The pvars R-Package: VAR Modeling for Heterogeneous Panels

Lennart Empting
Universität Duisburg-Essen

Abstract

pvars offers a seamless implementation of vector autoregressive (VAR) methods for heterogeneous panel data. The R-package comprises *panel cointegration rank tests* which can account for *cross-sectional dependence* and for *structural breaks* in the deterministic terms. The implemented *panel SVAR models* can be estimated under these specifications with *pooled cointegrating vectors* and identified by various *panel identification procedures*. In this article, we review these methods and present their modular implementation in R. Two empirical illustrations reproduce examples from the literature step-by-step and guide the **pvars** user into conducting own analyses.

Keywords: panel cointegration rank tests, panel vector autoregressive models, pooled cointegrating vectors, structural identification, independent components, R, **pvars**.

1. Introduction

The ever-increasing availability of macroeconomic data and recent developments in the multivariate time series analysis for panels have popularized panel vector autoregressive methods in applied econometric research. The **pvars** package summarizes a toolkit for the empirical analysis on panels in the narrow sense, where the same time series variables are repetitively observed across several individual entities such as countries. This three-dimensional data structure thus contrasts with panels in a wider sense, where time series are just arranged along the same periods. Although **pvars**' application is not bounded to financial or macroeconometric research, it addresses the data properties typically found in macroeconomic panels, namely (i) *mostly heterogeneous* (ii) *endogenous interactions* between (iii) potentially *non-stationary* variables. Since (iv) the *time dimension is distinctively larger* than the cross-section, the time series are prone to (v) a *complex deterministic term* with structural breaks in their mean and linear trend. The individual entities are usually subject to (vi) *cross-sectional dependencies*.[1]

In order to deal with these data properties, the econometric literature extends individual time series methods by the cross-sectional dimension under selective pooling assumptions. For example, the panel methods implemented by the Stata commands **xtcointtest** (StataCorp 2019), **xtwest** (Persyn and Westerlund 2008), and **xtpmg** (Blackburne and Frank 2007) orig-

[1]See Pedroni (2019) for an intuition and discussion of these panel properties and an overview on recent developments of single-equation and system-based methods. The data characteristics can be found in the applied econometrics of climate (Pretis 2020), energy (Smyth and Narayan 2015), and growth linked to financial development (Christopoulos and Tsionas 2004) or public capital (Empting and Herwartz 2025).

inate from the single-equation framework comprising residual-based cointegration tests and autoregressive distributive lag (ARDL) models in error-correction representation. In contrast, **pvars** relies on vector autoregressive (VAR) models as a system of equations. This approach has the advantage of avoiding restrictive exogeneity assumptions on the variables and modeling their dynamics and interactions explicitly. The model in vector error correction representation (VECM) can further accommodate multiple long-run relations and different types of deterministic regressors within the cointegration. Particularly the cointegration relations are reasonable candidates for a panel-wide pooling, while short-run dynamics are usually assumed to differ across individuals.

The synthesis of VAR models and methods for heterogeneous panels has not been implemented yet, but many packages for R (2020) can already contribute specifications results and individual counter-checks. Hence, they may be subsumed into two pillars for **pvars**: Firstly, the **vars**-ecosystem with **urca** (Pfaff 2008a), **vars** (Pfaff 2008b), and **svars** (Lange, Dalheimer, Herwartz, and Maxand 2021) covers individual VAR methods such as VECM and structural identification. The second pillar are the single-equation panel methods from **plm** (Croissant and Millo 2008) such as dynamic panel regression models and panel unit root tests. Univariate panel time series can be tested for common or idiosyncratic non-stationarity using the more specialized package **PANICr** (Bronder 2016). Like **pvars**, the R-package **panelvar** (Sigmund and Ferstl 2019) and its Stata-equivalent **pvar** (Abrigo and Love 2016) rest on the two pillars. However, they focus on stationary panels that contain more individuals than periods ($N > T$), and the implemented estimators rely on slope coefficients that are homogeneous across all individuals in the PVAR model. Hence, they rather suit microeconomic applications, and their primary task is to deal with the *Nickell bias* (1981).[2]

The objective of **pvars** is to close the gap in the **vars**-ecosystem and provide a seamless implementation of the fruitful panel VAR methodology. For this, three fields of VAR applications are integrated into **pvars**, namely panel estimation, cointegration testing, and structural identification. The implemented methods particularly noteworthy for each application field are (1) panel VECM with *pooled cointegrating vectors* from Breitung's (2005) two-step estimator. (2) The new panel cointegration rank tests by Arsova and Örsal (2017; 2018) respect cross-sectional dependencies stemming from common factors. Arsova and Örsal (2021) consider also structural breaks in the deterministic term. (3) Data-driven identification methods based on *independent component analysis* (ICA) are extended for panels by Calhoun, Adali, Pearlson, and Pekar (2001) and Herwartz and Wang (2024).

The remainder of this article is structured as follows: Section 2 is a review on the econometric methods which are relevant for using this R-package. Section 3 highlights the available functions in **pvars**, their implementation, and their library of auxiliary functions. In Section 4, we demonstrate two empirical applications of **pvars** to exemplary data from the reviewed articles. Finally, Section 5 summarizes this article.

2. Review of the econometric methodology

A main motivation for combining multivariate time series from individual entities by panel methods is to increase the sample size and thereby extend the available information set for a more precise estimation and higher test power. Table 1 provides an overview of the lit-

[2] See also Canova and Ciccarelli (2013) for a survey on panel VAR models.

erature of panel cointegration rank tests. In this unifying framework, the rows of the table classify the *pooling approach* according the different stages of hypothesis testing, at which the the cross-section information is presumed to be homogeneous. The listed pooling approaches allow for increasing heterogeneity between the individuals: (1) Panel-homogeneous cointegrating vectors enable a pooled estimation of these long-run coefficients. (2) The averaged statistic is standardized by moments of the distribution which the individual statistics have in common.[3] Finally, (3) the meta-analytical combination of individual p-values is the most flexible approach, where the commonality aims at a consistent test decision under an increasing number of individuals. If the alternative hypothesis is true, a non-vanishing share of individuals must actually exhibit this property.

Table 1: Panel tests for the rank of cointegration.

		First generation	**Second generation**		**Third generation**
	Cross-sectional dependence:	Independence	Correlated errors	PANIC (2004)	Correlated probits
Pooling approach:					
(1)	Pooled coefficients	Breitung (2005)	a) $LR(\Pi_C \mid \Pi_A)$	—	×
(2)	Averaged test statistics	Larsson et al. (2001) Örsal, Droge (2014)	a) $LR(\Pi_B \mid \Pi_A)$	Arsova, Örsal (2018) b) PMSB^{Z}	×
(3)	Meta-analytically combined p-values	Maddala, Wu (1999) Choi (2001)	—	Örsal, Arsova (2017) b) PMSB^{F}, PMSB^{C}	Hartung (1999) Arsova, Örsal (2021)

Not implemented in pvars: The panel tests by a) Groen, Kleibergen (2003) and b) Carrion-i Silvestre, Surdeanu (2011).

The columns of Table 1 represent the *generations*, through which panel tests have evolved, and reflect the increasing complexity of the assumed data generating process. The first articles propose the plain approach for combining independent individual tests and thus are naive towards any dependencies between the individual entities. This cross-sectional dependence can emerge e.g. between countries due to their trade and financial relations and is a research topic itself with regard to the global business cycle, the crude oil price, or other important commodities of the world market. The prevalent data property decreases the actual information content in macroeconomic panels compared to cross-independent samples of same size. First-generation tests then reject a correct null hypothesis more often than the nominal significance level and are thus "over-sized". In contrast, panel tests of the second generation maintain a correct test size and third-generation tests are robust against structural breaks additionally. For cointegration rank tests, the econometric literature considers structural breaks in the mean or linear trend of the cointegration relations. Empirical examples which are associated with shifts and with breaks in the linear time trend are the German reunification, the beginning of the Great Moderation, and its discussed end in the wake of the Great Recession. For a proper data representation, the VAR model must respect those changes in the long-run equilibrium, towards the error correction mechanism adjusts. Otherwise, the cointegration tests would miss to detect a reversion to the new equilibrium and thus understate the cointegration rank.

A third dimension of panel test construction is constituted by the underlying individual procedures, which have not been mapped onto Table 1. These *branches* are heterogeneous and

[3] In fact, the univariate counterpart by Im, Pesaran, and Shin (2003, p. 59, Remark 3.1) allows for individual-specific test distributions if their third moments exist. The approximating gamma distribution with its well-defined moments suggests that this holds true for cointegration rank tests. By the Lyapunov central limit theorem, the cross-sectional average of their first and second moments can then be used instead. In light of this, future studies may extend JMN (2000) and KN (2019) to panel tests with individual deterministic terms.

often used side-by-side in a panel test article such that they do not follow such a clear pattern as the other two panel test properties do. This does not only hold for the example of cointegration tests, but also for VAR estimation and structural identification. All panel applications are confronted with similar construction problems and thus follow these dimensions in principle. Since the individual methods elude any classification scheme for Table 1, Section 2.1 explains their technical details firstly and lays the foundation for Section 2.2 presenting the panel methods along the two dimensions of Table 1. The sections have the same structure according to estimation, testing, and identification as established in the individual and extended to the panel context respectively.

Notations. We adhere to the following rules. Subscript $t = 1, \dots, T$ denotes the periods and $k = 1, \dots, K$ refers to the endogenous variables in a multivariate time series. In order to simplify notation, Section 2.1 on the individual methods omits subscript $i = 1, \dots, N$ as the label specific to each of the N individuals. In contrast, Section 2.2 uses subscript i to distinguish individual-specific elements also from homogeneous elements. Without any individual identifier i, the latter must stay the same over the complete cross-section. Moreover, symbol $\xrightarrow{d}$ designates convergence in distribution. The $(K \times (K-r))$ matrix $A_\perp$ is the orthogonal complement of a $(K \times r)$ matrix A such that $\mathrm{rk}\{[A : A_\perp]\} = K$ indicates full rank and $A'_\perp A = 0$. By convention, the orthogonal complement of a nonsingular square matrix is the zero matrix 0 and the orthogonal complement of zero matrix 0 is an identity matrix I_K. With slight abuse of notation, integer $r_\perp$ refers to the number of stochastic trends in a multivariate time series, corresponding to the cointegration rank r.

2.1. Individual methods

The basic unit for all presented panel methods is an individual VAR model of the form

$$\boldsymbol{y}_t = \Phi \boldsymbol{d}_t + A_1 \boldsymbol{y}_{t-1} + ... + A_p \boldsymbol{y}_{t-p} + \boldsymbol{u}_t \quad \text{with} \quad \boldsymbol{u}_t \sim (0, \Sigma_u), \tag{1}$$

where $\boldsymbol{y}_t$ is a vector of K stacked time series, $A_j, j = 1, \dots, p$, are $K \times K$ coefficient matrices for the VAR process of order p, and Φ is a coefficient matrix for the deterministic regressors $\boldsymbol{d}_t$. The errors $\boldsymbol{u}_t$ are assumed to be serially, but not necessarily contemporaneously independent. Hence, the covariance matrix Σ_u is usually non-diagonal as the VAR model is estimated in reduced-form initially. Finding the structural shocks $\boldsymbol{\epsilon}_t = \mathsf{B}^{-1}\boldsymbol{u}_t$, which exhibit no contemporaneous correlation, is the matter of structural identification in the the decomposition problem $\Sigma_u = \mathsf{B}\mathsf{B}'$. The K variances of the structural shocks can be normalized to unity such that $\boldsymbol{\epsilon}_t \sim (0, I_K)$. Consequently, the unique identification of the impact matrix B requires at least $K(K-1)/2$ restrictions on the $K^2 - K$ covariances in Σ_u.

If $\boldsymbol{y}_t$ is integrated of order $I(1)$ at most and its first-differences $\Delta\boldsymbol{y}_t$ are thus stationary, the VAR Model (1) in levels can be rewritten in its vector error correction representation

$$\begin{aligned} \Delta\boldsymbol{y}_t &= \Phi\boldsymbol{d}_t + \Pi_K \boldsymbol{y}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta\boldsymbol{y}_{t-j} + \boldsymbol{u}_t \quad \text{with} \quad \boldsymbol{u}_t \sim (0, \Sigma_u) \\ &\text{using} \quad \Pi_K := -I_K + \sum_{j=1}^{p} A_j \quad \text{and} \quad \Gamma_j := -\sum_{j^*=j+1}^{p} A_{j^*},\ j = 1, \dots, p-1. \end{aligned} \tag{2}$$

The $K \times K$ coefficient matrix $\Pi_K = \alpha\beta'_K$ of the error correction term summarizes the $K \times r$ cointegrating matrix β_K and loading matrix α, which adjusts the disequilibria in the r long-run cointegration relations $\beta_K' \boldsymbol{y}_{t-1}$. The regressors $\boldsymbol{d}_t$ of the deterministic term $\Phi \boldsymbol{d}_t = [\Phi_1 : \Phi_2] \left(\boldsymbol{d}_{1t}', \boldsymbol{d}_{2t}'\right)'$ are either unrestricted ($\boldsymbol{d}_{2t}$) or restricted ($\boldsymbol{d}_{1t}$) to the cointegration relations such that the coefficients Φ_1 emerge from $\Pi := [\Pi_K : \Phi_1] = \alpha\, [\beta_K' : \beta_0'] = \alpha\beta'$. Finally, each Γ_j is a $K \times K$ matrix of coefficients for short-run effects at lag $j = 1, \ldots, p-1$.

Estimating cointegrated VAR models

Johansen (1991, 1996) provides a comprehensive collection of *Maximum-Likelihood* methods for the individual VECM (2). His ML-estimators and LR-tests are the foundation for any cointegrated VAR method of the **pvars** package. Hence, their definitions and components are throughout used in this article and shall be introduced here briefly. For convenience, firstly rewrite Model (2) in compact notation as

$$Z_0 \; = \alpha\beta' Z_1 + \Gamma Z_2 + U \quad \text{now with} \quad \boldsymbol{u}_t \sim \mathcal{N}(0, \Sigma_u), \tag{3}$$

where the matrix $\Gamma := [\Gamma_1 : \cdots : \Gamma_{p-1} : \Phi_2]$ lines up the coefficients for the short-run dynamics and for the unrestricted deterministic regressors. The observed time series are collected in the $K \times T$ matrix $Y := [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$ for a sample $1, \ldots, T$ and the same holds for the error matrix $U := [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T]$, which contains no presample periods by construction of the estimation with lagging regressors. Accordingly, the regressand matrix is given by $Z_0 := \Delta Y$ and the two regressor matrices with T columns alike by

$$Z_1 := \begin{bmatrix} \boldsymbol{y}_0 & \cdots & \boldsymbol{y}_{T-1} \\ \boldsymbol{d}_{11} & \ldots & \boldsymbol{d}_{1T} \end{bmatrix} \quad \text{and} \quad Z_2 \; := \begin{bmatrix} \Delta\boldsymbol{y}_0 & \ldots & \Delta\boldsymbol{y}_{T-1} \\ & \vdots & \\ \Delta\boldsymbol{y}_{1-p+1} & \ldots & \Delta\boldsymbol{y}_{T-p+1} \\ \boldsymbol{d}_{21} & \ldots & \boldsymbol{d}_{2T} \end{bmatrix}. \tag{4}$$

The variables for the error correction term enter Z_1 in levels. Z_2 stacks the first-differenced and the unrestricted deterministic regressors. If $p = 1$, all lagged $\Delta\boldsymbol{y}_t$ drop out of Z_2. In conditional VECM, vectors of L (weakly) exogenous variables $\boldsymbol{x}_t$ could be included as lagged $I(1)$ regressors into Z_1 as well as instantaneous and up to q lagged short-run effects into Z_2. The estimation would follow the same proceeding (Pesaran, Shin, and Smith 2000; Lütkepohl 2005, p. 398), but we skip these terms here for the sake of notational brevity.

ML-estimation.[4] The first order conditions from the maximized log-likelihood function

$$\begin{aligned} \ln \mathcal{L}\left(\alpha, \beta, \Gamma, \Sigma_u\right) = & -\frac{KT}{2} \ln\left(2\pi\right) - \frac{T}{2} \ln\left(\det(\Sigma_u)\right) \\ & -\frac{1}{2} \mathrm{tr}\left[\left(Z_0 - \alpha\beta' Z_1 - \Gamma Z_2\right)' \; \Sigma_u^{-1} \; \left(Z_0 - \alpha\beta' Z_1 - \Gamma Z_2\right)\right] \end{aligned} \tag{5}$$

are solved for the estimators of VECM (3), which corresponds to the following three steps:

1. *Concentrate out* short-run effects and the unrestricted deterministic term. The first step of Π's estimation is to remove the component $\Gamma \boldsymbol{z}_{2t}$ from $\Delta\boldsymbol{y}_t$ and $\boldsymbol{y}_t$ by OLS-regression of Z_0 resp. Z_1 on Z_2. Using the Frisch-Waugh Theorem, the projection

[4] See Johansen (1996, Ch. 6), Lütkepohl (2005, Ch. 7.2.3) and Lütkepohl (2006, Ch. 3.1) for a more detailed derivation and explanation.

matrix $M_2 := I_T - Z_2'(Z_2 Z_2')^{-1} Z_2$ generates the "long-run residuals" $R_0 = Z_0 M_2$ resp. $R_1 = Z_1 M_2$ such that Model (3) collapses into the plain error correction model

$$\begin{aligned} Z_0 M_2 &= \alpha\beta' Z_1 M_2 + \Gamma Z_2 M_2 + U M_2, \\ R_0 &= \alpha\beta' R_1 + \widetilde{U}. \end{aligned} \tag{6}$$

2. Estimate the concentrated Model (6) by *reduced-rank regression* (RRR) according to Anderson (1951).[5] For doing so, define the moment matrices

$$S_{00} := \frac{R_0 R_0'}{T}, \quad S_{01} := \frac{R_0 R_1'}{T}, \quad \text{and} \quad S_{11} := \frac{R_1 R_1'}{T}. \tag{7}$$

Further inserting the OLS estimator (9) for $\widetilde{\alpha}(\beta)$ into the concentrated likelihood function transforms the ML maximization into the generalized eigenvalue problem

$$\det\left(\lambda S_{11} - S_{01}' S_{00}^{-1} S_{01}\right) = 0. \tag{8}$$

This equation has K solutions with the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K \geq 0$, which are the squared *canonical correlation coefficients* and indicate the "strength" of cointegration in each solution. By the definition of cointegration, the correct coefficients $\alpha\beta'$ must recombine the non-stationary R_1 to the stationary R_0. Hence, the estimator $\widetilde{\beta}$ equals the eigenvectors of the r largest λ_j and is super consistent converging with rate T instead of the usual $\sqrt{T}$ of the remaining OLS estimates. The $r_\perp$ eigenvectors with the smallest λ_j are associated with the underlying stochastic trends and thus must cancel out from the henceforth *rank-restricted* VECM.

3. Maximize the log-likelihood function (5) conditional on a given β, for which the super-consistent RRR-estimate $\widetilde{\beta}$ can be inserted. The solution for the remaining estimators of the reduced-form VECM (3) are in fact *conditional OLS*, namely

$$\begin{aligned} \widetilde{\Pi} &= \widetilde{\alpha}\beta' = \left[R_0 \left(\beta' R_1\right)' \left(\left(\beta' R_1\right)\left(\beta' R_1\right)'\right)^{-1}\right]\beta' = S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta', \\ \widetilde{\Gamma} &= \left(Z_0 - \widetilde{\Pi} Z_1\right) Z_2'(Z_2 Z_2')^{-1}, \\ \widetilde{U} &= Z_0 - \widetilde{\Pi} Z_1 - \widetilde{\Gamma} Z_2 = R_0 - \widetilde{\Pi} R_1. \end{aligned} \tag{9}$$

The residual covariance matrix is given by $\widetilde{\Sigma}_u = \frac{\widetilde{U}\widetilde{U}'}{T} = S_{00} - \widetilde{\Pi} S_{01}'$ for ML estimation and is further corrected for n, the number of coefficients per equation, in the OLS estimator $\widehat{\Sigma}_u = \frac{T}{T-n} \cdot \widetilde{\Sigma}_u$. From Eq. (9), it becomes clear that only the loadings $\widetilde{\alpha}(\beta)$ adjust towards the normalization of β, while all other estimators are conditional on the complete matrix product $\widetilde{\Pi} = \widetilde{\alpha}\beta'$. This shows that the cointegrated VAR process (2) and its impulse response functions (IRF) are invariant to any eligible normalization under a chosen rank r of Π.

GLS-based trend adjustment. For the unique purpose of determining the cointegration rank, Lütkepohl and Saikkonen (2000) suggest to estimate and remove the deterministic term prior to an LM-test. Saikkonen and Lütkepohl (2000c) apply the LR-statistic (15) to the "detrended" time series, which may further improve the test power. This *SL-procedure*

[5] See also Anderson (2003, Ch. 12.7) and Izenman (1975).

entails a series of publications proposing different specifications of the deterministic term, which can be subsumed by the additive data generating process

$$\begin{aligned} \boldsymbol{y}_t &= M_\mu \boldsymbol{d}_t + \boldsymbol{y}_t^{st}, \\ \boldsymbol{y}_t^{st} &= A_1 \boldsymbol{y}_{t-1}^{st} + \ldots + A_p \boldsymbol{y}_{t-p}^{st} + \boldsymbol{u}_t^{st}. \end{aligned} \tag{10}$$

Therein, the VAR process of the "pure" stochastic component $\boldsymbol{y}_t^{st}$ is usually abbreviated by a lag polynomial $A(L) = I_K - A_1 L - \ldots - A_p L^p$. The $K \times n_\mu$ matrix M_μ collects the coefficients of the deterministic term in the moving-average representation of VAR (1). Specifications of $\boldsymbol{d}_t$ pursuant to the SL literature are listed in Table 2 and illustrated in Section 4.2.

All SL-procedures consider a *GLS estimator* to determine M_μ. It becomes feasible by a preceding ML-estimation of VECM (2), for which Table 2 indicates also the conforming deterministic cases. Each hypothesis $r_{H0} = 0, \ldots, K-1$ in the test sequence then requires its specific rank-restriction on the ML-estimates and an own GLS estimation of M_μ. For this, the ML-estimates of the rank-restricted VECM are converted into $\widetilde{A}_1, \ldots, \widetilde{A}_p$ of the VAR in levels and inserted into the transformed model

$$Q' A(L) \boldsymbol{y}_t = Q' A(L) M_\mu \boldsymbol{d}_t + Q' \boldsymbol{u}_t^{st} \quad \text{for} \quad t = 1, \ldots, T. \tag{11}$$

Therein, $\boldsymbol{d}_t$ and $\boldsymbol{y}_t$ are set to initial zeros in any presample period $t \leq 0$ and the feasible transformation matrix $\widetilde{Q}$ is obtainable by inserting ML-estimates into

$$Q = \left[\Sigma_u^{-1} \alpha \left(\alpha' \Sigma_u^{-1} \alpha \right)^{-1/2} : \alpha_\perp \left(\alpha'_\perp \Sigma_u \alpha_\perp \right)^{-1/2} \right]. \tag{12}$$

The left-multiplication of $Q'A(L)$ to (10) in (11) allows for subtracting the confounding effects of the VAR dynamics and transforms the residual covariance matrix into I_K. After vectorizing Model(11), M_μ can thus be estimated by OLS. Alternatively, Saikkonen and Lütkepohl (2000c, p. 438) show that

$$QQ' \;=\; \Sigma_u^{-1} \alpha \left(\alpha' \Sigma_u^{-1} \alpha \right)^{-1} \alpha' \Sigma_u^{-1} + \alpha_\perp \left(\alpha'_\perp \Sigma_u \alpha_\perp \right)^{-1} \alpha'_\perp \;=\; \Sigma_u^{-1}. \tag{13}$$

Accordingly, the GLS estimator based on the estimated covariance matrix $\widetilde{\Sigma}_u$ can be applied directly to the data unaltered by $\widetilde{Q}$, but still corrected for $\widetilde{A}(L)$. Both proceedings lead to identical estimation results $\widehat{M}_\mu$. Finally, the deterministic term is subtracted and the VECM of the stochastic component $\hat{\boldsymbol{y}}_t^{st} = \boldsymbol{y}_t - \widehat{M}_\mu \boldsymbol{d}_t$ is estimated by ML again – now without any deterministic term. The usual LR-statistic (15) can assess the sole null hypothesis of r_{H0}, but its test distribution (18) differs from the Johansen procedure.

Deterministic term. Johansen and Nielsen (2018) distinguish between the *innovative* and the *additive* formulation of the deterministic term in Model (2) for the Johansen procedure resp. in Model (10) for the SL-procedure. Table 2 adopts these labels and lists the model specifications. Like in a stable VAR process, the deterministic regressors $\boldsymbol{d}_t$ in VECM (2) can contain either no deterministic component, an intercept, or an additional linear trend. Beyond these "standard" types, non-stationarity and cointegration do complicate the estimation and testing in the presence of deterministic terms. For example, if a constant is assigned to $\boldsymbol{d}_{2t}$, the unit roots in the VECM do not only accumulate the innovations $\boldsymbol{u}_t$ into stochastic trends, but also this deterministic constant into a linear trend. On the other hand, the deterministic

Table 2: Verified specifications of the deterministic term.

Deterministic regressors		**Type**		**Literature**
restricted $\boldsymbol{d}_{1t}$	unrestricted $\boldsymbol{d}_{2t}$	innovative[a)]	additive[b)]	
• ***conventional:***				
none	none	*Case 1*	(needless)	—
constant	none	*Case 2*	*SL_mean*	L&S (2000, p. 185)
none	constant	*Case 3*	*SL_ortho*	S&L (2000a) is irrelevant for **pvars**
linear trend	constant	*Case 4*	*SL_trend*	L&S (2000)
none	constant & linear trend	*Case 5*	—	—
• ***period-specific:***				
none	seasonal dummies	+ *Case 2-5*	+ all	Johansen (1996, p. 84) || ...
impulse dummy	none	+ all	+ all	... TSL (2008, p. 348)
shift dummy	impulse dummies	+ *Case 2* [c)]	+ *SL_mean, SL_trend*	JMN (2000, Ch. 3.2) || S&L (2000b)[d)]
trend break	shift & impulse dummies	+ *Case 4*	+ *SL_trend*	JMN (2000, Ch. 3.1) || TSL (2008)

[a)] Case labeling in accordance with Juselius (2007, Ch. 6.3) and Hlouskova and Wagner (2010, p. 193).
[b)] Overview on SL-test specifications from Trenkler (2008, p. 24, Tab. 1; p. 25, Tab. 2).
[c)] For shifts in *Case 4*, Johansen *et al.* (JMN, 2000, Ch. 4) assess r_{H0} under trend breaks and then restrictions on β_0.
[d)] Lütkepohl, Saikkonen, and Trenkler (TSL, 2004) propose estimators of the unknown shift periods $\boldsymbol{\tau}$.

regressors $\boldsymbol{d}_{1t}$ are restricted to the stationary cointegration relation and, thus, a linear trend in $\boldsymbol{d}_{1t}$ does not generate a quadratic trend in the data. In the additive formulation of the SL-procedure, the stochastic component contains the unit roots so that they do not interfere with the deterministic component. Irrespective of the accumulating innovations, the deterministic term thereby retains its intended specification, as it is visible in the data, and can be easily subtracted after FGLS estimation. For this, each row in Table 2 juxtaposes the specifications of the additive model and the corresponding term in the innovative VECM (2) which provides the parameters in Eq. (11) to (13) for feasible GLS estimation.

Additionally to the conventional types, there can be period-specific shifts in the mean and breaks in the linear trend as well as impulse dummies for a single period or a repeated pattern of dummies for deterministic seasonality. Both test procedures retain the conventional asymptotics in Eq. (19) if solitary impulse dummies are added or if the seasonal dummies in $\boldsymbol{d}_{2t}$ are centered around zero and thus accumulate along a constant. In contrast, only the distributions $Z_{r_\perp}$ of the SL-procedure are invariant to the inclusion of shift dummies irrespective of their known or estimated shift period and both procedures must cope with the nuisance introduced by broken trend slopes. Johansen, Mosconi, and Nielsen (JMN, 2000), Kurita and Nielsen (KN, 2019) as well as Trenkler, Saikkonen, and Lütkepohl (TSL, 2008) provide solutions for the Johansen- and SL-procedure respectively. Note that, beyond these specification whose asymptotics are verified by the literature of Table 2 and described in Section 2.1.2, **pvars** accepts all period-specific extensions technically and does not check their validity for the cointegration tests.

Like for the conventional types in Table 2, any regressor $\boldsymbol{d}_{1t}$ in the cointegration term has a first-differenced counterpart in $\boldsymbol{d}_{2t}$. While the additive model separates clearly between deterministic and stochastic components, the innovative VECM (2) mixes the structural breaks into the autoregressive dynamics. Hence, the innovative model implies not only a single, but additional lagged first-differences of the break over the periods $\tau, \tau+1, \ldots, \tau+(p-1)$ after the break occurring in period τ. Their coefficients depend on the other model parameters and would require non-linear estimation. Against this, all authors of the reviewed literature accept a minor loss in degrees of freedom and prefer to estimate these coefficients separately without those restrictions. Like TSL (2008, p. 335), **pvars** sets lagged impulse dummies for

$\tau, \ldots, \tau+p-1$ in $\boldsymbol{d}_{2t}$ even after a trend break because, in combination with the shift in τ, they control for these non-linear dynamics the same way as lagged shift dummies. In return, a subsequent FGLS estimation of M_μ in the additive Model (10) uses the regressor matrix $\boldsymbol{d}_t$ without these lagged first-differences. Besides the conventional term, it contains only the trend break and shift at the very same period τ as illustrated in Eq. (38).

Testing the cointegration rank

In order to determine the cointegration rank r in the K-dimensional System (2), all testing procedures imply a sequential decision making according to the hypotheses

$$H_0 : \mathrm{rk}\,(\Pi) = r_{H0} \quad \text{versus} \quad H_1 : \mathrm{rk}\,(\Pi) > r_{H0}, \quad r_{H0} = 0, \ldots, K-1. \tag{14}$$

As long as a null hypothesis is rejected, r_{H0} is increased by 1 and tested again. The procedure stops for an accepted $r = r_{H0}$ or if the maximal $r_{H0} = K-1$ has been rejected in favor of $r = K$. Correspondingly, the number of stochastic trends[6] $r_\perp = K - r_{H0}$ declines in each step from K down to 1. Note that, under a given r_{H0}, the *maximum eigenvalue test* builds on the H_1 of $\mathrm{rk}(\Pi) = r_{H0}+1$ and the *trace test* on $\mathrm{rk}(\Pi) = K$. Although both tests are applicable to panel extensions, only variants of the trace test have been adopted by the panel literature and are thus introduced here.

LR-test. Johansen (1988) develops a nowadays predominant *likelihood ratio* test which compares the maximized likelihood $\mathcal{L}(r_{H0})$ from the rank-restricted model against the likelihood $\mathcal{L}(K)$ from the full-rank model with K variables. This trace statistic denotes as

$$\begin{aligned} \lambda^{\mathrm{LR}}\,(r_{H0}) &= -2\left[\ln \mathcal{L}(r_{H0}) - \ln \mathcal{L}(K)\right] \\ &= -T \sum_{j=r_{H0}+1}^{K} \ln(1-\hat{\lambda}_j) \;\xrightarrow{d}\; Z_{r_\perp}. \end{aligned} \tag{15}$$

The estimated eigenvalues $\hat{\boldsymbol{\lambda}}$ are the squared canonical correlation coefficients obtained from the RRR in Section 2.1.1 and resemble the ordered eigenvalues of coefficient matrix Π. Correspondingly, this test assesses whether any of the $r_\perp$ smallest eigenvalues is significantly different from zero and thus increases the rank of Π. Under H_0, the asymptotic test distribution is a function of $r_\perp$-dimensional Brownian motions as outlined in Eq. (18).

LM-test. Albeit its shadow existence in the individual time series methodology, the *Lagrange multiplier* test adopted by Luukkonen, Ripatti, and Saikkonen (1999) has proven useful for the panel test procedure proposed by Breitung (2005) as it can serve as a vehicle to introduce the panel-homogeneous cointegrating vectors into an individual testing procedure. In order to make the cointegration rank r_{H0} testable by LM, the concentrated Model (6) is extended by $\phi\beta_\perp' R_1$ and multiplied by $\alpha_\perp'$. Consequently in the auxiliary model

$$\begin{aligned} R_0 &= \alpha\beta' R_1 + \phi\beta_\perp' R_1 + U^{LM} \\ \text{resp.} \quad \alpha_\perp' R_0 &= \phi^*\beta_\perp' R_1 + \alpha_\perp' U^{LM}, \end{aligned} \tag{16}$$

the extension is $H_0 : \alpha_\perp'\phi = \phi^* = 0_{r_\perp \times r_\perp}$ under r_{H0}, but converts $\Pi = \alpha\beta' + \phi\beta_\perp'$ into a full-rank matrix if the data suggest the alternative $H_1 : \mathrm{rk}(\Pi) = K$. After inserting suitable

[6]Note that in the R code of **pvars** and also often in the considered literature the variable d denotes the number of stochastic trends instead of the less ambiguous $r_\perp$.

estimates for the orthogonal complements,[7] the coefficient matrix ϕ^* is estimated by OLS with regressand $E := \widehat{\alpha_\perp}' R_0$ and regressor $W := \widehat{\beta_\perp}' R_1$. The test statistic for $H_0 : \phi^* = 0$ is constructed according to the here all-equivalent LR-, Wald-, or LM-principle and is thus directly calculated with

$$\lambda^{\text{LM}}(r_{H0}) = T \text{ tr} \left[EW' (WW')^{-1} WE' (EE')^{-1} \right] \xrightarrow{d} Z_{r_\perp}. \tag{17}$$

This LM-test statistic and Johansen's LR-test statistic from Eq. (15) are asymptotically identical and so are their test distributions.

Test distribution.[8] Under H_0 and $T \to \infty$, the theoretical distributions converge to

$$Z_{r_\perp} = \text{tr} \left[\int_0^1 \left(\mathrm{d}W^\circ_{r_\perp} \right) W^{\bullet\prime}_{r_\perp} \left(\int_0^1 W^\bullet_{r_\perp} W^{\bullet\prime}_{r_\perp} \right)^{-1} \int_0^1 W^\bullet_{r_\perp} \left(\mathrm{d}W^\circ_{r_\perp} \right)' \right] \tag{18}$$

$$\text{with} \quad W^\bullet_{r_\perp} := \begin{cases} W_{r_\perp}(s) & \text{for } \textit{Case 1} \text{ or } \textit{SL\_mean} \\ W_{r_\perp}(s) - \int_0^1 W_{r_\perp}(s)\mathrm{d}s & \text{for } \textit{Case 2} \\ \left[\begin{smallmatrix} W_{r_\perp - 1}(s) \\ s \end{smallmatrix} \right] - \int_0^1 \left[\begin{smallmatrix} W_{r_\perp - 1}(s) \\ s \end{smallmatrix} \right] \mathrm{d}s & \text{for } \textit{Case 3} \\ W_{r_\perp}(s) - sW_{r_\perp}(1) & \text{for } \textit{Case 4} \text{ or } \textit{SL\_trend} \end{cases}$$
$$\text{and} \quad \mathrm{d}W^\circ_{r_\perp} := \begin{cases} \mathrm{d}W_{r_\perp}(s) & \text{for } \textit{SL\_mean} \text{ or any } \textit{Case} \\ \mathrm{d}W_{r_\perp}(s) - \mathrm{d}sW_{r_\perp}(1) & \text{for } \textit{SL\_trend}. \end{cases} \tag{19}$$

The $r_\perp$-dimensional vector $W_{r_\perp}(s)$ stacks the independent standard Brownian motions. In Eq. (18), their vector products are integrated over their complete domain $s \in [0, 1]$. If the VECM accommodates deterministic regressors, $W^\bullet_{r_\perp}(s)$ must be specified accordingly. Eq. (19) lists the Brownian motions resp. bridges for the conventional cases from Table 2. In contrast to those, trend breaks are specific to the periods of their occurrence and the relative position of these periods within the sample affects $Z_{r_\perp}$ additionally. Also note that any unrestricted deterministic term which is accumulated under $I(1)$, e.g. the constant in *Case 3*, dominates the stochastic trends and thus replaces a Brownian motion in $W_{r_\perp}(s)$.

The asymptotic distribution of $Z_{r_\perp}$ is non-standard, therefore critical values have been simulated e.g. by Osterwald-Lenum (1992). For this, $W_{r_\perp}$ is substituted by a repetitive simulation of $r_\perp$-dimensional vectors of Gaussian random walks with sufficiently[9] large sample size. Also, $Z_{r_\perp}$ can be approximated by the gamma distribution $\Gamma(\mathsf{s}, \mathsf{r})$, whose continuous probability density function offers the advantage of retrieving p-values conveniently. In order to do so, its shape- and rate-parameters are equated with

$$\mathsf{s} = \frac{\mathbb{E}(Z_{r_\perp})^2}{\text{Var}(Z_{r_\perp})} \qquad \text{and} \qquad \mathsf{r} = \frac{\mathbb{E}(Z_{r_\perp})}{\text{Var}(Z_{r_\perp})}. \tag{20}$$

Simulated values for the moments $\mathbb{E}(Z_{r_\perp})$ and $\text{Var}(Z_{r_\perp})$ can be found for example in Larsson *et al.* (2001) and Breitung (2005).[10] Like the unknown theoretical distribution $Z_{r_\perp}$, their

[7] For his panel cointegration test, Breitung (2005) uses the individual first-step estimate for α_i and the pooled second-step estimate for β_K under rank-restriction r_{H0}. In **pvars**, their orthogonal complements are then calculated via the QR-decomposition as done in the **MASS** package (Venables and Ripley 2002).

[8] See Johansen (1996, Ch. 11.2), Lütkepohl (2005, Ch. 8.2), and Trenkler (2008, p. 24, Eq. 2.9).

[9] For example, Johansen (1996, Ch. 15) recommends $T = 400$. Breitung (2005, p. 171, App. B) employs $T = 500$ and Örsal and Droge (2014) $T = 1000$ in order to simulate the respective moments of $Z_{r_\perp}$.

[10] Larsson *et al.* (2001) provide moments of $Z_{r_\perp}$ only for *Case 1* and Breitung (2005) for *Case 2* to *Case 4* in order to center and scale the LR-bar panel test statistic as in Eq. (28). In **pvars**, these moments are stored in the list object `coint_moments`.

simulation depends on *(i)* the number of stochastic trends $r_\perp$ under H_0, *(ii)* the deterministic term, and *(iii)* the number of weakly exogenous variables in a conditional VECM. Hence, the different test specifications imply a grid of simulation setups to run.

More flexibility towards adapting these specifications is offered by response surface approximation of the moments as employed by Doornik (1998). For each *Case*, he estimates response surface coefficients[11] of a polynomial regression model $f(r_\perp)$ which "predicts" the moments of $Z_{r_\perp}$ for Johansen (1996). Trenkler (2008) derives response surface coefficients specifically for the trend-adjusted tests of Eq. (10). TSL (2008) tabulate these coefficients for trend-adjusted tests which accommodate structural trend breaks in the cointegration relationship. Likewise, JMN (2000) consider structural breaks in the constant and in the linear trend of the innovative model, which KN (2019) generalize for conditional VECM by Doornik (1998, Ch. 9). Although both models, additive and innovative, can cope with several breaks in principle, the regression models $f\left(r_\perp, \frac{\tau}{T}\right)$ by JMN (2000), KN (2019), and TSL (2008) contain polynomials of up to two break periods $\boldsymbol{\tau}$ only. **pvars** resorts to the respective approximation automatically in order to comply with the test specification selected by the user.

Identifying structure

Based on the given rank-restriction r and structural shocks $\boldsymbol{\epsilon}_t = \mathsf{B}^{-1}\boldsymbol{u}_t$, the VECM (2) can be transformed according to the *Granger representation theorem* (GRT) into

$$\begin{aligned} \boldsymbol{y}_t &= \Xi\mathsf{B}\sum_{j=1}^{t}\boldsymbol{\epsilon}_j + \sum_{j=0}^{\infty}\Xi_j^*\mathsf{B}\boldsymbol{\epsilon}_{t-j} + \Xi\Phi\sum_{j=1}^{t}\boldsymbol{d}_j + \sum_{j=0}^{\infty}\Xi_j^*\Phi\boldsymbol{d}_{t-j} + \boldsymbol{y}_0^* \quad \text{with} \\ \Xi &= \beta_\perp\left[\alpha_\perp'\left(I_K - \sum_{j=1}^{p-1}\Gamma_j\right)\beta_\perp\right]^{-1}\alpha_\perp'. \end{aligned} \tag{21}$$

The initial vector $\boldsymbol{y}_0^*$ and the deterministic terms are usually dropped in order to trace the responses to an isolated structural impulse $\epsilon_{(k)}$. The number of stochastic trends $\mathrm{rk}(\Xi) = K - r = r_\perp$ follows from the $K \times r_\perp$ orthogonal complements $\alpha_\perp$ and $\beta_\perp$ for the $K \times r$ loadings α and cointegrating vectors β_K. Their different specifications illustrate the role of the components in (21) and their options for structural restrictions:

- If $r_\perp = 0$, the stochastic trends cancle out and the sole multiplier matrices $\Xi_j^*\mathsf{B}$ reflect the IRF of the implied *stable* SVAR model. Hence, just-identification requires the $K(K-1)/2$ restrictions to be imposed on the instantaneous effects B exclusively. A simple example is Sims (1980), who assumes recursive (short-run) causality such that the structural effects on impact can be calculated by $\widehat{\mathsf{B}} = \mathrm{chol}\left(\widehat{\Sigma}_u\right)$.
- If $r_\perp = K$, VECM (2) loses its error correction term and is estimated as a first-differenced VAR model by OLS (9). Since $\alpha_\perp = \beta_\perp = I_K$, matrix $\Xi = \left(I_K - \sum_{j=1}^{p-1}\Gamma_j\right)^{-1} = \Gamma(1)^{-1}$ is full-rank and $K(K-1)/2$ long-run restrictions are sufficient for identifying this SVAR of *growth rates* $\Delta\boldsymbol{y}_t$. In this example, recursive long-run causality is imposed on the structural long-run effects $\Xi\mathsf{B}$ such that $\widehat{\mathsf{B}} = \widetilde{\Gamma}(1)\ \mathrm{chol}\left(\widehat{\Sigma}_\infty\right)$ is calculated via the long-run covariance $\Sigma_\infty := \Xi\mathsf{B}\mathsf{B}'\Xi' = \Gamma(1)^{-1}\Sigma_u\Gamma(1)'^{-1}$.

[11] In **pvars**, the response surface coefficients are stored in the list object `coint_rscoef`.

- A cointegration rank r implies $r_\perp$ permanent shocks at minimum and conversely r transitory shocks at maximum. Identifications schemes in the tradition of King, Plosser, Stock, and Watson (1991) assume that $\Xi\mathsf{B}$ is rank-deficient due to r columns of $\mathbf{0}_K$. Each $\mathbf{0}_K$ suppresses the stochastic trends and thus defines a transitory shock in $\boldsymbol{\epsilon}_t$. Other formulations of the rank-deficiency may lead to fewer transitory shocks.

ML-estimation.[12] The impact matrix B can be estimated by ML via $\mathcal{L}(\alpha, \beta, \Gamma, \mathsf{B})$ as an extension of Eq. (5). Inserting the reduced-form estimates from Section 2.1.1 facilitates a conditional ML-estimation of B as Step **4.** with the concentrated likelihood function

$$\ln \mathcal{L}_c(\mathsf{B}) = -\frac{KT}{2}\ln(2\pi) - \frac{T}{2}\ln(\det(\mathsf{B}))^2 - \frac{1}{2}\mathrm{tr}\left[\mathsf{B}'^{-1}\mathsf{B}^{-1}\widetilde{\Sigma}_u\right]. \tag{22}$$

This constrained maximization problem has no closed-form solution. Hence, Breitung, Brüggemann, and Lütkepohl (2004) maximize $\mathcal{L}_c(\mathsf{B})$ with respect to the free parameters subject to the identifying restrictions numerically by the scoring algorithm of Amisano and Giannini (1997). If the number of short- and long-run restrictions suffice just-identification only, the equality $\widetilde{\Sigma}_u = \widetilde{\mathsf{B}}\widetilde{\mathsf{B}}'$ holds and the scoring algorithm serves as a non-linear equation solver.

2.2. Panel methods

Estimating cointegrated VAR models

In macroeconomic panels, the idiosyncrasies of countries can subvert the objective to increase estimates' precision with pooled data. Deterministic effects like intercepts are usually kept individual-specific in panel data analysis, but even the assumption of homogeneous slope coefficients can be too restrictive. Pesaran and Smith (1995) demonstrates how systematic differences between individual dynamics lead to inconsistent estimates and consequently deny poolability. As an economically plausible middle ground, Pesaran, Shin, and Smith (1999) propose a selective pooling for single-equation models, where only the long-run relationship is homogeneous across all individuals, while the short-run dynamics are considered as heterogeneous. The following panel estimators adopt this principle into a system of equations and restrict the r cointegrating vectors $\beta_i = \beta\ \forall\ i$. Although the individual VECM (2) does not require the basic normalization of the error correction term as single-equation models do, a panel-wide normalization becomes necessary in order to extract the homogeneous β from the individual coefficient matrices $\Pi_i = \alpha_i\beta'$.

Two-step estimator.[13] Breitung (2005) extends the individual two-step estimator by Ahn and Reinsel (1990) to a panel estimator for β. Based on the panel-wide normalization $\Pi_i = \alpha_i[I_r : \mathbf{B}] = [\alpha_i I_r : \alpha_i\mathbf{B}]\ \forall\ i$, he transforms the concentrated Model (6) into

$$R_{0,i} - \alpha_i I_r R_{1,i}^{(1)} = \alpha_i \mathbf{B} R_{1,i}^{(2)} + U_i \quad \text{with} \quad R_{1,i} = \begin{bmatrix} R_{1,i}^{(1)} \\ R_{1,i}^{(2)} \end{bmatrix} \tag{23}$$

[12] See Vlaar (2004), who imposes linear short- and long-run restriction with restriction matrices for the cases of just- and over-identification. Compare also to Hamilton (1994, Ch. 11, p. 331), Lütkepohl (2006, Ch. 3.2, p. 84), and Kilian and Lütkepohl (2017, Ch. 11.2.2 and p. 315) as a full-information MLE.

[13] See also Lütkepohl (2005, Ch. 7.2.2) or Brüggemann and Lütkepohl (2005) about GLS estimation of the individual cointegration parameters. In fact, Eq. (25) reduces to this GLS estimator if there is only $N = 1$ individual entity. Moreover, Lütkepohl (2005, Ch. 7.3.2) and Breitung (2005, p. 159) consider restricted β.

partitioned into $R_{1,i}^{(1)}$ as the r upper time series and $R_{1,i}^{(2)}$ as the remaining $r_\perp$ time series and n_1 restricted deterministic regressors. The left-multiplication $\alpha_i^{(+)}\alpha_i = I_r$ eliminates the individual loadings α_i in Eq. (23) such that Model (6) is further transformed into

$$\alpha_i^{(+)}\left(R_{0,i} - \alpha_i R_{1,i}^{(1)}\right) = \mathbf{B}R_{1,i}^{(2)} + \alpha_i^{(+)}U_i \quad \text{using} \quad \alpha_i^{(+)} = \left(\alpha_i'\Sigma_{u,i}^{-1}\alpha_i\right)^{-1}\alpha_i'\Sigma_{u,i}^{-1}. \tag{24}$$

The *pooled OLS estimator* of the homogeneous $\mathbf{B}$, i.e. the right $r \times (r_\perp + n_1)$ block of the normalized and transposed cointegrating matrix, yields the second step given by

$$\widehat{\mathbf{B}}_{2S} = R_0^{(+)}R_1^{(2)\prime}\left(R_1^{(2)}R_1^{(2)\prime}\right)^{-1} \tag{25}$$

$$\begin{aligned} &\text{with regressand} \quad R_{0,i}^{(+)} := \alpha_i^{(+)}R_{0,i} - R_{1,i}^{(1)} && \text{collected in} \quad R_0^{(+)} := \left[R_{0,1}^{(+)} : \cdots : R_{0,N}^{(+)}\right] \\ &\text{and regressor} \quad R_{1,i}^{(2)} && \text{collected in} \quad R_1^{(2)} := \left[R_{1,1}^{(2)} : \cdots : R_{1,N}^{(2)}\right]. \end{aligned}$$

For feasible estimation, inserting any $\sqrt{T}$-consistent estimates for α_i and $\Sigma_{u,i}$ leaves the limiting distribution of the estimator (25) unaffected. This first step may be the individual ML-estimation of each VECM, which concentrates out the short-run effects and unrestricted deterministic terms $\Gamma_i Z_{2,i}$ from the observed time series as well. Under rank-restriction r, the second-step estimates $\widehat{\mathbf{B}}_{2S}$ thereby $T\sqrt{N}$-consistently (Breitung 2005, p. 156, Th. 1). Overall, this procedure thus follows the steps of the ML-estimation, but sets a pooled estimation on top of the individual reduced-rank regressions (8). Like in third step of the individual ML-estimation, the cointegrating matrix $\widehat{\beta}' = \left[I_r : \widehat{\mathbf{B}}_{2S}\right]$ can be reintroduced into the conditional OLS (9) of the individual parameters, for which product moments are already available from the first step. Thereof, the estimates for α_i and $\Sigma_{u,i}$ – now conditional on the pooled $\widehat{\beta}$ – can be used in Eq. (25) again. Hlouskova and Wagner (2010, p. 195) expand this principle and iterate conditional OLS and second-step estimation alternately until reaching convergence of the estimates $\widehat{\mathbf{B}}_{2S}$.

Deterministic term. Since the first estimation step (6) removes the unrestricted regressors $\boldsymbol{d}_{2,it}$ from the error correction Model (2), the coefficients Φ_{2i} are individual-specific by construction. In consequence, the homogeneity restriction $\beta_{K,i} = \beta_K$ in

$$\Pi_i := [\Pi_{K,i} : \Phi_{1,i}] = \alpha_i\,[\beta_K' : \beta_{0i}'] = \alpha_i\left[I_r : \mathbf{B}^{(2)} : \mathbf{B}_i^{(3)}\right] \tag{26}$$

implies one additional choice for specifying the innovative deterministic term in Table 2:[14] The coefficients for $D_{1,i}M_{2i}$ can be defined as having some homogeneous $\beta_{0i} = \beta_0\ \forall\ i$ or individual-specific effects β_{0i}. For example, *Case 2* with heterogeneous intercepts β_{0i} and $p_i = 1 \forall i$, i.e. without the need to correct for lagged short-run effects by M_{2i}, adds *fixed-effects* to the auxiliary Model (24). The regressors of such individual-specific $\mathbf{B}_i^{(3)}$ are partialled out from the pooled regression beforehand, while $\mathbf{B}^{(2)}$ is simply extended by the $r \times n_1$ homogeneous β_0' and estimated by the pooled $\widehat{\mathbf{B}}_{2S}$ of (25). Note however that $\Phi_{1i} = \alpha_i\beta_0'$ remains heterogeneous irrespective of the homogeneity restriction on β_0.

[14] Groen and Kleibergen (2003, Ch. 4.1) give an overview on the different combinations that emerge from this single additional option.

Testing the cointegration rank

Like the individual tests in Section 2.1.2, the following panel tests evaluate the hypothetical cointegration rank $r_{H0} = 0, ..., K-1$ within a sequential testing procedure. Each of the N individuals must have observations on the same K variables so that the hypothesis pair on the heterogeneous ranks r_i in the individual VAR processes $\boldsymbol{y}_{it}$ can be stated as

$$\begin{aligned} H_0 : \bar{r} = r_{H0} \quad &\text{versus} \quad H_1 : \text{rk}\,(\Pi_i) > r_{H0} \text{ for some } i \\ \text{with } \bar{r} &= \max\{\text{rk}\,(\Pi_i) \mid i = 1, .., N\}. \end{aligned} \tag{27}$$

While a statistical heterogeneity in r_i lets the sequential procedure find the maximum rank among the individual processes, the theoretical context might allow for the more restrictive assumption of homogeneous cointegration ranks across all individuals $i = 1, \ldots, N$. Accordingly, the hypothesis pair (27) would adopt a common rank $r_i = \bar{r} = r$. Carrion-i Silvestre and Surdeanu (2011, p. 9) assume this "for the panel data based procedure to make sense" and the homogeneous cointegrating vectors $\beta_{K,i} = \beta_K$ in Breitung (2005) and Groen and Kleibergen (2003) actually have the statistical implication of $r_i = r$.

Pooled cointegrating vectors. The more restrictive assumption of homogeneous β_K allows for a pooled estimation to further increase the test power. Only in this case, the *pooling* and the *combination* approach of the panel test differ. After estimating individual models with pooled β_K, Breitung (2005) combines the individual LM-test (17) results by their averaged statistics (28). Originally, he does not make use of the gamma distribution to derive p-values of the LM-tests. Since this option follows straightforwardly from the asymptotic equivalence of LM- and LR-test, **pvars** implements their meta-analytical combination, too.[15] Thereby, both combination approaches are available to shrink the vectors of N test results into scalar decisions. The following section refers to these combination approaches, which are congruent with the pooling approaches of Table 1.

Combination approaches. Extending the combination idea from Im *et al.* (2003) to the multivariate case of K variables, Larsson *et al.* (2001) use the *cross-sectional average* of individual test statistics $\lambda_i^\bullet$, which are standardized by the first and second moment. Accordingly, the so-called LR-bar test statistic $\Upsilon_{\overline{LR}}$ is calculated as

$$\begin{aligned} \overline{\lambda}(r_{H0}) &= \frac{\sum_{i=1}^{N} \lambda_i^\bullet (r_{H0})}{N}, \\ \Upsilon_{\overline{LR}} &= \frac{\sqrt{N}\left(\overline{\lambda}(r_{H0}) - \mathbb{E}\,(Z_{r_\perp})\right)}{\sqrt{\text{Var}\,(Z_{r_\perp})}} \xrightarrow{d} \mathcal{N}(0,1) \end{aligned} \tag{28}$$

and follows asymptotically a standard normal distribution under H_0 as $T \to \infty$ and subsequently $N \to \infty$. This holds also if $T, N \to \infty$ simultaneously, but T has a faster limiting rate such that $\sqrt{N}/T \to 0$. The combination approach itself is well-established for all kinds of panel tests. The key contribution by Larsson *et al.* (2001) is to prove that the moments of the asymptotic distribution $Z_{r_\perp}$ exist.[16] Likewise, Örsal and Droge (2014) present and utilize a proof for the detrending panel SL-test with the deterministic type *SL_trend*. Additionally,

[15] Choi (2001, p. 268) shows for the univariate case of panel unit root tests that especially the inverse normal test has a favorable power in comparison to the panel test of averaged statistics.

[16] See also the corrigendum by Örsal and Droge (2011).

pvars offers the specification *SL_mean* because the existence of the related moments follows directly from the identity of $Z_{r_\perp}$ for *SL_mean* and *Case 1* – see Eq. (19).

Following the *meta-analytical combination* of p-values, Choi (2001) uses different combination methods on the individual p-values denoted by p_i. The first combination test, which Fisher (1932) originally proposed and Maddala and Wu (1999) embraced for tests in non-stationary panels, is the inverse chi-square test. Its statistic is stated as

$$P = -2\sum_{i=1}^{N} \ln(p_i) \xrightarrow{d} \chi^2(2N) \tag{29}$$

under the asymptotics of $T_i \to \infty$ while N being fixed. In the limit of N, the distribution of P degenerates however. Hence for panels with many[17] individuals, Choi (2001) standardizes P by the moments of the χ^2-distribution and proposes the modified statistic

$$\begin{aligned} P_m &= \frac{\sum_{i=1}^{N}(\ln(p_i)+1)}{\sqrt{N}} \\ &= \frac{P-2N}{\sqrt{4N}} \xrightarrow{d} \mathcal{N}(0,1), \end{aligned} \tag{30}$$

when $T_i \to \infty$ and then additionally $N \to \infty$. Furthermore, assuming the same limiting data dimensions, Choi (2001) uses the inverse normal test

$$Z = \frac{\sum_{i=1}^{N}\Phi^{-1}(p_i)}{\sqrt{N}} \xrightarrow{d} \mathcal{N}(0,1) \tag{31}$$

from Stoufer, Suchman, DeVinney, and Williams (1949) for non-stationary panel data analysis. Here, $\Phi^{-1}(\bullet)$ denotes the probit function, i.e. the inverse of the standard normal cumulative distribution.

PANIC. The *Panel Analysis of Non-stationarity in Idiosyncratic and Common components* by Bai and Ng (2004) accounts for cross-sectional dependency stemming from unobserved common factors, which could be stationary, integrated, or both coexistent. Arsova and Örsal (2017; 2018) employ PANIC on VAR systems so that their multivariate extension then allows for tests on the coingration rank within the idiosyncratic VECM of the individuals. The additive model is stated as

$$\begin{aligned} \boldsymbol{y}_{it} &= \Lambda_i' \boldsymbol{F}_t + \boldsymbol{y}_{it}^{id}, \\ \boldsymbol{y}_{it}^{id} &= \boldsymbol{\mu}_{0i} + \boldsymbol{\mu}_{1i} t + \boldsymbol{y}_{it}^{dt}, \\ \boldsymbol{y}_{it}^{dt} &= A_{i1}\boldsymbol{y}_{i,t-1}^{dt} + \ldots + A_{i,p_i}\boldsymbol{y}_{i,t-p_i}^{dt} + \boldsymbol{u}_{it}^{dt} \\ (1-L)\boldsymbol{F}_t &= C(L)\boldsymbol{u}_t^F. \end{aligned} \tag{32}$$

The idea of PANIC is to perform separate analyses on the common factor $\boldsymbol{F}_t$ and the idiosyncratic component $\boldsymbol{y}_{it}^{id}$, both being part of the observable multivariate time series $\boldsymbol{y}_{it}$. Hence firstly, factors $\boldsymbol{f}_t$ and heterogeneous loadings Λ_i are estimated via PCA on the first-differenced data matrix of dimension $(T-1) \times (K \cdot N)$. If unknown, the number of underlying factors

[17] Choi (2001, p. 268) states for panel unit root tests that, even with samples of $N = 100$, P_m does not follow its asymptotic distribution closely. In contrast to the two Fisher tests, the test size of the inverse normal test and of the averaged statistics is more robust to any choice of N.

can be chosen according to information criteria by Bai and Ng (2002) or Onatski (2010) as illustrated in Section 4.1. The factors $\hat{\boldsymbol{F}}_t = \sum_{j=1}^{t} \hat{\boldsymbol{f}}_j$ with initial $\hat{\boldsymbol{f}}_1 = \mathbf{0}_L$ are re-accumulated into levels to estimate the idiosyncratic component $\hat{\boldsymbol{y}}_{it}^{id} = \boldsymbol{y}_{it} - \hat{\Lambda}_i' \hat{\boldsymbol{F}}_t$. While PCA generates orthogonal factors $\hat{\boldsymbol{f}}_t$ normalized to an identity covariance matrix, the re-accumulated factors $\hat{\boldsymbol{F}}_t$ are usually *oblique* and described by a single VAR model. The VAR process of $\boldsymbol{F}_t$ can further contain common structural breaks from $\boldsymbol{y}_{it}$ such that, for instance, Örsal (2017) resorts to JMN (2000) for testing the cointegration rank of $\hat{\boldsymbol{F}}_t$.

Since $\boldsymbol{F}_t$ is assumed to be the exclusive source of cross-sectional dependence, the idiosyncratic VECM of $\boldsymbol{y}_{it}^{id}$ are independent across individuals i and the basic methodology of first-generation tests is valid again. Örsal and Arsova (2017) use the approach of meta-analytical combination of the idiosyncratic p-values and Arsova and Örsal (2018) the averaged test statistics. Accordingly, the nested hypotheses of Eq. (27) about r_{H0} refer to idiosyncratic cointegration within $\boldsymbol{y}_{it}^{id}$ and, for conclusions on conintegration within the observable $\boldsymbol{y}_{it}$, the non-stationary properties of $\boldsymbol{F}_t$ need to receive attention. Individual unit root and cointegration tests can be applied to the common factors $\boldsymbol{F}_t$ and their VAR process. If all series in $\boldsymbol{F}_t$ are stationary, panel test results on $\boldsymbol{y}_{it}^{id}$ hold equally for $\boldsymbol{y}_{it}$. If $\boldsymbol{F}_t$ contains stochastic trends, those enter the the observable variables $\boldsymbol{y}_{it}$, too. In this case, they can confound the interpretation, for example, if some of the heterogeneous loadings Λ_i are zero and thus do not let stochastic trends in $\boldsymbol{F}_t$ drive the variables $\boldsymbol{y}_{it}$ equally.

For testing the cointegration rank in each idiosyncratic component, Arsova and Örsal (2017; 2018) consider two individual procedures: The Johansen test applied to $\boldsymbol{y}_{it}^{id}$ directly and the SL-test applied to $\boldsymbol{y}_{it}^{dt}$ after the GLS-based trend adjustment. The SL-test based on the defactored and detrended series $\boldsymbol{y}_{it}^{dt}$ assesses only the null hypothesis of no cointegration. For $r_{H0} > 0$ in the sequential test, the $r_\perp = K - r_{H0}$ stochastic trends in the rank-restricted VECM are calculated as $\beta_{\perp i}' \boldsymbol{y}_{it}^{id}$. Arsova and Örsal (2017; 2018) then apply the SL-test to the presumably $r_\perp$ stochastic trends under the H_0 of no cointegration in order to test the equivalent H_0 of rank r_{H0} in the idiosyncratic process of $\boldsymbol{y}_{it}^{id}$.

Defactoring adds a linear trend to the estimated series $\hat{\boldsymbol{y}}_{it}^{id}$ inevitably, which affects the distribution of $Z_{r_\perp}$ for both, SL- and Johansen test. For deriving individual p-values via the gamma approximation as in Eq. (20) or for standardizing the LR-bar panel test statistic as in Eq. (28), the panel tests rely on the moments of $Z_{r_\perp}$ which Arsova and Örsal (2018, Appendix, Tab. A.1) have simulated and tabulated. The moments depend on $r_\perp$ and the prescribed deterministic trend, but not on the common factors $\boldsymbol{F}_t$. Likewise, they can be calculated via the response surface regression model by Trenkler (2008) using the coefficients for the case of a deterministic trend.[18] Note that Örsal and Arsova (2017) do not propose to combine individual p-values from defactored Johansen tests originally. However, **pvars** implements this in accordance with the convergence result of $Z_{r_\perp}$ implied by Arsova and Örsal (2018, p. 1041, Th. 3.4).

Correlated probits. Cross-sectional dependence between the individual entities induces correlation between the probits $\Phi^{-1}(p_i)$ of the individual cointegration tests. In order to correct for the this and robustify the inverse normal test of Eq. (31), Arsova and Örsal (2021) combine individual p-values from the SL-procedure with Hartung's (1999) modifications and

[18]For an illustration, compare the results of the R commands:

```
R> pvars:::coint_moments$SL_trend[12:1, ]
R> pvars:::aux_CointMoments(dim_K=12, rs_coef=pvars:::coint_rscoef[["SL_trend"]])
```

with their own *correlation-augmented inverse normal* (CAIN) test. The panel test statistic for the modified combination approach is given by

$$\begin{aligned} t(\tilde{\rho}) &= \frac{\sum_{i=1}^{N} \Phi^{-1}(p_i)}{\sqrt{N + (N^2 - N) \cdot \tilde{\rho}^{\bullet}_{\text{probit}}}} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{with} \\ \tilde{\rho}^{\text{HA}}_{\text{probit}} &= \hat{\rho}^{*}_{\text{probit}} + \kappa \cdot \sqrt{\frac{2}{(N+1)}} \left(1 - \hat{\rho}^{*}_{\text{probit}}\right) \quad \text{or} \\ \tilde{\rho}^{\text{CAIN}}_{\text{probit}} &= g\left(\rho_{\epsilon}, K, r_{H0}\right). \end{aligned} \tag{33}$$

The correction factor $\tilde{\rho}^{\text{HA}}_{\text{probit}}$ by Hartung (1999) is based on his unbiased and consistent estimator $\hat{\rho}^{*}_{\text{probit}} = \max\{-\frac{1}{N-1},\ 1 - s^2_{N-1}\left(\Phi^{-1}(p_i)\right)\}$ of the probit correlation, where $s^2_{N-1}(\cdot)$ denotes the unbiased sample variance. For the scaling factor κ, he proposes $\kappa_1 = 0.2$ and, additionally, $\kappa_2 = 0.1 \cdot \left(1 + \frac{1}{N-1} - \hat{\rho}^{*}_{\text{probit}}\right)$ "working mainly for smaller $\hat{\rho}^{*}_{\text{probit}}$" (Hartung 1999, p. 853). He does not state any decision criterion to distinguish between the intervals of smaller and larger values of $\hat{\rho}^{*}_{\text{probit}}$ however.

The correction factor $\tilde{\rho}^{\text{CAIN}}_{\text{probit}}$ by Arsova and Örsal (2021) is the empirical estimate of the average correlation between the individual probits. For translating cross-sectional dependency within the data panel into $\tilde{\rho}^{\text{CAIN}}_{\text{probit}}$, the authors provide response surface regression coefficients for the regression model $g\left(\rho_{\epsilon}, K, r_{H0}\right)$. Therein, the number of endogenous variables K and the tested cointegration rank r_{H0} follow directly from the model resp. test specification. The *average absolute pairwise cross-sectional correlation* ρ_{ϵ} between the residuals is estimated for $r_{H0} = 0$ only and then used for all r_{H0} of the sequential testing procedure up to $r_{H0} = K - 1$. The estimator for this mean sample correlation of the residuals within the *same* variable $k = 1, \dots, K$ and between the different individuals is

$$\hat{\rho}_{\epsilon} = \frac{2}{K \cdot N \cdot (N-1)} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=i+1}^{N} |\hat{\rho}_{ki,kj}|. \tag{34}$$

Since Arsova and Örsal (2021) assume that the pairwise correlations between the residuals of *different* variables converge to zero for $N \to \infty$, the sample correlations $\hat{\rho}_{ki,k^*j},\ k \neq k^*$, are omitted in Eq. (34). In doing so, the cross-sectional correlation between variables does not reduce the average of the presumably stronger "within"-correlations, which mitigates the hazard of underestimating ρ_{ϵ}. Only if the empirical "between"-correlation surpasses "within"-correlation, $\hat{\rho}_{\epsilon}$ could actually understate cross-sectional dependence.

If the individual, potentially heterogeneous break periods $\boldsymbol{\tau}_i$ and the number of breaks are known, this third-generation panel test can also respect trend breaks in the cointegration relationship. For this, the individual GLS-based trend-adjustment by Trenkler *et al.* (2008) removes the deterministic component from the observed time series of Eq. (10) including up to two trend breaks. Then, the standard LR-test procedure is applied under consideration of the trend-break specific test distribution. The individual p-values are combined as described in Eq. (33) in order to account for the cross-sectional dependence.

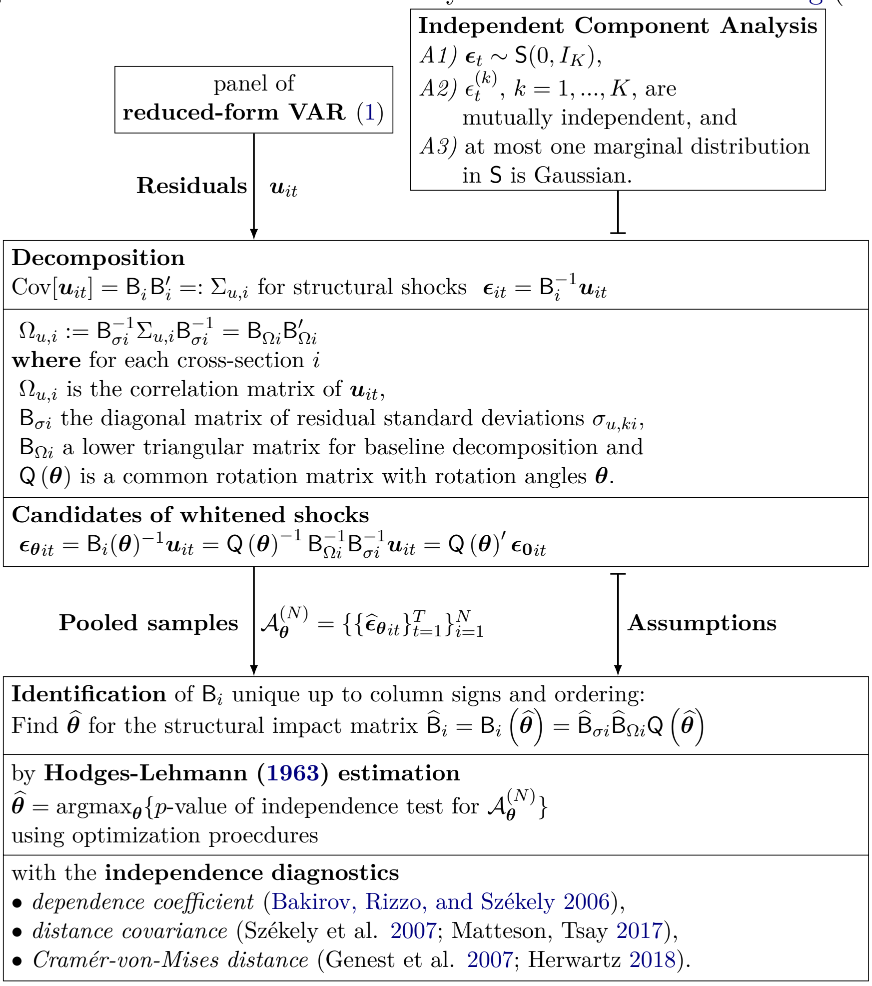
Identifying structure

In comparison to the individual time series analysis, the cross-section dimension of the panel data enables additional ways to identify structural shocks $\boldsymbol{\epsilon}_{it} = \mathsf{B}_i^{-1}\boldsymbol{u}_{it}$ from the K reduced-form errors $\boldsymbol{u}_{it}$ of VAR (1). The identification across the individuals $i = 1, \dots, N$ may start with normalizing $\boldsymbol{u}_{it}$ to unit-variance by $\mathsf{B}_{\sigma i} = \operatorname{diag}(\sigma_{u,1i}, \ldots, \sigma_{u,Ki})$ such that the structural decomposition concentrates on the residual correlation matrix

$$\begin{aligned} \Omega_{u,i} &:= \mathsf{B}_{\sigma i}^{-1} \Sigma_{u,i} \mathsf{B}_{\sigma i}^{-1} = \mathsf{B}_{\Omega i} \mathsf{Q}_i \mathsf{Q}_i' \mathsf{B}_{\Omega i}' \\ \text{with } \mathsf{B}_{\Omega i} &:= \operatorname{chol}(\Omega_{u,i}) \text{ and } \mathsf{Q}_i \mathsf{Q}_i' = \mathsf{Q}_i \mathsf{Q}_i^{-1} = I_K. \end{aligned} \tag{35}$$

The usual Cholesky decomposition $\mathsf{B}_i(\mathbf{0}) := \operatorname{chol}(\Sigma_{u,i}) = \mathsf{B}_{\sigma i} \cdot \operatorname{chol}(\Omega_{u,i})$ generates *whitened* shocks $\boldsymbol{\epsilon}_{\mathbf{0}it} = \mathsf{B}_i(\mathbf{0})^{-1} \boldsymbol{u}_{it}$, which already conform with the defining $\boldsymbol{\epsilon}_{\mathbf{0}it} \sim (0, I_K)$ after scaling $\left(\mathsf{B}_{\sigma i}^{-1}\right)$ and de-correlating $\left(\mathsf{B}_{\Omega i}^{-1}\right)$ the zero-mean vector $\boldsymbol{u}_{it}$. Further, matrix Q_i is orthogonal and can be a product of $K(K-1)/2$ Givens rotation matrices, which map from the tuple of rotation angles $\boldsymbol{\theta}_i$ to the total set of candidate shocks $\boldsymbol{\epsilon}_{\boldsymbol{\theta}it} = \mathsf{Q}(\boldsymbol{\theta}_i)^{-1} \boldsymbol{\epsilon}_{\mathbf{0}it}$ and their impact matrices $\mathsf{B}_i(\boldsymbol{\theta}_i) = \mathsf{B}_i(\mathbf{0})\, \mathsf{Q}(\boldsymbol{\theta}_i)$. As the identification procedures rely on estimated $\widehat{\boldsymbol{u}}_{it}$, the candidate samples $\mathcal{A}_{\boldsymbol{\theta}}^{(\mathrm{i})} = \{\widehat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}it}\}_{t=1}^{T}$ of dimension $K \times T$ involve the usual issues of individual SVAR with small or medium-sized time series. The panel methods presented in the following offer the advantage of *(i)* increasing the power of dependence tests in ICA and of *(ii)* accommodating structural information from cross-sectional dependence.

Figure 1: Panel SVAR identification by ICA in Herwartz and Wang (2024).



Common rotation. Herwartz and Wang (2024) propose and evaluate the pooled identification procedure stylized in Figure 1. An individual decomposition $\boldsymbol{\epsilon}_{\boldsymbol{\theta}it} = \mathsf{Q}(\boldsymbol{\theta})^{-1}\,\mathsf{B}_{\Omega i}^{-1}\mathsf{B}_{\sigma i}^{-1}\boldsymbol{u}_{it}$ allows for a common rotation $\mathsf{Q}(\boldsymbol{\theta})$ of the whitened shocks while accounting for heterogeneous variances and cross-variable correlations in $\boldsymbol{u}_{it}$. Among competing *pooled samples* $\mathcal{A}_{\boldsymbol{\theta}}^{(N)} = \{\{\widehat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}it}\}_{t=1}^{T}\}_{i=1}^{N}$ of dimension $K \times (T \cdot N)$, the ICA then determines the least dependent shocks $\widehat{\boldsymbol{\epsilon}}_{\widehat{\boldsymbol{\theta}}it}$ with optimal $\widehat{\boldsymbol{\theta}}$, from which $\widehat{\mathsf{B}}_i$ is recovered for each individual.

Common shocks. Calhoun *et al.* (2001) propose *group ICA* and apply this to panel data originating from functional Magnetic Resonance Imaging (fMRI) of brains. In the same empirical context, Risk, Matteson, Ruppert, Eloyan, and Caffo (2014) evaluate ICA algorithms, which can be applied to the panel of reduced-form errors $\boldsymbol{u}_{it}$ alike. Accordingly, their model $\boldsymbol{u}_{it} = \mathsf{B}_i\boldsymbol{\epsilon}_t + \boldsymbol{e}_{it}$ consists of L common shocks $\boldsymbol{\epsilon}_t$ and some idiosyncratic noise $\boldsymbol{e}_{it}$ (Risk *et al.* 2014, p. 227). In a two-step PCA, they firstly whiten each individual sample $[\widehat{\boldsymbol{u}}_{i1} : \ldots : \widehat{\boldsymbol{u}}_{iT}]' = \mathsf{U}_i\mathsf{D}_i\mathsf{V}_i'$ by compact *singular value decomposition* such that $\mathcal{A}_i := \sqrt{T}\mathsf{U}_i'$ has an identity covariance matrix I_K. They further factorize the $T \times (K \cdot N)$ concatenated samples $[\mathcal{A}_1' : \ldots : \mathcal{A}_N'] = \mathsf{U}\mathsf{D}\mathsf{V}'$ and utilize the first L columns of left singular-vectors to construct a $L \times T$ baseline sample $\mathcal{A}_{\mathbf{0}}^{(2S)} := \sqrt{T}\mathsf{U}_{1:L}'$. The ICA, which becomes noise-free after this data reduction, then determines the least dependent common shocks $\widehat{\boldsymbol{\epsilon}}_{\widehat{\boldsymbol{\theta}}t}$. The multivariate least squares regression of $\widehat{\boldsymbol{u}}_{it}$ on $\widehat{\boldsymbol{\epsilon}}_{\widehat{\boldsymbol{\theta}}t}$ recovers the $K \times L$ impact matrices $\widehat{\mathsf{B}}_i,\ \forall i = 1\ldots,N$.

ICA. If at most one of the shocks is Gaussian, the *independent component analysis* as established by Comon (1994) can determine the rotation angles $\widehat{\boldsymbol{\theta}}^{\bullet}$ and thereby identify the impact matrix $\widehat{\mathsf{B}}_i^{\bullet}$ of the methods $\bullet \in \{(i),(N),(2S)\}$. For this purpose, dependence measures $\mathcal{D}(\cdot)$ discriminate between the candidates of whitened shocks $\mathcal{A}_{\boldsymbol{\theta}}^{\bullet} = \mathsf{Q}(\boldsymbol{\theta})^{-1}\,\mathcal{A}_{\mathbf{0}}^{\bullet}$ by dependencies beyond the second moment. In the spirit of Hodge-Lehmann estimation (1963), a minimization procedure $\widehat{\boldsymbol{\theta}}^{\bullet} = \text{argmin}_{\boldsymbol{\theta}}\ \mathcal{D}(\mathcal{A}_{\boldsymbol{\theta}}^{\bullet})$ finds the rotation angels $\widehat{\boldsymbol{\theta}}^{\bullet}$ of the least dependent shocks.

ICA can identify shocks and impact matrices up to to scaling, column signs, and ordering only. For example in the case of $K = 2$, the relevant interval of a full rotation $\theta \in (0, 2\pi]$ reduces to a quadrant, e.g. $\theta \in (0, \pi/2]$, since any exceeding rotation just permutes the ordering and reverses signs. If $\theta \in (\pi/2, 2\pi]$, the results of $\mathcal{D}(\mathcal{A}_{\boldsymbol{\theta}}^{\bullet})$ including the minima would be identical to those of the first quadrant. Against this ambiguity, a common practice for the unique identification of B_i is to *(1)* choose the column ordering which maximizes the sum of the absolute diagonal elements and *(2)* then switch signs of those columns whose main diagonal element turns out to be negative. Under $\boldsymbol{\epsilon}_{it} \sim (0, I_K)$, each shock is thereby attributed to the variable on which it has the strongest effect on impact.

ICA-based identification procedures for the individual SVAR are already implemented in **svars**. As we focus on their embedding into a panel framework, we just list them here briefly and refer the reader to the accompanying vignette (Lange *et al.* 2021) for a comprehensive overview and to the Monte Carlo study (Herwartz, Lange, and Maxand 2022) for a performance assessment. The following dependence measures $\mathcal{D}(\cdot)$ and optimization procedures are adopted in **pvars**: The *Cramér-von-Mises* (cvm) *distance* by Genest *et al.* (2007) is used in **svars**' two-step optimization procedure with **copula** (Kojadinovic and Yan 2010) and has been exemplarily applied for individual SVAR by Herwartz (2018). The *distance covariance* (dCov) by Székely *et al.* (2007) is used in the gradient algorithm of **steadyICA** (Risk, James, and Matteson 2015) and has been applied for SVAR by Matteson and Tsay (2017). The *dependence coefficient* (dCoef) by Bakirov *et al.* (2006) is not used in **svars** and **pvars**.[19]

[19]Note that dCoef and dCov are implemented in **energy** by Rizzo and Szekely (2022).

3. Implementation

For each field of VAR application and for the supporting tools, Table 3 displays **pvars**' core functions, the S3-class of their output object and their dependencies within and to other packages. In particular, several classes and their corresponding methods are imported from **svars**. In addition to the familiar methods such as `print()` and `summary()`, **pvars** offers the method `toLatex()` for conveniently formatting **pvars** results into Latex objects, thus minimizing the risk of reporting errors from tedious copy-pasting.

Table 3: Package design of **pvars**.

Function	Class	Branch	Methods	Description	Literature
3.1 Testing the cointegration rank					
`pcoint.JO`	`pcoint`	`coint.JO`	`print,`	Panel Johansen tests	Larsson *et al.* (2001), Choi (2001)
`pcoint.BR`	`pcoint`	`coint.JO`	`summary,`	Panel test with pooled β	Breitung (2005)
`pcoint.SL`	`pcoint`	`coint.SL`	`toLatex`	Panel SL-tests	Örsal, Droge (2014)
`pcoint.CAIN`	`pcoint`	`coint.SL`		Correlation augmented tests	Arsova, Örsal (2021), Hartung (1999)
3.2 Estimating VAR models					
`pvarx.VAR`	`pvarx`	`VAR`[a)]	`irf, print,`	Mean-group estimation	Rebucci (2010), Pesaran, Smith (1995)
`pvarx.VEC`	`pvarx`	`VECM`	`summary`	Pooled cointegrating vectors β	Breitung (2005), Pesaran *et al.* (1999)
3.3 Identifying structure					
`pid.chol`	`pid`	`id.chol`[a)]	`irf,`	Recursive causality	Sims (1980)
`pid.grt`	`pid`	`id.grt`[c)]	`print,`	Long- & short-run restrictions	Breitung *et al.* (2004)
`pid.iv`	`pid`	`id.iv`	`summary,`	Proxy SVAR	Empting et al. (2025)
`pid.dc`	`pid`	`id.dc`[a)]	`toLatex`	ICA by distance covariance	Calhoun *et al.* (2001) and
`pid.cvm`	`pid`	`id.cvm`[a)]		ICA by Cramér-von-Mises dist.	Herwartz, Wang (2024)
3.4 Supporting tools					
`speci.factors`	`speci`	—	`print`	Criteria for number of factors	Bai, Ng (2002; 2004), Onatski (2010)
`speci.VAR`	`speci`	—	`print`	Criteria for lags p and periods $\boldsymbol{\tau}$	Bai, Perron (1998; 2003), Yang (2002)
`irf.pvarx`	`svarirf`[b)]	`irf.varx`	`plot, print`	Mean-Group IRF [S3-METHOD]	Sims (1980), Gambacorta et al. (2014)
`PP.system`	`svarirf`[b)]	—	`plot, print`	Persistence profiles	Lee, Pesaran (1993) and
`PP.variable`	`svarirf`[b)]	—	`plot, print`	(Structural) persistence profiles	Pesaran, Shin (1996)
`sboot.mg`	`sboot`[b)]	—	`print, summary,`	Mean-group inference	Pesaran, Smith (1995)
`sboot.pmb`	`sboot`[b)]	`sboot.mb`	`plot, toLatex`	Panel-block bootstrap	Empting *et al.* (2025)

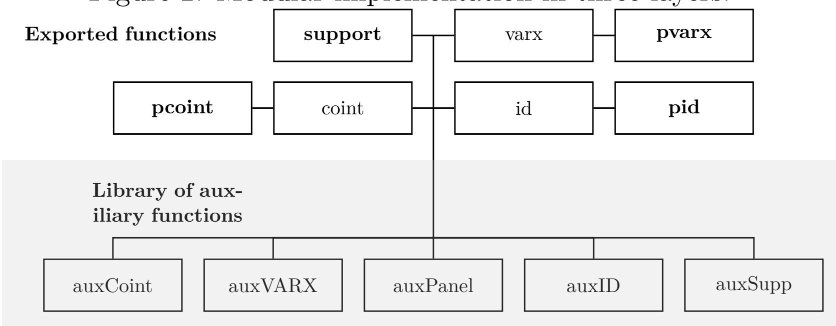
a) Consider the R-packages **vars** and **svars** for these functions. Like `id.dc`, **pvars**' `pid.dc` uses **steadyICA** by Risk *et al.* (2015).
b) This class and its methods are imported from the R-package **svars** by Lange *et al.* (2021).
c) Its scoring algorithm is part of the `SVEC` function from the R-package **vars** by Pfaff (2008b).

Figure 2 illustrates **pvars**' modular implementation, which leads to the three layers of dependency between the library of auxiliary functions, the functions for individual econometric procedures, and their panel extensions. In this tree-like structure, the `aux`-functions are the hidden "roots" and the individual functions are "leafy branches". Some branches of the individual `id.*()` functions are a "graftage" from **vars** and **svars**. The individual `coint.*()` functions are more flexible than existing implementations of cointegration rank tests and extend the set of available specification options. Finally attached to the branches, the "fruits" of this package are the panel functions for each field of VAR application. These are printed in bold to emphasize their novelty and **pvars**' main contribution.

The idea of modular programming is to break monolithic and repetitive code down into functional sub-entities, which achieves easier maintenance, better testability, and reusability. Especially the multivariate panel procedures benefit from this kind of implementation because their functions reduce to simple re-arrangements of auxiliary functions. These can be repetitively applied over subsets of the $K \times T \times N$ data array. Whenever sensible, the R-functions are vectorized in order to enable more flexible input arguments and faster matrix computation of the multivariate VAR models. In consequence, the in- and outputs are mainly matrix objects, which thereby serve as an interface between the auxiliary modules.[20] With

[20]Note that **pvars** does not export auxiliary functions to the global environment. If the user wishes to

Figure 2: Modular implementation in three layers.



automatic unit tests by **testthat** (Wickham 2011), `pcoint.*()` and subordinated modules are checked against reproduced examples of the literature. Functions outside this hierarchy are checked against consistency with the other **vars** packages, identities derived from econometric theory, and simulation results from Empting *et al.* (2025).

Argument structure. In order to comply with the modular functions and their repetitive call over data subsets, several conventions have been established for the input arguments of the auxiliary, individual, and panel functions. The arguments whose names are marked with the prefix `L.*` are an R-object of class `list` with N elements of repeated structure. Most prominently, `L.data` is the panel data in *list-format* and, as a repetitive `list`, contains N `data.frame`s of dimension $T_i \times K$. In these multivariate time series, the variables must have an identical order $k = 1, \dots, K$ for each individual $i = 1, \dots, N$. All successions are binding for any other `L.*` object within the environment of a given function. Hence even in the global environment, the user should stick to a universal succession for all `L.*` objects in order to avoid confusion.

Further conventions for prefixes are `n.*` and `dim_*`, which denote integers and refer to a quantity resp. the dimension of a matrix or array. Straightforwardly, `dim_K`, `dim_T`, and `dim_N` signify the data dimensions K, T, and N for instance. Arguments with prefix `t_*` define $\boldsymbol{\tau}$ for the period-specific deterministic regressors. As a `list`, they collect optional vectors of integers for trend breaks (named `t_break`), shift dummies (`t_shift`), and impulse dummies (`t_impulse`) as well as a single integer `n.season`, which indicates the seasonal frequency for the centered dummies. Each integer in those vectors specifies a period τ at which a break, shift, or impulse occurs within the interval $1, \dots, T$ of the complete panel data set (i.e. including the presamples). The following sections describe how the mandatory and optional input arguments are employed in the **pvars** functions for cointegration tests (Section 3.1), estimation (Section 3.2), and structural identification (Section 3.3).

3.1. Testing the cointegration rank

The panel cointegration tests are implemented in accordance with the three dimensions of panel test construction as presented in Table 1. *(i)* Each `pcoint.*()` function always assesses all hypotheses $r_{H0} = 0, \dots, K-1$ and uses all combination approaches available for the respective underlying individual procedure. *(ii)* The data generating process by contrast, which covers also the panel test generation, must be defined via the arguments of `pcoint.*()`.

construct own functions, she needs to call an auxiliary function `pvars:::aux_*()` via a triple colon.

These arguments are kept consistent across the functions or refer to the underlying procedure. *(iii)* The panel functions `pcoint.*()` usually loop over their internal function `cointf()` as the underlying procedure. `cointf()` arranges the auxiliary functions in the same way as its respective branch `coint.*()`, but has been modified for the panel test. Hence, if an individual sample does not conform with the panel result, the user may consult the individual counterpart `coint.*()` for a closer inspection of the peculiar individual entity. For this purpose, she can also consider the persistence profiles `PP.*()`. Within the scope of PANIC, the test functions `coint.*()` for the single VAR system can assess cointegration between the common factors $\boldsymbol{F}_t$ of Model (32), too.

The branch of Johansen. The individual Johansen procedure is available as the function

```
coint.JO(y, dim_p, x, dim_q, type, t_D1=NULL, t_D2=NULL)
```

By default, the R function performs both LR-test variants on the multivariate time series `y` with lag order `dim_p`. For the *trace test* as in Eq. (15) and exclusively for the individual *maximum eigenvalue test* with simple specifications, p-values are approximated by the gamma distribution. Optionally, weakly exogenous variables `x` with lag order `dim_q` are incorporated. The character argument `type` defines the conventional deterministic term according to the labels for the *innovative model* in Table 2. Optional breaks of the deterministic term in the periods `t_D1` are treated in accordance with JMN (2000) or KN (2019). The panel-extensions of the Johansen procedure are implemented by

```
pcoint.JO(L.data, lags, type, t_D1=NULL, t_D2=NULL, n.factors=FALSE)
pcoint.BR(L.data, lags, type, t_D1=NULL, t_D2=NULL, n.iterations=FALSE)
```

The input argument `L.data` requires a data panel in *list-format* as explained above for the argument structure of **pvars** functions in general. `lags` defines the lag order p_i of the individual VAR models in levels and is either a vector of N integers or a single integer for a common lag order $p = p_i$. For assigning the heterogeneous lag orders to each individual, the integers p_i must have the same succession $i = 1, \ldots, N$ as `L.data`. The optional argument `n.factors` can activate the PANIC-defactoring, where the common components with the chosen number of factors $\boldsymbol{F}_t$ are subtracted. Then, the idiosyncratic cointegration rank tests are fixed to the distribution moments for *SL_trend* irrespective of the specification of `type`. Specifically `pcoint.BR()` uses the LM-test from Eq. (17) instead, where `n.iterations` defines the number of repetition in the two-step estimation of β_K. Note that any deterministic term is equipped with a heterogeneous effect β_{0i} in order comply with the Brownian bridges in the individual $Z_{r\perp}$. The corresponding option `idx_pool` in `pvarx.VEC()` is thus disabled in `pcoint.BR()` and just cancels out.

The branch of Saikkonen and Lütkepohl. The individual SL-procedure is available via

```
coint.SL(y, dim_p, type_SL, t_D=NULL)
```

The arguments therein are the same as in the individual Johansen procedure except for `type_SL`, which requires a label for the *additive model* from Table 2. Optional breaks of the deterministic trend in the periods `t_D` are treated in accordance with TSL (2008). The panel-extensions of the SL procedure are implemented by

```
pcoint.SL(L.data,   lags, type="SL_trend", t_D=NULL, n.factors=FALSE)
pcoint.CAIN(L.data, lags, type="SL_trend", t_D=NULL)
```

Again, the specifications of the input arguments are identical to those of the Johansen procedure `pcoint.JO` except for the `type` of the additive model. The default is `"SL_trend"` as Örsal and Droge (2014) propose for their panel SL-test, Arsova and Örsal (2021) for the CAIN-test, and Örsal and Arsova (2017; 2018) enforce after the defactoring in PANIC.

3.2. Estimating VAR models

The position of the individual VAR Model (1) as the basic econometric unit of Section 2 is reflected in the R-implementation by its class `varx`. Accordingly, *(i)* panel VAR estimates of class `pvarx` contain a repeated list `L.varx` of individual VAR, *(ii)* SVAR estimates `id` inherit from the parent class `varx`, and *(iii)* panel SVAR estimates `pid` define their `L.varx` as a list of individual `id` objects. Hence, each individual VAR embedded in the panel estimates can be separately inspected by the methods for `varx` objects.

In **pvars**, `varx` also serves as an intermediary class to ensure compatibility to the other packages of the **vars**-ecosystem. Either estimated via **pvars**' `VECM()` or coerced from **vars**' `varest` and `vec2var` objects via `as.varx()`, the `varx` objects can then enter the same functions since the class obeys a unifying construction plan for different VAR model types. If the user wishes to employ **pvars** function to VAR objects of other classes, she may simply specify accordant `as.varx()`-methods instead of altering the original **pvars** function. A list of class `varx` contains the coefficient matrix `$A` for the *full-system* and *level-representation* VAR (1), its residual covariance matrix `$SIGMA`, and a structural impact matrix `$B`. In reduced-form VAR objects, the latter is just a placeholder `$B` $= I_K$ such that `irf()` generates *forecast-error impulse responses* (Lütkepohl 2005, p. 52). In SVAR object of class `id`, `$B` is the result of an identification procedure. If a cointegration rank-restriction or conditional estimation is employed, the estimates and specifications of these VAR representations are stored in the slots `$VECM`, `$PARTIAL`, and `$MARGINAL` and then transformed to the top-level `$A`.

Panel of VAR models. Extended from the branch of individual VAR resp. VECM, the estimators for the VAR models of heterogeneous panels

```
pvarx.VAR(L.data, lags, type, t_D=NULL, D=NULL)
pvarx.VEC(L.data, lags, dim_r, type, t_D1=NULL, t_D2=NULL,
          D1=NULL, D2=NULL, idx_pool=FALSE, n.iterations=FALSE)
```

use data panels in list-format `L.data` to estimate a list of individual VAR models `$L.varx`. The specifications of the VAR processes are the lag orders `lags`, the `type` of the conventional deterministic term, and optional deterministic regressors activated in the periods `t_D` resp. `t_D1` and `t_D2`. While these arguments of `pvarx.VEC()` comply with the labels in Table 2 and with $\boldsymbol{d}_{1t}$ and $\boldsymbol{d}_{2t}$ in Model (2), `pvarx.VAR()` is in accord with **vars**' well-known `VAR()` function and accepts a `"const"`, a linear `"trend"`, `"both"`, or `"none"` in $\boldsymbol{d}_t$ of each individual Model (1). Customized regressors can be included as a common single data matrix or a `list` of individual data matrices (including the presample) via `D` in VAR models and via restricted `D1` and unrestricted `D2` in VECM. Unlike `t_D`, `t_D1`, and `t_D2`, these arguments do not add accompanying lagged regressors to $\boldsymbol{d}_{2t}$ automatically.

In the next step, the individual coefficients are combined by cross-sectional averages. This provides *(i)* Pesaran and Smith's (1995) mean-group (MG) estimation as suggested by Canova and Ciccarelli (2013) and assessed by Rebucci (2010) for VAR or *(ii)* Pesaran et al.'s (1999) pooled mean-group (PMG) estimation if Breitung's (2005) two-step estimator has been selected. The latter is activated if some variables `idx_pool` are restricted to have homogeneous coefficients in the `dim_r` cointegrating vectors `$beta`. For this, the switching algorithm can be used with further `n.iterations`. If all elements of `idx_pool` are in $[0, \ldots, r]$, the coefficients up to the uniform upper block I_r are throughout heterogeneous and estimated with the individual estimator by Ahn and Reinsel (1990). All resulting panel estimates such as `$A` or `$beta` are stored in top-level slots of the `pvarx` objects.

3.3. Identifying structure

Equivalently to **svars**' `id.*()` procedures for individual VAR objects, **pvars**' `pid.*()` functions are applied to `pvarx` objects containing the panel estimates of the reduced-form VAR. Accordingly, the arguments to control the underlying identification procedures are identical to those of the **svars** package. Lange *et al.* (2021) describe the implementation and application of the identification procedures in detail.

Imposed. Theoretical considerations like recursive causality may imply restrictions which can be imposed uniformly on B_i of each individual VAR model. Some ensuing functions of Section 3.4 accept a simple list of individual **svars** objects, too, which will produce identical results under these restrictions. However, to enable the full functionality of the **pvars** package, the following functions are added as panel equivalents into the `pid.*()` canon.

```
pid.chol(x, order_k=NULL)
pid.grt(x, LR=NULL, SR=NULL, start=NULL, max.iter=100, conv.crit=1e-07, maxls=1.0)
```

The function `pid.chol()` is the direct extension of **svars**' `id.chol()`, where the optional argument `order_k` allows for specifying alternative causal orderings in the Cholesky decomposition. The panel function `pid.grt()` and its individual counterpart perform the ML estimation (22) for SVECM under short and long-run restrictions on the $K \times K$ matrices `SR` resp. `LR`. Using the scoring algorithm from **vars**' `SVEC()`, they have identical arguments to tune the optimization procedure as Pfaff (2008b, Sec. 3.2) describes in detail.[21] If the input object `x` contains pooled cointegrating coefficients $\beta_k = \beta_{ki}$, those are used to calculate the orthogonal complement $\beta_\perp$ in the structural identification of Eq. (21).

Data-driven. Residual structure in $\boldsymbol{u}_{it}$ such as non-Gaussianity allows data-driven identification. For this purpose, **pvars** offers the panel applications of ICA. Via the argument `combine`, the user can select a strictly individual identification of B_i as in **svars** (using `"indiv"`), a common rotation of pooled shocks by Herwartz and Wang (2024) (`"pool"`), or `n.factors` common shocks across the individuals by Calhoun *et al.* (2001) (`"group"`).

```
pid.dc(x,  combine, n.factors=NULL, n.iterations=100, PIT=FALSE)
pid.cvm(x, combine, n.factors=NULL, dd=NULL, itermax=500, steptol=100, iter2=75)
```

The panel identification functions pass the combined samples of whitened shocks to the ICA procedures. Like `id.dc()` in **svars**, `pid.dc()` uses the gradient algorithm from **steadyICA** to

[21] This function integrates **vars**' `SVEC()` into the **pvars** system on the panel level. `SVEC()` cannot be applied to objects of class `cajo-test`, i.e. **urca**'s VECM object with restricted α or β, although these restrictions contribute to the identification of structural shocks. As an individual counterpart, `id.grt()` is applied to `varx` objects, allowing for complex deterministic terms, the MB bootstrap, and `svarirf` methods.

minimize the distance covariance (dCov) with respect to the rotation angles $\boldsymbol{\theta}$. Their joint option `PIT` activates *probability integral transformation*, which transforms the marginal densities of the structural shocks before evaluating dCov. The maximum number of iterations is `n.iterations=100` by default. The panel function `pid.cvm()` uses the procedure from **svars**' `id.cvm()` to minimize the CvM distance. For both CvM functions, **copula**'s (Kojadinovic and Yan 2010) `indepTestSim()` simulates the distribution of test statistics under independence, which is either provided via `dd` or called internally if `dd=NULL`. The external provision of `dd` saves computation time if simulated once and then used for multiple calls of `id.cvm()` or `pid.cvm()` on `x` with identical sample dimensions. The remaining arguments control the two-step optimization procedure for $\widehat{\boldsymbol{\theta}}$. In a first step, the differential evolution algorithm from **DEoptim** (Mullen, Ardia, Gil, Windover, and Cline 2011) determines preliminary angles $\boldsymbol{\theta}^*$ within `itermax` iterations under a tolerance of `steptol`. The second step further optimizes the test statistic in `iter2` iterations locally around $\boldsymbol{\theta}^*$.

All `pid.*()` functions extend their `pvarx` input to the class `pid`. Therein, the structural impact matrices $\widehat{\mathsf{B}}_{\mathsf{i}}$ have been assigned to each SVAR object in `$L.varx` and their mean-group estimates to the top-level slot `$B` next to `$A`. By default, the ICA-based panel functions order the columns of all $\widehat{\mathsf{B}}_{\mathsf{i}}$ uniformly pursuant to the aforementioned convention of the SVAR literature. Consequently, the main diagonal of the mean-group `$B` holds the maximum absolute estimates as positive coefficients $\widehat{\mathsf{b}}_{kk}$. Mean-group statistics of $\widehat{\mathsf{B}}_i$ as presented in Bernoth and Herwartz (2021) and Herwartz and Wang (2024) can be viewed via `pid`'s `summary()` method and exported via `toLatex()`.

3.4. Supporting tools

Dynamic analysis. In the **vars**-ecosystem, several tools of dynamic analysis are already available such as *impulse response functions* (IRF), *forecast error variance decomposition* (FEVD) and *historical decomposition* (HD). **pvars** extends this list by *mean-group IRF* (Gambacorta *et al.* 2014, p. 627). Given a VAR object `x` of class `pvarx` or `pid`, the method

```
irf(x, n.ahead=20, normf=NULL, w=NULL)
```

calculates the cross-sectional average of individual responses for each period after the initial impulse. The function optionally provided in `normf` normalizes the shock size of these impulses. Vector `w` with names, N logical, or N numeric elements allows to select a subset of the N individuals resp. to apply real-valued weights in the mean-group estimation.

Persistence profiles (PP) by Pesaran and Shin (1996) are particularly useful for panel cointegration analysis. Given an individual VECM, they map the speed of convergence to the long-run equilibrium after an impulse shock and thus allow to counter-check the individual error correction behavior under a common cointegration rank or pooled cointegration matrix. While a reversion to the r long-run equilibria is the defining property of cointegration, explosive roots can emerge from ignored breaks in the deterministic term on the individual level and contaminate the panel results. In **pvars**, the functions

```
PP.system(x,   n.ahead=20)
PP.variable(x, n.ahead=20, shock=NULL)
```

calculate PP initiated by *system-wide shocks* resp. by *variable-specific shocks* based on the Cholesky decomposition of $\widehat{\Sigma}_{u,i}$. *Structural shocks* are derived from $\widehat{\mathsf{B}}_i$ if `x` is a structural

VECM object of class `id` instead of the reduced-form `varx`. The matrix `shock` controls via its K rows which shocks are selected and combined. Both tools, `PP.*()` and `irf()`, return `svarirf` objects, for which **svars** provides the `plot` method to visualize the responses over a horizon of up to `n.ahead` time periods.

Bootstrapping. Bootstrap procedures are a standard tool for VAR modeling to reconstruct the sampling distribution and perform inference. For estimating standard errors of point estimates and confidence bands of structural impulse-responses, **pvars** provides recursive-design bootstrap procedures. In particular, the following functions implement the moving-block bootstrap for individual VAR models (Brüggemann, Jentsch, and Trenkler 2016) and the panel-block bootstrap (Empting *et al.* 2025) respectively.

```
sboot.mb(x, b.length=1, n.ahead=20, n.boot=500, n.cores=1,
          deltas=cumprod((100:0)/100), normf=NULL)
sboot.pmb(x, b.dim=c(1, 1), n.ahead=20, n.boot=500, n.cores=1,
          deltas=cumprod((100:0)/100), normf=NULL, w=NULL)
```

Given an estimated (panel) SVAR object `x` of class `id` reps. `pid`, the bootstrap functions iterate `n.boot` times re-estimating the VAR models, their structural matrices B_i, and impulse responses over a horizon of up to `n.ahead` periods. The arguments from the SVAR object itself (i.e. model specifications, estimation and identification methods, optional restrictions on $\alpha_i\beta'$ like rank and weak exogeneity) are passed to and fixed over the bootstrap iterations. In order to speed up their computation by parallel processing, more than one CPU core can be assigned to the procedure via `n.cores`. The resulting **svars** object of class `sboot` allows to plot IRF confidence bands via **svars**' `plot()` method. Confidence intervals for parameters A or B can be viewed via `summary()` and exported via `toLatex()`.

If input `x` contains bias-correction terms `PSI_bc` resp. `L.PSI_bc`, both functions perform a bias-corrected bootstrap. For example, objects from a first bootstrap contain such terms and thus enable the bootstrap-after-bootstrap of individual (Kilian 1998) or panel VAR models (Empting *et al.* 2025), where the weights `deltas` control a successive stationarity correction. `plot()` then displays small-sample corrected IRF and their confidence bands.

The argument `b.dim` defines the dimensions $\left(b_{(t)}, b_{(i)}\right)$ of the panel blocks for temporal and cross-sectional resampling. The default `c(1, 1)` specifies an *iid.* resampling in both dimensions, `c(1, FALSE)` a temporal resampling, and `c(FALSE, 1)` a cross-sectional resampling. Choosing integers $b_{(t)},\ b_{(i)} > 1$ assembles blocks of consecutive residuals to capture residual structure like ARCH or cross-sectional correlation. Moreover, `sboot.mb()` complements **svars**' `mb.boot()` (Lange *et al.* 2021, Sec. 3.6 and 4.2) and accepts individual SVAR objects identified by `id.grt()` (Breitung *et al.* 2004) or `id.iv()` (Jentsch and Lunsford 2022). Here, the default `b.length=1` implies the residual *iid.* bootstrap as implemented in **vars**' `irf()`, while a single integer $b_{(t)} > 1$ defines the length of temporal blocks for a moving-block bootstrap.

4. Empirical illustrations

Several empirical illustrations accompany the package to demonstrate its application. In the `help()` for functions, the examples provide chunks of R-code for directly copy-pasting unit-tested reproductions. In this section, we focus on the workflow of **pvars** and therefore guide the user to first organize the $K \times T \times N$ data array, then perform a panel cointegration analysis, and finally export the results to Latex. Specifically, the reproduced example of Örsal and Arsova (2017) in Section 4.1 illustrates how to perform the *Panel Analysis of Nonstationarity in Idiosyncratic and Common components* (PANIC), and the reproduced example of Arsova and Örsal (2021) in Section 4.2 how to specify deterministic terms. The R-code for both illustrations is assembled in the file *pvars_reproductions.R* in **pvars**' *examples* folder. A comprehensive illustration can be found in Empting and Herwartz (2025), who go through the pretests and VAR-applications 3.1 to 3.4 successively.

Data format. The exemplary data sets of the **pvars** package are panels in the popular *long-format*, where all N multivariate time series have been transposed into $T \times K$ matrices and stacked into an $(N \cdot T) \times (2 + K)$ `data.frame` object. The two additional columns `id_i` and `id_t` contain `factor` elements, which serve as identifiers for individual i and time period t. Accordingly, each observation $\boldsymbol{y}_{it}$ is stored in a single row. The `factor` variables preserve the predefined `levels` order $1, \ldots, N$ within the complete long-format data panel or its subsets.[22] In the following, we consider the data set `MERM` and firstly extract the names of the variables $k = 1, \ldots, K$ and countries $i = 1, \ldots, N$.

```
R> library("pvars")
R> data("MERM")
R> names_k = colnames(MERM)[-(1:2)]
R> names_i = levels(MERM$id_i)
R> head(MERM, n=3)

    id_i     id_t          s          m          y         p
1 Brazil 1995_Jan -0.1660546 -3.094546 0.07401953 0.3357538
2 Brazil 1995_Feb -0.1731636 -3.054644 0.07127137 0.3422039
3 Brazil 1995_Mar -0.1176580 -3.055017 0.06986985 0.3539417
```

Naturally, **pvars**' modular implementation works well with panel data in *list-format*, where each of the N listed elements is an individual matrix of $T \times K$ time series. This can be constructed by either writing separate time series into the `list` object or transforming the long-format[23] data panel via `sapply()`.

```
R> L.data = sapply(names_i, FUN=function(i)
+      ts(MERM[MERM$id_i==i, names_k], start=c(1995, 1), frequency=12),
+      simplify=FALSE)
```

Here, the individual matrices in `L.data` have been defined as time series objects `ts` with `frequency=12` for monthly observations starting in January 1995. Although the functions in **pvars** do not require this, the `ts`-definition simplifies the workflow when using further packages like **ggplot** (Wickham 2016). The panel functions yet resort to the names of the listed time series as labels for individual results. `sapply()` assigns this definition directly, but `names()` can enforce this subsequently, too, as an alternative transformation requires:

[22] Thereby, we also preempt the data management of R versions older than release 4.0.0, which would coerce `character` vectors into `factor` columns automatically and sort their `levels` alphabetically. For instance, this could lead to mismatches when switching between label standards as in the case of Switzerland with the ISO-3166 abbreviation "CHE".

[23] *Wide-format* panels may be transformed into long-format first. The function `melt()` of the **reshape2** package (Wickham 2007) can perform this task. Consider his vignette for a more detailed explanation of these two data formats and for an additional, third way to transform data into list-format.

```
R> L.data = lapply(names_i, FUN=function(i) MERM[MERM$id_i==i, names_k])
R> names(L.data) = names_i
```

Either way, the data set is now readily prepared for the econometric analysis with **pvars**.

4.1. The monetary exchange rate model: Conduct a PANIC

Örsal and Arsova (2017) illustrate the PANIC analysis of the *monetary exchange rate model* (MERM), according to which the nominal exchange rate s_{it} between two countries forms a long-run relationship with their relative level of money supply and their relative level of output. As Dąbrowski, Papież, and Śmiech (2014) propose, Örsal and Arsova adopt the log-linear model

$$s_{it} = \mu_{0i} + \mu_{1i}t + \beta_{i1}\left(m_{it} - m_t^*\right) + \beta_{i2}\left(y_{it} - y_t^*\right) + \beta_{i3}\left[\left(p_{it} - p_{it}^T\right) - \left(p_t^* - p_t^{T*}\right)\right] + u_{it}, \quad (36)$$

where the variables for the USA as the preselected reference country are marked with an asterisk. The natural logarithm of the dollar exchange rate for a country i is denoted by s_{it}, the logarithmized nominal money supply by m_{it}, and the logarithmized industrial production index by y_{it}. Moreover, they have included the natural logarithm of consumer price index p_{it} and producer price index p_{it}^T for country i and likewise for the USA.

Data. As `head(MERM)` has shown for the illustrative transformation of the data format, the data set `MERM` contains $K = 4$ variables. These already summarize each log-ratio of Model (36) and thus enter the additive Model (32) directly as the observed time series $\boldsymbol{y}_{it}$. The monthly observations cover the period 1995/01 – 2007/12 ($T = 156$) for $N = 19$ countries and are listed in `L.data` after transforming their data format. The names of `L.data`'s 19 elements provide the labels for the "individuals" in Table 4.

Approximate factor model. The first step of PANIC is to estimate the approximate factor model in Eq. (32), which splits $\boldsymbol{y}_{it}$ into common and idiosyncratic components $\Lambda_i' \boldsymbol{F}_t$ resp. $\boldsymbol{y}_{it}^{id}$. The factor model considers the data panel just as a collection of time series without individual structure. Both dimensions $T \times NK$ of the data are assumed to be large and both components of the model may involve mixes of $I(0)$ and $I(1)$ series. First-differencing these data panels beforehand is a valid choice to estimate the factor model by PCA (Bai and Ng 2004) and to determine its number of common factors $\boldsymbol{F}_t$ by the eigenvalues.[24] The information criteria in `[[1]]` however ignore the individual structure of our panel and thus tend to pick up the domestic dependencies between the $K = 4$ variables within countries. Since we are interested in the factors describing cross-sectional dependence only, we prefer the specification procedure by Onatski (2010). His *edge distribution* `ED`[25] is more robust against domestic dependencies because it looks for a characteristic kink in the ordered eigenvalues. `ED` also works with the original `L.data` in levels irrespective of the components' order of integration. In order to find the optimal number of factors within the discrete interval $\{0, \ldots, 20\}$, we enter the R function

```
R> speci.factors(L.data, k_max=20, n.iterations=4)
```

[24] See Corona, Poncela, and Ruiz (2017) for an overview and Monte Carlo results. An exception is the set of $IPC(k)$ criteria by Bai (2004), who seek to distinguish non-stationary factors from stationary idiosyncratic series. Accordingly, `speci.factors()` suppresses their result if `differenced=TRUE` is selected.

[25] **phtt** by Bada and Liebl (2014) with `OptDim(obj, criteria="ED")` has been removed from CRAN.

```
### Optimal number of common factors ###
[[1]]
   PC IC IPC
p1 20 20   7
p2 20 20   7
p3 20 20   5

[[2]]
ER GR ED
1  2  8
```

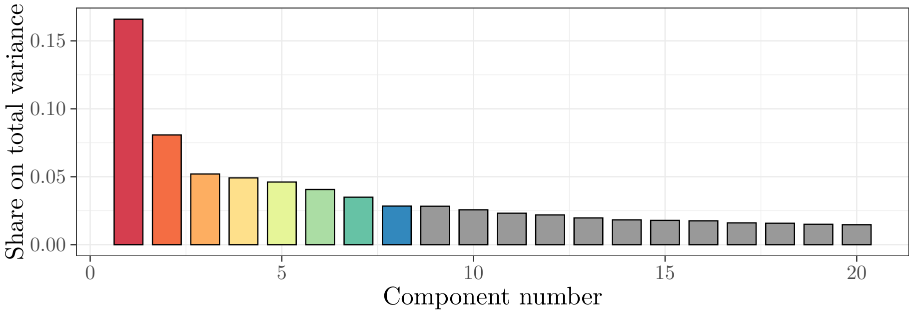
A numerical result for `ED` indicates that the default of `n.iterations=4` allows Onatski's (2010) edge distribution to converge. In case of an `NA`, the user needs to increase the number of iterations, but small numbers are often sufficient.

The result of eight common factors can be visualized and checked in a scree plot. In accordance with PANIC of the `pcoint` functions, we may consider the factor model estimated with the first-differenced and standardized data now:

```
R> R.fac0 = speci.factors(L.data, k_max=20, n.iterations=4,
+    differenced=TRUE, centered=TRUE, scaled=TRUE, n.factors=8)

R> library("ggplot2")
R> pal = c("#999999", RColorBrewer::brewer.pal(n=8, name="Spectral"))
R> lvl = levels(R.fac0$eigenvals$scree)
R> ggplot(R.fac0$eigenvals[1:20, ]) +
+    geom_col(aes(x=n, y=share, fill=scree), color="black", width=0.75) +
+    scale_fill_manual(values=pal, breaks=lvl, guide="none") +
+    labs(x="Component number", y="Share on total variance", title=NULL) +
+    theme_bw()
```

Figure 3: Scree plot.



In the resulting plot of Figure 3, the first eight eigenvalues for the relevant components are colored.[26] They still account for almost 50% of the total variation in the first-differenced and centered $K \cdot N$ time series, but the PCA of the original data attributes over 98% to the first

[26]The vector graphics in this Latex document have been generated by the **tikzDevice** package (Sharpsteen and Bracken 2020), which prints R plots as a Ti*k*Z environment into ".tex" files.

principal component alone. It appears that this all-dominating component is a linear trend in the time series, which is removed after first-differencing and centering. Now, the eighth and ninth eigenvalue are inconsiderable, while the first two exhibit more pronounced kinks. Indeed, the *eigenvalue ratios* `ER` and *growth rates* `GR` by Ahn and Horenstein (2013) as well as `ED` by Onatski (2010) hint at one resp. two common factors.

```
R> R.fac0$selection[[2]]

ER GR ED
1  1  2
```

For the reproduction, we yet proceed with the decision by Örsal and Arsova (2017) and adhere to the conservative choice of eight common factors in order to ensure cross-sectional independence for the panel test.

Panel cointegration tests. The approximate factor model with `n.factors=8` yields a non-stationary,[27] idiosyncratic remainder $\hat{\boldsymbol{y}}_{it}^{id}$, to which Örsal and Arsova (2017) apply the panel SL-tests. To reproduce their results, we specify **pvars**' function `pcoint.SL()` as follows. Due to the defactoring and in accordance with the econometric Model (36), the $N = 19$ individual testing procedures therein must take care of deterministic trends. The lag order p_i of each idiosyncratic VAR model is chosen from the discrete interval $\{1, \ldots, 4\}$ by the minimized *Akaike information criterion*. Here, we enter the results directly, but **vars** functions may determine them from the data matrices in `R.fac0$L.idio`, too.

```
R> R.lags = c(2, 2, 2, 2, 1, 2, 2, 4, 2, 3, 2, 2, 2, 2, 2, 1, 1, 2, 2)
R> R.pcsl = pcoint.SL(L.data, lags=R.lags, type="SL_trend", n.factors=8)
R> toLatex(R.pcsl)
```

The method `toLatex()` prints the `pcoint` results as a `tabular` for Latex, which has been encapsulated in Latex' float environment in order to create Table 4. The table reports the individual and panel results for each hypothesis $r_{H0} = 0, \ldots, K-1$, which refer to Table 5 in Örsal and Arsova (2017, p. 68). All four combination approaches under the independence assumption of the idiosyncratic VAR processes have been used. Comparing the p-values to a significance level of $\alpha = 5\%$, all sequential panel test procedures reject the hypotheses up to $r_{H0} = 1$ and thus confirm the presence of a single cointegration relation in $\boldsymbol{y}_{it}^{id}$.

Cointegration rank of the factors. Having determined the idiosyncratic cointegration rank, the PANIC turns then to the cointegration within the eight common factors $\boldsymbol{F}_t$. The `$CSD`-slot of the `pcoint` object contains the estimates for the cross-sectional dependence and is identical for the PANIC analysis of any `pcoint` function. Therein, the eigenvalues of the PCA are stored in the vector `eigenvals` and the cumulated common factors in the matrix `Ft` of dimension `dim_T` × `n.factors`. These multivariate time series shall be plotted firstly to get an overview as in Figur 4 of Örsal and Arsova (2017, p. 71). For this, we define the factor matrix `Ft` as a `ts` object with the same specifications as the observed time series and

[27]Note that Bronder's (2016) R-package **PANICr** for single-equation PANIC methods, i.e. unit root and residual-based cointegration tests, has been removed from CRAN lately. The methods rely on the same estimator for the common factors, that is a principal component analysis on the first-differenced variables, where the deterministic component has been removed. Consequently, the auxiliary function `aux_ComFact()` can also be used for constructing own functions for these single-equation methods.

Table 4: Panel cointegration rank tests for MERM.

Individual		statistics				p-values			
	lags	$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$	$r_{H0}=3$	$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$	$r_{H0}=3$
Brazil	2	41.606	13.963	8.660	3.618	0.115	0.832	0.473	0.256
Canada	2	43.889	19.816	6.253	0.413	0.070	0.406	0.747	0.940
Colombia	2	25.210	13.042	4.744	1.914	0.875	0.881	0.891	0.558
Czech Republic	2	29.238	17.663	7.712	0.583	0.689	0.568	0.580	0.901
Denmark	1	37.925	18.804	7.758	2.450	0.230	0.480	0.575	0.442
Hungary	2	32.372	16.940	8.401	0.863	0.510	0.624	0.502	0.829
India	2	24.846	14.801	6.280	1.889	0.887	0.780	0.744	0.564
Indonesia	4	26.911	12.640	4.211	1.687	0.806	0.899	0.928	0.612
Israel	2	36.282	21.678	6.217	0.561	0.301	0.285	0.751	0.906
Japan	3	28.154	15.395	6.966	1.975	0.746	0.739	0.667	0.544
Korea	2	57.469	13.816	7.835	2.817	0.002	0.840	0.566	0.374
Mexico	2	28.996	20.596	6.638	0.682	0.702	0.352	0.704	0.876
Norway	2	43.766	19.633	6.715	1.367	0.072	0.419	0.696	0.694
Poland	2	60.457	28.641	6.618	0.557	0.001	0.048	0.707	0.907
South Africa	2	21.298	10.053	6.868	4.730	0.968	0.976	0.678	0.147
Sweden	1	32.127	7.147	3.386	0.777	0.524	0.998	0.969	0.851
Switzerland	1	28.419	13.682	3.413	1.699	0.733	0.848	0.968	0.609
Turkey	2	48.692	27.137	14.041	0.325	0.021	0.075	0.094	0.958
United Kingdom	2	50.253	24.440	4.287	3.031	0.014	0.152	0.923	0.339
Panel		statistics				p-values			
		$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$	$r_{H0}=3$	$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$	$r_{H0}=3$
LRbar		2.305	-1.346	-2.635	-2.095	0.011	0.911	0.996	0.982
Choi P		70.515	29.377	17.050	20.338	0.001	0.841	0.999	0.992
Choi Pm		3.730	-0.989	-2.403	-2.026	0.000	0.839	0.992	0.979
Choi Z		-1.914	1.528	2.639	2.124	0.028	0.937	0.996	0.983

use the related plotting method via `autoplot()`. The package **ggfortify** (Tang, Horikoshi, and Wenxuan 2016) provides a comprehensive set of unified methods for **ggplot2** graphics.

```
R> library("ggfortify")
R> Ft = ts(R.pcsl$CSD$Ft, start=c(1995, 1), frequency=12)
R> autoplot(Ft, facets=FALSE, size=1.5) + theme_bw() +
+      scale_color_brewer(palette="Spectral") +
+      labs(x=NULL, y=NULL, color="Factor", title=NULL)
```

Figure 4: Estimated common factors.

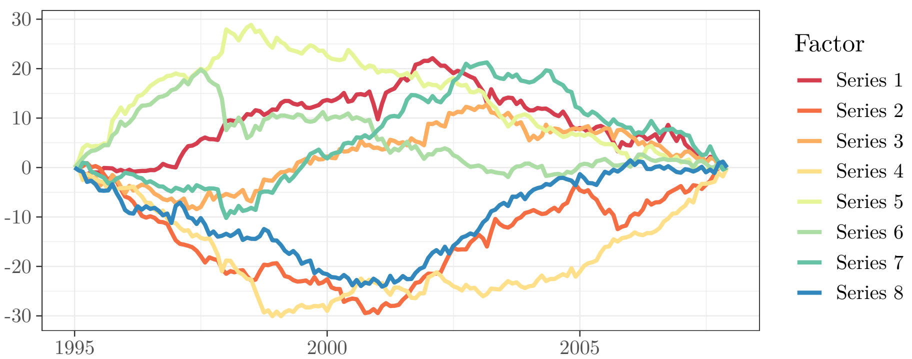


Figure 4 depicts the `autoplot()` result for the common factors, which are ordered according to their decreasing eigenvalue of the PCA. Using `palette="Spectral"` for factors' color assignment, the transition from "warm" to "cold" colors reflects decreasing power of the factors to explain the common variation within the observed time series. In comparison to the figure from the original study, the factors $\hat{\boldsymbol{F}}_t$ may have switched signs and be multiplied by a scalar such that only their dynamics are informative. However, the PANIC analysis on both components is invariant to this differing normalization of loadings and factors $\hat{\Lambda}_i' \hat{\boldsymbol{F}}_t$: *(1)* The idiosyncratic components are calculated via this complete term of common components, which the normalization does not alter. *(2)* The common factors $\hat{\boldsymbol{F}}_t$ exhibit the same strength of cointegration, i.e. $\hat{\boldsymbol{\lambda}}$ of the reduced-rank regression (8), irrespective of any scaling weight multiplied to a time series of $\hat{\boldsymbol{F}}_t$.

In order to determine the cointegration rank within $\boldsymbol{F}_t$, we employ the single VECM (2) with a linear trend in the cointegration space. The proceeding is the same as any individual analysis of the country-specific time series. The AIC, as provided by **vars**' `VARselect()` or **pvars**' `speci.VAR()`, hints at a lag order of $p_{F_t} = 2$ for the VAR Model (1) without rank-restriction. We perform both procedures, Johansen (1996) and SL (2000c), using an unrestricted intercept and a linear trend restricted to the coingreation (*Case 4*, see Table 2) and the corresponding intercept and linear trend in the additive Model (10) (*SL_trend*).

```
R> R.Ftsl = coint.SL(R.pcsl$CSD$Ft, dim_p=2, type_SL="SL_trend")
R> R.Ftjo = coint.JO(R.pcsl$CSD$Ft, dim_p=2, type="Case4")
R> toLatex(R.Ftsl, R.Ftjo, write_ME=TRUE,
+     add2header=c("\\textbf{SL:} Trace test",
+                  "\\textbf{Johansen:} Trace test"))
```

Both results are jointly exported to Latex via the `toLatex()` method for `coint` objects. In order to distinguish them in the joint table, the optional argument `add2header` labels the procedure in the respective column. Depending on the input `coint` objects, the user could indicate individual names or different specifications manually here as well. In contrast to the panel test functions, the two functions of the `coint` branches conduct the maximum eigenvalue LR-test additionally. The argument `write_ME` controls whether these results shall enter the Latex table. This feature is particular useful for test procedures on more complex data generation processes with trend breaks or weakly exogenous variables. In these cases, the distribution moments of $Z_{r_\perp}$ and thus p-values would be available for the trace test only such that the slot `$pvals_ME` of the `coint` object would report a vector of `NA`s. However, since the empirical example implies standard specifications, the complete Table 5 has been printed with `write_ME=TRUE`.

Table 5: Cointegration rank tests for the common factors.

	SL: Trace test		Max. eigenvalue test		**Johansen:** Trace test		Max. eigenvalue test	
r_{H0}	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
0	202.448	0.000	63.178	0.001	251.612	0.000	74.245	0.000
1	116.540	0.080	40.822	0.124	177.367	0.000	57.414	0.006
2	85.131	0.125	38.750	0.043	119.952	0.034	38.361	0.206
3	47.891	0.622	20.551	0.637	81.591	0.147	30.719	0.297
4	27.169	0.794	14.439	0.716	50.873	0.379	19.469	0.699
5	14.737	0.784	10.300	0.642	31.404	0.427	14.536	0.684
6	3.719	0.955	3.465	0.894	16.868	0.433	9.713	0.656
7	1.406	0.684	1.406	0.684	7.155	0.338	7.155	0.339

The table shows results for the thereby four different tests on the cointegration rank within the eight common factors. In accordance with Table 6 in Örsal and Arsova (2017, p. 68), the trace tests in the SL- and the Johansen procedure stop at an r_{H0} of one resp. three cointegration relations and thus suggest at least five global stochastic trends driving the observable variables $\boldsymbol{y}_{it}$ in the PANIC Model (32). Correspondingly, the maximum eigenvalue test procedures suggest ranks of one resp. two at a significance level of 5%. The range of five to seven global stochastic trends are also in accord with the $IPC(k)$ criteria.

4.2. The exchange rate pass-through: Specify the deterministic term

The empirical example from Arsova and Örsal (2016, Ch. 6) about the *exchange rate pass-through* (ERPT) shows how to perform cointegration rank analysis under the consideration of structural breaks in the cointegration relation. Their example itself is based on the data set and theoretical model from Banerjee and Carrion-i-Silvestre (2015), who assess a single long-run relationship between the logarithmized time series of import price mp_{it}, exchange rate er_{it}, and foreign price fp_t in the linear model

$$mp_{it} = \mu_{i0} + \mu_{i1}t + \beta_{i1} \cdot er_{it} + \beta_{i2} \cdot fp_t + u_{it}. \tag{37}$$

Data.[28] The data set on the $K = 3$ variables $\boldsymbol{y}_{it}$ consists of monthly observations over the period 1995/01 – 2005/03 ($T = 123$) for $N = 7$ Euro-area countries and nine different industries. However for the illustration, we reproduce the empirical analysis for the industry *"chemicals and related products"* only and it is up to the user to try out cointegration tests for the remaining industries. Their time series are also stored in **pvars**' data set `ERPT`. In order to subset and transform the long-format panel `ERPT` into the necessary list-format, we follow the same steps as explained at the beginning of Section 4 and select the variables for the chemical industry denoted by serial number 5. Only for an easier comparison with the results from Arsova and Örsal (2016), the country names `names_i` are re-ordered according to the original literature. This has no consequences on the implementation.

```
R> library("pvars")
R> data("ERPT")
R> names_k = c("lpm5", "lfp5", "llcusd")
R> names_i = levels(ERPT$id_i)[c(1,6,2,5,4,3,7)]
R> L.data  = sapply(names_i, FUN=function(i)
+     ts(ERPT[ERPT$id_i==i, names_k], start=c(1995, 1), frequency=12),
+     simplify=FALSE)
```

Over the considered sample period, Arsova and Örsal (2016) suspect a level shift and trend break in May 2002 motivated by the appreciation of the Euro against the US-Dollar. The authors attribute this persistent change to effects from the outside of Model (37), namely *(i)* the aftermath of the terrorist attacks on the World Trade Center on 11 September 2001, *(ii)* hesitant investors on the US markets, *(iii)* the declining importance of US exports on the world markets, and *(iv)* the fully established Euro after the national currencies had been withdrawn from circulation in March 2002. Since the complete data set including the presample periods comprises the time interval 1995/01 – 2007/12, the single break period τ of 2002/05 is counted to be the 89th one within the interval $1, \ldots, 123$.

[28] The balanced panel `ERPT` as used by Arsova and Örsal (2016) contains less individuals than Banerjee and Carrion-i-Silvestre (2015) actually provide. The countries Austria, Finland, and Portugal are omitted because Eurostat as the primary data source has reported some missing values in these time series.

Specify the deterministic term. In order to accommodate this structural break, the CAIN-test considers individual TSL-procedures by Trenkler *et al.* (2008), which allows for individual-specific deterministic components in the additive Model (10) given by

$$M_{\mu i}\boldsymbol{d}_{it} = [\boldsymbol{\mu}_{0i} : \boldsymbol{\mu}_{1i} : \boldsymbol{\delta}_{0i} : \boldsymbol{\delta}_{1i}]\,\boldsymbol{d}_{it} = \boldsymbol{\mu}_{0i} + \boldsymbol{\mu}_{1i}t + \boldsymbol{\delta}_{0i}d_{it} + \boldsymbol{\delta}_{1}b_{it}. \tag{38}$$

Besides the conventional constant 1 and linear trend t, the $n_i \times 1$ vectors $\boldsymbol{d}_{it}$ stack the period-specific shift dummies d_{it} and trend breaks b_{it}. These regressors are $d_{it} = b_{it} = 0$ firstly and activated to be $d_{it} = 1$ and $b_{it} = t - \tau_i + 1$ when $t \geq \tau_i$. The empirical example holds the special characteristics that there is only a single break and it occurs at the very same period $\tau_i = \tau = 89$ in each i. Therefore, the shift, which must accompany the single break, is the sole $d_t = \Delta b_t$ and the deterministic regressors $\boldsymbol{d}_t = (1, t, \Delta b_t, b_t)'$ are in fact identical for each i. This has no further consequences for the flexible CAIN-TSL test itself, but its implementation `pcoint.CAIN()` then requires only the single value $\tau = 89$ instead of a list of N specifications for the argument `t_D`. Accordingly, the objects `R.t_D` and `L.t_D` lead to identical results. The latter can serve as a basis for adding individual-specific regressors d_{it}. In this example, the explicit formulation for France is yet redundant as **pvars** adds any accompanying shift from `t_break` automatically.

```
R> R.t_D = list(t_break=89)
R> L.t_D = sapply(names_i, function(i) list(t_break=89), simplify=FALSE)
R> L.t_D$France$t_shift = c(89)
```

For the feasible GLS-detrending,[29] Table 2 shows how **pvars** constructs $\boldsymbol{d}_{1,it}$ and $\boldsymbol{d}_{2,it}$ in the preceding reduced-rank regressions of $\Delta\boldsymbol{y}_{it}$ on $\boldsymbol{z}_{1,it} = \left(\boldsymbol{y}'_{i,t-1}, \boldsymbol{d}'_{1,it}\right)'$ corrected for $\boldsymbol{z}_{2,it} = \left(\Delta\boldsymbol{y}'_{i,t-1}, \ldots, \Delta\boldsymbol{y}'_{i,t-p_i+1}, \boldsymbol{d}'_{2,it}\right)'$. The deterministic regressors $\boldsymbol{d}_{1,it} = (t-1, b_{t-1})'$ for the cointegration relations consist of a linear trend and its break in $\tau = 89$. In addition to the unrestricted constant, **pvars** includes the shift Δb_t and lags of the impulse dummy $d^{\text{im}}_{\tau,t}$ into $\boldsymbol{d}_{2,it} = \left(1, \Delta b_t, d^{\text{im}}_{\tau,t}, \ldots, d^{\text{im}}_{\tau,t-(p_i-1)}\right)'$ automatically. In converse notation, these lags can be easily recognized as a solitary impulse $d^{\text{im}}_{\bullet,t} = 1$ in each period $\bullet \in \{\tau, \ldots, \tau + (p_i - 1)\}$.

Panel cointegration tests. Arsova and Örsal (2021) determine the individual lag orders p_i with the modified AIC by Qu and Perron (2007) under the null hypothesis of no cointegration. The feasible estimation of $M_{\mu i}$ in Eq. (38) and the LR-test based on an individual VECM with detrended time series then proceed as generally described in Section 2.1. Finally, the panel statistics (33) by Hartung (1999) and Arsova and Örsal (2021) combine the individual p-values under the three correction factors to account for cross-sectional dependence. We conduct the complete CAIN test procedure with the `pcoint` command

```
R> R.lags = c(3, 3, 3, 4, 4, 3, 4)  # by modified AIC
R> R.cain = pcoint.CAIN(L.data, lags=R.lags, t_D=R.t_D, type="SL_trend")
R> R.cain$CSD$rho_tilde

            r_H0 = 0  r_H0 = 1  r_H0 = 2
Hartung_K1 0.8123894 0.3973220 0.5726069
Hartung_K2 0.7954536 0.3583592 0.5403518
CAIN       0.1252727 0.1310724 0.1448498
```

[29]See TSL (2008, p. 335) for more details and Arsova and Örsal (2021, Ch. 6) for another panel example.

```
R> R.cain$CSD$rho_eps
```

```
[1] 0.6333148
```

In the S3-slot `$CSD`, the matrix `rho_tilde` contains the correction factors for each panel test and hypothesis $r_{H0} = 0, \ldots, K-1$. Further, `rho_eps` stores the *average absolute cross-sectional correlation* ρ_ϵ calculated only for $r_{H0} = 0$. Its high value indicates strong cross-sectional dependence for the empirical example. Arsova and Örsal (2016, p. 15) attribute this to the common foreign price series fp_t and similar exchange rate dynamics er_{it}. After the Euro exchange rate was fixed on 31 December 1998, the variable er_{it} has been identical for the considered European countries from January 1999 onward.

```
R> toLatex(R.cain)
```

Table 6: Panel cointegration rank tests for ERPT.

Individual		statistics			*p*-values		
	lags	$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$	$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$
France	3	35.006	13.709	3.812	0.024	0.248	0.425
Netherlands	3	32.854	11.754	5.514	0.044	0.402	0.220
Germany	3	36.446	20.365	2.438	0.015	0.029	0.666
Italy	4	39.369	18.309	1.245	0.006	0.060	0.888
Ireland	4	32.834	19.624	1.814	0.045	0.038	0.786
Greece	3	34.559	16.713	3.900	0.027	0.102	0.412
Spain	4	28.230	9.743	2.108	0.146	0.598	0.730
Panel		statistics			*p*-values		
		$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$	$r_{H0}=0$	$r_{H0}=1$	$r_{H0}=2$
Hartung $K1$		-2.034	-1.477	0.335	0.021	0.070	0.631
Hartung $K2$		-2.052	-1.531	0.343	0.020	0.063	0.634
CAIN		-3.725	-2.033	0.517	0.000	0.021	0.697

As in Section 4.1, the `pcoint` object is printed by `toLatex()` and encapsulated in Latex' `table` environment. Table 6 displays the individual and panel results for each hypothesis $r_{H0} = 0, \ldots, K-1$, which Arsova and Örsal (2016, p. 22) report in Table 7 and 8 for the industry of *"5: Chemicals and related products"*, too. Comparing the p-values to a significance level of $\alpha = 5\%$, the sequential testing procure of CAIN rejects the hypotheses up to $r_{H0} = 2$ and thus suggests the presence of overall two cointegration relations.

5. Summary

This article has presented a set of VAR methods for panel data (Section 2), described their implementation in **pvars** (Section 3), and illustrated their application (Section 4). The R-package comprises *panel cointegration rank tests* as well as *estimators of VAR models for heterogeneous panels* and *panel methods for structural identification* of the reduced-form VAR models. In this context, **pvars** addresses typical properties of financial and macroeconomic panel data, in particular cross-sectional dependence and structural breaks in the deterministic term. Finally, **pvars** supplements functions for model specification and dynamic analysis which are not provided by other packages of the **vars**-ecosystem, namely various *criteria for the number of common factors*, *persistence profiles*, *mean-group IRF*, and *moving-block bootstrap procedures* for panel SVAR models.

Future research may extend bootstrapped tests for the individual cointegration rank to panel tests. Swensen (2006, 2009) and Cavaliere, Rahbek, and Taylor (2012, 2014) construct bootstrapped tests for the individual Johansen procedure. Trenkler (2009) and Cavaliere, Taylor, and Trenkler (2013) compare bootstrapped SL-procedures in view of deterministic components. In particular, the bootstrap-after-bootstrap procedure by Cavaliere, Taylor, and Trenkler (2015) exhibits improved small-T sample performance and robustness against conditional heteroskedasticity and serial correlation in comparison to the asymptotic test procedure. An extension based on panel blocks could respect additional cross-sectional dependence and would be easily accommodated into **pvars**. Yet, these forms of residual structure can often be attributed to common or extraordinary shocks and thus treated by common factors or deterministic dummies (Juselius 2007, Ch. 6.7). Their rigorous implementation throughout the different VAR applications is readily available in **pvars**.

References

Abrigo M, Love I (2016). "Estimation of Panel Vector Autoregression in Stata." *Stata Journal*, **16**(3), pp. 778–804. doi:10.1177/1536867X1601600314.

Ahn SK, Horenstein A (2013). "Eigenvalue Ratio Test for the Number of Factors." *Econometrica*, **81**(3), pp. 1203–1227. doi:10.3982/ECTA8968.

Ahn SK, Reinsel GC (1990). "Estimation for Partially Nonstationary Multivariate Autoregressive Models." *Journal of the American Statistical Association*, **85**(411), pp. 813–823. doi:10.1080/01621459.1990.10474945.

Amisano G, Giannini C (1997). *Topics in Structural VAR Econometrics.* 2nd edition. Springer. ISBN 978-3-642-60623-6.

Anderson TW (1951). "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions." *Annals of Mathematical Statistics*, **22**(3), pp. 327–351. doi:10.1214/aoms/1177729580.

Anderson TW (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley Series in Probability and Statistics, 3rd edition. Wiley. ISBN 0471360910,9780471360919.

Arsova A, Örsal DDK (2016). "A Panel Cointegrating Rank Test with Structural Breaks and Cross-Sectional Dependence." *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Time Series Econometrics No. D01-V3*, German Economic Association.

Arsova A, Örsal DDK (2018). "Likelihood-based Panel Cointegration Test in the Presence of a Linear Time Trend and Cross-Sectional Dependence." *Econometric Reviews*, **37**(10), pp. 1033–1050. doi:10.1080/07474938.2016.1183070.

Arsova A, Örsal DDK (2021). "A Panel Cointegrating Rank Test with Structural Breaks and Cross-Sectional Dependence." *Econometrics and Statistics*, **17**(C), pp. 107–129. doi:10.1016/j.ecosta.2020.05.002.

Bada O, Liebl D (2014). "**phtt**: Panel Data Analysis with Heterogeneous Time Trends in R." *Journal of Statistical Software*, **59**(6), pp. 1–33. ISSN 1548-7660. doi:10.18637/jss.v059.i06.

Bai J (2004). "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data." *Journal of Econometrics*, **122**(1), pp. 137–183. doi:10.1016/j.jeconom.2003.10.022.

Bai J, Ng S (2002). "Determining the Number of Factors in Approximate Factor Models." *Econometrica*, **70**(1), pp. 191–221. doi:10.1111/1468-0262.00273.

Bai J, Ng S (2004). "A PANIC Attack on Unit Roots and Cointegration." *Econometrica*, **72**(4), pp. 1127–1177. doi:10.1111/j.1468-0262.2004.00528.x.

Bai J, Perron P (1998). "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, **66**(1), pp. 47–78. doi:10.2307/2998540.

Bai J, Perron P (2003). "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics*, **18**(1), pp. 1–22. doi:10.1002/jae.659.

Bakirov NK, Rizzo ML, Székely GJ (2006). "A Multivariate Nonparametric Test of Independence." *Journal of Multivariate Analysis*, **97**, pp. 1742–1756. doi:10.1016/j.jmva.2005.10.005.

Banerjee A, Carrion-i-Silvestre JL (2015). "Cointegration in Panel Data with Structural Breaks and Cross-Section Dependence." *Journal of Applied Econometrics*, **30**(1), pp. 1–23. doi:10.1002/jae.2348.

Bernoth K, Herwartz H (2021). "Exchange Rates, Foreign Currency Exposure and Sovereign Risk." *Journal of International Money and Finance*, **117**(C), 102454. doi:10.1016/j.jimonfin.2021.102454.

Blackburne E, Frank M (2007). "Estimation of Nonstationary Heterogeneous Panels." *Stata Journal*, **7**(2), pp. 197–208. doi:10.1177/1536867X0700700204.

Breitung J (2005). "A Parametric Approach to the Estimation of Cointegration Vectors in Panel Data." *Econometric Reviews*, **24**, pp. 151–173. doi:10.1081/ETC-200067895.

Breitung J, Brüggemann R, Lütkepohl H (2004). "Structural Vector Autoregressive Modeling and Impulse Responses." In H Lütkepohl, M Krätzig (eds.), *Applied Time Series Econometrics*, Themes in Modern Econometrics, chapter 4, pp. pp. 159–196. Cambridge University Press. ISBN 9780521547871.

Bronder S (2016). "**PANICr**: Panel Analysis of Nonstationarity in Idiosyncratic and Common Components." *R package version 1.0*, CRAN. URL https://CRAN.R-project.org/package=PANICr.

Brüggemann R, Jentsch C, Trenkler C (2016). "Inference in VARs with Conditional Heteroskedasticity of Unknown Form." *Journal of Econometrics*, **191**(1), pp. 69–85. doi:10.1016/j.jeconom.2015.10.004.

Brüggemann R, Lütkepohl H (2005). "Practical Problems with Reduced-Rank ML Estimators for Cointegration Parameters and a Simple Alternative." *Oxford Bulletin of Economics and Statistics*, **67**(5), pp. 673–690. doi:10.1111/j.1468-0084.2005.00136.x.

Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001). "A Method for Making Group Inferences from Functional MRI Data using Independent Component Analysis." *Human Brain Mapping*, **16**(2), pp. 131–131. doi:10.1002/hbm.10044.

Canova F, Ciccarelli M (2013). "Panel Vector Autoregressive Models: A Survey." *Advances in Econometrics*, **32**, pp. 205–246. doi:10.1108/S0731-9053(2013)0000031006.

Carrion-i Silvestre JL, Surdeanu L (2011). "Panel Cointegration Rank Testing with Cross-Section Dependence." *Studies in Nonlinear Dynamics & Econometrics*, **15**(4), pp. 1–43. doi:10.2202/1558-3708.1825.

Cavaliere G, Rahbek A, Taylor R (2012). "Bootstrap Determination of the Co-Integration Rank in Vector Autoregressive Models." *Econometrica*, **80**(4), pp. 1721–1740. doi:10.3982/ECTA9099.

Cavaliere G, Rahbek A, Taylor R (2014). "Bootstrap Determination of the Co-Integration Rank in Heteroskedastic VAR Models." *Econometric Reviews*, **33**(5-6), pp. 606–650. doi:10.1080/07474938.2013.825175.

Cavaliere G, Taylor R, Trenkler C (2013). "Bootstrap Cointegration Rank Testing: The Role of Deterministic Variables and Initial Values in the Bootstrap Recursion." *Econometric Reviews*, **32**(7), pp. 814–847. doi:10.1080/07474938.2012.690677.

Cavaliere G, Taylor R, Trenkler C (2015). "Bootstrap Co-integration Rank Testing: The Effect of Bias-Correcting Parameter Estimates." *Oxford Bulletin of Economics and Statistics*, **77**(5), pp. 740–759. doi:10.1111/obes.12090.

Choi I (2001). "Unit Root Tests for Panel Data." *Journal of International Money and Finance*, **20**, pp. 249–272. doi:10.1016/S0261-5606(00)00048-6.

Christopoulos D, Tsionas M (2004). "Financial Development and Economic Growth: Evidence from Panel Unit Root and Cointegration Tests." *Journal of Development Economics*, **73**(1), pp. 55–74. doi:10.1016/j.jdeveco.2003.03.002.

Comon P (1994). "Independent Component Analysis, a new Concept?" *Signal Processing*, **36**, pp. 287–314. doi:10.1016/0165-1684(94)90029-9.

Corona F, Poncela P, Ruiz E (2017). "Determining the Number of Factors after Stationary Univariate Transformations." *Empirical Economics*, **53**(1), pp. 351–372. doi:10.1007/s00181-016-1158-5.

Croissant Y, Millo G (2008). "Panel Data Econometrics in R: The **plm** Package." *Journal of Statistical Software*, **27**(2). doi:10.18637/jss.v027.i02.

Doornik J (1998). "Approximations to the Asymptotic Distributions of Cointegration Tests." *Journal of Economic Surveys*, **12**(5), pp. 573–93. doi:10.1111/1467-6419.00068.

Dąbrowski MA, Papież M, Śmiech S (2014). "Exchange Rates and Monetary Fundamentals in CEE Countries: Evidence from a Panel Approach." *Journal of Macroeconomics*, **41**, pp. 148–159. ISSN 0164-0704. doi:10.1016/j.jmacro.2014.05.005.

Empting LFT, Herwartz H (2025). "Revisiting the 'Productivity of Public Capital': VAR Evidence on the Heterogeneous Dynamics in a New Panel of 23 OECD Countries." *Working paper*, Universität Göttingen.

Empting LFT, Maxand S, Öztürk S, Wagner K (2025). "Inference in Panel SVARs with Cross-Sectional Dependence of Unknown Form." *Working paper*, Universität Göttingen / Viadrina.

Fisher RA (1932). *Statistical Methods for Research Workers*, volume 5 of *Biological Monographs and Manuals*. Oliver and Boyd.

Gambacorta L, Hofmann B, Peersman G (2014). "The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis." *Journal of Money, Credit and Banking*, **46**(4), pp. 615–642. doi:10.1111/jmcb.12119.

Genest C, Quessy JF, Rémillard B (2007). "Asymptotic Local Efficiency of Cramér–von Mises Tests for Multivariate Independence." *Annals of Statistics*, **35**(1), pp. 166–191. doi:10.1214/009053606000000984.

Groen JJ, Kleibergen F (2003). "Likelihood-based Cointegration Analysis in Panels of Vector Error-Correction Models." *Journal of Business & Economic Statistics*, **21**(2), pp. 295–318. doi:10.1198/073500103288618972.

Hamilton JD (1994). *Time Series Analysis*. Princeton University Press. ISBN 0691042896.

Hartung J (1999). "A Note on Combining Dependent Tests of Significance." *Biometrical Journal*, **41**(7), pp. 849–855. doi:10.1002/(SICI)1521-4036(199911)41:7<849::AID-BIMJ849>3.0.CO;2-T.

Herwartz H (2018). "Hodges–Lehmann Detection of Structural Shocks: An Analysis of Macroeconomic Dynamics in the Euro Area." *Oxford Bulletin of Economics and Statistics*, **80**(4), pp. 736–754. doi:10.1111/obes.12234.

Herwartz H, Lange A, Maxand S (2022). "Data-driven Identification in SVARs — When and how can Statistical Characteristics be used to unravel Causal Relationships?" *Economic Inquiry*, **60**(2), pp. 668–693. doi:10.1111/ecin.13035.

Herwartz H, Wang S (2024). "Statistical Identification in Panel Structural Vector Autoregressive Models based on Independence Criteria." *Journal of Applied Econometrics*, **39**(4), pp. 620–639. doi:10.1002/jae.3044.

Hlouskova J, Wagner M (2010). "The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study." *Econometric Reviews*, **29**(2), pp. 182–223. doi:10.1080/07474930903382182.

Hodges JL, Lehmann EL (1963). "Estimates of Location based on Rank Tests." *Annals of Mathematical Statistics*, **34**(2), pp. 598–611. doi:10.1214/aoms/1177704172.

Im KS, Pesaran MH, Shin Y (2003). "Testing for Unit Roots in Heterogeneous Panels." *Journal of Econometrics*, **115**, pp. 53–74. doi:10.1016/S0304-4076(03)00092-7.

Izenman AJ (1975). "Reduced-Rank Regression for the Multivariate Linear Model." *Journal of Multivariate Analysis*, **5**(2), pp. 248–264. doi:10.1016/0047-259X(75)90042-1.

Jentsch C, Lunsford KG (2022). "Asymptotically Valid Bootstrap Inference for Proxy SVARs." *Journal of Business & Economic Statistics*, **40**(4), pp. 1876–1891. doi:10.1080/07350015.2021.1990770.

Johansen S (1988). "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics & Control*, **12**(2-3), pp. 231–254. doi:10.1016/0165-1889(88)90041-3.

Johansen S (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, **59**(6), pp. 1551–1580. ISSN 00129682, 14680262. doi:10.2307/2938278.

Johansen S (1996). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Advanced Texts in Econometrics. Oxford University Press, USA. ISBN 0198774508,9780198774501,0198774494,9780198774495,9780191525063.

Johansen S, Mosconi R, Nielsen B (2000). "Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend." *Econometrics Journal*, **3**(2), pp. 216–249. doi:10.1111/1368-423X.00047.

Johansen S, Nielsen M (2018). "The Cointegrated Vector Autoregressive Model with General Deterministic Terms." *Journal of Econometrics*, **202**(2), pp. 214–229. doi:10.1016/j.jeconom.2017.10.003.

Juselius K (2007). *The Cointegrated VAR Model: Methodology and Applications*. Advanced Texts in Econometrics, 2nd edition. Oxford University Press, USA. ISBN 0199285675,9780199285679.

Kilian L (1998). "Small-sample Confidence Intervals for Impulse Response Functions." *Review of Economics and Statistics*, **80**(2), pp. 218–230. doi:10.1162/003465398557465.

Kilian L, Lütkepohl H (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press. doi:10.1017/9781108164818.

King R, Plosser C, Stock J, Watson M (1991). "Stochastic Trends and Economic Fluctuations." *American Economic Review*, **81**(4), pp. 819–40.

Kojadinovic I, Yan J (2010). "Modeling Multivariate Distributions with Continuous Margins using the **copula** R Package." *Journal of Statistical Software*, **34**(9), pp. 1–20. doi:10.18637/jss.v034.i09.

Kurita T, Nielsen B (2019). "Partial Cointegrated Vector Autoregressive Models with Structural Breaks in Deterministic Terms." *Econometrics*, **7**(4), pp. 1–35. ISSN 2225-1146. doi:10.3390/econometrics7040042.

Lange A, Dalheimer B, Herwartz H, Maxand S (2021). "**svars**: An R Package for Data-Driven Identification in Multivariate Time Series Analysis." *Journal of Statistical Software*, **97**(5), pp. 1–34. ISSN 1548-7660. doi:10.18637/jss.v097.i05.

Larsson R, Lyhagen J, Lothgren M (2001). "Likelihood-based Cointegration Tests in Heterogeneous Panels." *Econometrics Journal*, **4**(1), pp. 109–142. doi:10.1111/1368-423X.00059.

Lee KC, Pesaran MH (1993). "Persistence Profiles and Business Cycle Fluctuations in a Disaggregated Model of U.K. Output Growth." *Ricerche Economiche*, **47**(3), pp. 293–322. doi:10.1016/0035-5054(93)90032-X.

Luukkonen R, Ripatti A, Saikkonen P (1999). "Testing for a Valid Normalization of Cointegrating Vectors in Vector Autoregressive Processes." *Journal of Business & Economic Statistics*, **17**(2), pp. 195–204. doi:10.2307/1392475.

Lütkepohl H (2005). *New Introduction to Multiple Time Series Analysis.* 2nd edition. Springer. ISBN 3540401725,9783540401728.

Lütkepohl H (2006). "Structural Vector Autoregressive Analysis for Cointegrated Variables." *AStA Advances in Statistical Analysis*, **90**(1), pp. 75–88. doi:10.1007/s10182-006-0222-4.

Lütkepohl H, Saikkonen P (2000). "Testing for the Cointegrating Rank of a VAR Process with a Time Trend." *Journal of Econometrics*, **95**(1), pp. 177–198. doi:10.1016/S0304-4076(99)00035-4.

Lütkepohl H, Saikkonen P, Trenkler C (2004). "Testing for the Cointegrating Rank of a VAR Process with Level Shift at Unknown Time." *Econometrica*, **72**(2), pp. 647–662. doi:10.1111/j.1468-0262.2004.00505.x.

Maddala GS, Wu S (1999). "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test." *Oxford Bulletin of Economics and Statistics*, **61**, pp. 631–652. doi:10.1111/1468-0084.0610s1631.

Matteson DS, Tsay RS (2017). "Independent Component Analysis via Distance Covariance." *Journal of the American Statistical Association*, **112**(518), pp. 623–637. doi:10.1080/01621459.2016.1150851.

Mullen KM, Ardia D, Gil DL, Windover D, Cline J (2011). "**DEoptim**: An R Package for Global Optimization by Differential Evolution." *Journal of Statistical Software*, **40**(6), pp. 1–26. doi:10.18637/jss.v040.i06.

Nickell S (1981). "Biases in Dynamic Models with Fixed Effects." *Econometrica*, **49**(6), pp. 1417–26. doi:10.2307/1911408.

Onatski A (2010). "Determining the Number of Factors from Empirical Distribution of Eigenvalues." *Review of Economics and Statistics*, **92**(4), pp. 1004–1016. doi:10.1162/REST_a_00043.

Osterwald-Lenum M (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics." *Oxford Bulletin of Economics and Statistics*, **54**(3), pp. 461–72. doi:10.1111/j.1468-0084.1992.tb00013.x.

Pedroni P (2019). "Panel Cointegration Techniques and Open Challenges." In M Tsionas (ed.), *Panel Data Econometrics: Theory*, chapter 10, pp. pp. 251–287. Academic Press. ISBN 978-0-12-814367-4.

Persyn D, Westerlund J (2008). "Error-Correction-Based Cointegration Tests for Panel Data." *Stata Journal*, **8**(2), pp. 232–241. doi:10.1177/1536867X0800800205.

Pesaran MH, Shin Y (1996). "Cointegration and Speed of Convergence to Equilibrium." *Journal of Econometrics*, **71**(1-2), pp. 117–143. doi:10.1016/0304-4076(94)01697-6.

Pesaran MH, Shin Y, Smith RJ (2000). "Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables." *Journal of Econometrics*, **97**, pp. 293–343. doi:10.1016/S0304-4076(99)00073-1.

Pesaran MH, Shin Y, Smith RP (1999). "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels." *Journal of the American Statistical Association*, **94**(446), pp. 621–634. doi:10.2307/2670182.

Pesaran MH, Smith R (1995). "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels." *Journal of Econometrics*, **68**(1), pp. 79–113. doi:10.1016/0304-4076(94)01644-F.

Pfaff B (2008a). *Analysis of Integrated and Cointegrated Time Series with R*. Use R!, 2nd edition. Springer. ISBN 0387759662,9780387759661. URL http://CRAN.R-project.org/package=urca.

Pfaff B (2008b). "VAR, SVAR and SVEC Models: Implementation within R Package **vars**." *Journal of Statistical Software*, **27**(4), pp. 1–32. doi:10.18637/jss.v027.i04.

Pretis F (2020). "Econometric Modelling of Climate Systems: The Equivalence of Energy Balance Models and Cointegrated Vector Autoregressions." *Journal of Econometrics*, **214**(1), pp. 256–273. doi:10.1016/j.jeconom.2019.05.013.

Qu Z, Perron P (2007). "A Modified Information Criterion for Cointegration Tests based on a VAR Approximation." *Econometric Theory*, **23**(04), pp. 638–685. doi:10.1017/S0266466607070284.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rebucci A (2010). "Estimating VARs with Long Stationary Heterogeneous Panels: A Comparison of the Fixed Effect and the Mean Group Estimators." *Economic Modelling*, **27**(5), pp. 1183–1198. doi:10.1016/j.econmod.2010.03.001.

Risk BB, James NA, Matteson DS (2015). "**steadyICA**: ICA and Tests of Independence via Multivariate Distance Covariance." *R package version 1.0*, CRAN. URL https://CRAN.R-project.org/package=steadyICA.

Risk BB, Matteson DS, Ruppert D, Eloyan A, Caffo BS (2014). "An Evaluation of Independent Component Analyses with an Application to Resting-State fMRI." *Biometrics*, **70**(1), pp. 224–236. ISSN 0006-341X. doi:10.1111/biom.12111.

Rizzo M, Szekely G (2022). "**energy**: E-Statistics: Multivariate Inference via the Energy of Data." *R package version 1.7-9*, CRAN. URL https://CRAN.R-project.org/package=energy.

Saikkonen P, Lütkepohl H (2000a). "Testing for the Cointegrating Rank of a VAR Process with an Intercept." *Econometric Theory*, **16**(3), pp. 373–406. doi:10.1017/S0266466600163042.

Saikkonen P, Lütkepohl H (2000b). "Testing for the Cointegrating Rank of a VAR Process with Structural Shifts." *Journal of Business & Economic Statistics*, **18**(4), pp. 451–64. doi:10.1080/07350015.2000.10524884.

Saikkonen P, Lütkepohl H (2000c). "Trend Adjustment Prior to Testing for the Cointegrating Rank of a Vector Autoregressive Process." *Journal of Time Series Analysis*, **21**(4), pp. 435–456. doi:10.1111/1467-9892.00192.

Sharpsteen C, Bracken C (2020). "**tikzDevice**: R Graphics Output in LaTeX Format." *R package version 0.12.3.1*, CRAN. URL https://CRAN.R-project.org/package=tikzDevice.

Sigmund M, Ferstl R (2019). "Panel Vector Autoregression in R with the Package **Panelvar**." *Quarterly Review of Economics and Finance*. doi:10.2139/ssrn.2896087.

Sims CA (1980). "Macroeconomics and Reality." *Econometrica*, **48**(1), pp. 1–48. doi:10.2307/1912017.

Smyth R, Narayan P (2015). "Applied Econometrics and Implications for Energy Economics Research." *Energy Economics*, **50**(C), pp. 351–358. doi:10.1016/j.eneco.2014.07.023.

StataCorp (2019). *Stata: Statistical Software: Release 16*. StataCorp LLC, College Station, Texas. URL https://www.stata.com/.

Stoufer SA, Suchman EA, DeVinney Leland Cand Star SA, Williams RM (1949). *The American Soldier*, volume 1 of *Studies in Social Psychology in World War II*. Princeton University Press. doi:10.1177/000271624926500124.

Swensen AR (2006). "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models." *Econometrica*, **74**(6), pp. 1699–1714. doi:10.1111/j.1468-0262.2006.00723.x.

Swensen AR (2009). "Corrigendum to "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models"." *Econometrica*, **77**(5), pp. 1703–1704. doi:10.3982/ECTA8201.

Székely GJ, Rizzo ML, Bakirov NK (2007). "Measuring and Testing Dependence by Correlation of Distances." *Annals of Statistics*, **35**(6), pp. 2769–2794. doi:10.1214/009053607000000505.

Tang Y, Horikoshi M, Wenxuan L (2016). "**ggfortify**: Unified Interface to Visualize Statistical Results of Popular R Packages." *The R Journal*, **8**(2), pp. 474–485. doi:10.32614/RJ-2016-060.

Trenkler C (2008). "Determining p-Values for Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms." *Computational Statistics*, **23**(1), pp. 19–39. doi:10.1007/s00180-007-0066-8.

Trenkler C (2009). "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms." *Econometric Theory*, **25**(1), pp. 243–269. doi:10.1017/S0266466608090087.

Trenkler C, Saikkonen P, Lütkepohl H (2008). "Testing for the Cointegrating Rank of a VAR Process with Level Shift and Trend Break." *Journal of Time Series Analysis*, **29**(2), pp. 331–358. doi:10.1111/j.1467-9892.2007.00558.x.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer. ISBN 0-387-95457-0. URL http://www.stats.ox.ac.uk/pub/MASS4/.

Vlaar PJ (2004). "On the Asymptotic Distribution of Impulse Response Functions with Long-Run Restrictions." *Econometric Theory*, **20**(5), pp. 891–903. doi:10.1017/S0266466604205047.

Wickham H (2007). "Reshaping Data with the **reshape** Package." *Journal of Statistical Software*, **21**(12), pp. 1–20. doi:10.18637/jss.v021.i12.

Wickham H (2011). "**testthat**: Get Started with Testing." *The R Journal*, **3**(1), pp. 5–10. doi:10.32614/RJ-2011-002.

Wickham H (2016). ***ggplot2****: Elegant Graphics for Data Analysis*. Use R!, 2nd edition. Springer. ISBN 331924275X,9783319242750. URL https://ggplot2.tidyverse.org.

Yang M (2002). "Lag Length and Mean Break in Stationary VAR Models." *Econometrics Journal*, **5**(2), pp. 374–387. doi:10.1111/1368-423X.00089.

Örsal DDK (2017). "Analysing Money Demand Relation for OECD Countries using Common Factors." *Applied Economics*, **49**(60), pp. 6003–6013. doi:10.1080/00036846.2017.1371842.

Örsal DDK, Arsova A (2017). "Meta-Analytic Cointegrating Rank Tests for Dependent Panels." *Econometrics and Statistics*, **2**(C), pp. 61–72. doi:10.1016/j.ecosta.2016.10.001.

Örsal DDK, Droge B (2011). "Corrigendum to 'Likelihood-based cointegration tests in heterogeneous panels' (Larsson R., J. Lyhagen and M. Löthgren, Econometrics Journal, 4, 2001, 109–142)." *Econometrics Journal*, **14**, pp. 121–125. doi:10.1111/j.1368-423X.2010.00327.x.

Örsal DDK, Droge B (2014). "Panel Cointegration Testing in the Presence of a Time Trend." *Computational Statistics & Data Analysis*, **76**(C), pp. 377–390. doi:10.1016/j.csda.2012.05.017.

Affiliation:

Lennart Empting
Faculty of Economics
Universität Duisburg-Essen
Universitätsstraße 12

45141 Essen, Germany
E-mail: lennart.empting@vwl.uni-due.de
URL: https://www.github.com/Lenni89