(1):106–128.

Cavaliere, G., De Angelis, L., Rahbek, A., and Taylor, A. (2018). Determining the cointegration rank in heteroskedastic VAR models of unknown order. *Econometric Theory*, 34(2):349–382.

Centoni, M. and Cubadda, G. (2003). Measuring the business cycle effects of permanent and transitory shocks in cointegrated time series. *Economics Letters*, 80(1):45–51.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Cubadda, G. (2007). A unifying framework for analysing common cyclical features in cointegrated time series. *Computational Statistics & Data Analysis*, 52(2):896–906.

Cubadda, G. and Guardabascio, B. (2019). Representation, estimation and forecasting of the multivariate index-augmented autoregressive model. *International Journal of Forecasting*, 35(1):67–79.

Cubadda, G., Guardabascio, B., and Grassi, S. (2025). The time-varying multivariate autoregressive index model. *International Journal of Forecasting*, 41(1):175–190.

Cubadda, G., Guardabascio, B., and Hecq, A. (2017). A vector heterogeneous autoregressive index model for realized volatility measures. *International Journal of Forecasting*, 33(2):337–344.

Cubadda, G. and Hecq, A. (2011). Testing for common autocorrelation in data-rich environments. *Journal of Forecasting*, 30(3):325–335.

Cubadda, G. and Hecq, A. (2022a). Dimension reduction for high dimensional vector autoregressive models. *Oxford Bulletin of Economics and Statistics*, 84(5):1123–1152.

Cubadda, G. and Hecq, A. (2022b). Reduced rank regression models in economics and finance. In *Oxford handbook of economic forecasting*. Oxford University Press.

Cubadda, G., Hecq, A., and Palm, F. (2009). Studying co-movements in large multivariate data prior to multivariate modelling. *Journal of Econometrics*, 148(1):25–35.

Cubadda, G. and Mazzali, M. (2024). The vector error correction index model: Representation, estimation and identification. *Econometrics Journal*, 27:126–150.

Deistler, M. and Wagner, M. (2017). Cointegration in singular ARMA models. *Economics Letters*, 155(4):39–42.

Engle, R. and Kozicki, S. (1993). Testing for common features (with comments). *Journal of Business & Economic Statistics*, 11(4):369–395.

Fernández-Villaverde, J., Rubio-Ramírez, J., and Schorfheide, F. (2016). Solution and estimation methods for DSGE models. In Taylor, J. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 527–724. North Holland.

Forni, M. and Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics*, 66(1):124–136.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2009). Opening the black box: structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347.

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory*, 17(6):1113–1141.

Gonzalo, J. and Pitarakis, J. (1999). Dimensionality effect in cointegration analysis. *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive WJ Granger*, pages 212–229.

Hautsch, N., Okhrin, O., and Ristig, A. (2023). Maximum-likelihood estimation using the zig-zag algorithm. *Journal of Financial Econometrics*, 21(4):1346–1375.

Hecq, A., Margaritella, L., and Smeekes, S. (2023). Granger causality testing in high-dimensional VARs: A post-double-selection procedure. *Journal of Financial Econometrics*, 21(3):915–958.

Hsu, N. J., Hung, H. L., and Chang, Y. M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, 52(7):3645–3657.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press.

Johansson, K., Ogut, M. G., Pelger, M., Schmelzer, T., and Boyd, S. (2023). A simple method for predicting covariance matrices of financial returns. *Foundations and Trends in Econometrics*, 12(4):324–407.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, 186(2):325–344.

Koop, G. and Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, 71:101–116.

Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Lippi, M. (2019). Time-domain approach in high-dimensional dynamic factor models. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Presss.

Oberhofer, W. and Kmenta, J. (1974). A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica*, 42(3):579–590.

Patton, A. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218–238.

Reinsel, G. (1983). Some results on multivariate autoregressive index models. *Biometrika*, 70(1):145–156.

Reinsel, G., Velu, R., and Chen, K. (2022). *Multivariate Reduced-Rank Regression. Theory, Methods and Applications*. Springer Nature.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Stock, J. and Watson, M. (2016). Dynamic factor models, factor-augmented vector autoregressions and structural vector autoregressions in macroeconomics. In Taylor, J. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2A, pages 415–525. North Holland.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Vahid, F. and Engle, R. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 8(4):341–360.

RECENT PUBLICATIONS BY *CEIS Tor Vergata*

Beyond Colors: Communication and Social Identity in Natural Groups
Giovanni Di Bartolomeo, Stefano Papa and Alessandra Pelloni
CEIS *Research Paper*, 610 September 2025

Technology Spillovers from the Final Frontier: A Long-Run View of U.S. Space Innovation
Luisa Corrado, Stefano Grassi and Aldo Paolillo
CEIS *Research Paper*, 609 August 2025

Identification, Estimation and Inference in High-Frequency Event Study Regressions
Alessandro Casini and Adam McCloskey
CEIS *Research Paper*, 608 July 2025

When Fewer Bids Increase Competition: Buyer Surplus Enhancing Mergers in Single-Award Procurement Auctions
Gian Luigi Albano, Walter Ferrarese and Roberto Pezzuto
CEIS *Research Paper*, 607 July 2025

The State-Dependent Consumption Response to Government Spending in US: a Markov-Switching TANK Model with Sticky Wages
Francesco Morelli
CEIS *Research Paper*, 606 June 2025

Catastrophic Risk Perception and the Shared Burden Effect: An Experimental Study Using Prospect Theory
Riccardo Manghi and Daniela Di Cagno
CEIS *Research Paper*, 605 June 2025

New Keynesian Models, Puzzles and Omitted Effects
Francesco Morelli
CEIS *Research Paper*, 604 June 2025

A Foundation for Universalisation in Games
Enrico Mattia Salonia
CEIS *Research Paper*, 603 June 2025

On the Estimation of Climate Normals and Anomalies
Tommaso Proietti and Alessandro Giovannelli
CEIS *Research Paper*, 602 June 2025

Meritocracy as an End and as a Means
Enrico Mattia Salonia
CEIS *Research Paper*, 601 May 2025

DISTRIBUTION

Our publications are available online at www.ceistorvergata.it

DISCLAIMER

The opinions expressed in these publications are the authors' alone and therefore do not necessarily reflect the opinions of the supporters, staff, or boards of CEIS Tor Vergata.

MEDIA INQUIRIES AND INFORMATION

For media inquiries, please contact Barbara Piazzi at +39 06 72595652/01 or by e-mail at piazzi@ceis.uniroma2.it. Our web site, www.ceistorvergata.it, contains more information about Center's events, publications, and staff.

DEVELOPMENT AND SUPPORT

For information about contributing to CEIS Tor Vergata, please contact at +39 06 72595601 or by e-mail at segr.ceis@economia.uniroma2.it