ESTIMATING AND TESTING FOR A COMMON CO-INTEGRATION SPACE IN LARGE PANEL VECTOR AUTOREGRESSIONS

LAURENT A. F. CALLOT
CREATES
AARHUS UNIVERSITY
LCALLOT@CREATES.AU.DK

Abstract. This paper proposes a maximum likelihood estimator for a common co-integration space in large panels of co-integrated Vector Autoregressive models. The method pioneered by Pesaran (2006) and further refined in Dees et al. (2007) is used to reduce the dimension of the parameter space of the model and control for cross section dependence. The common co-integration space is estimated using standard optimization methods The common co-integration space is estimated using standard optimization methods.. Test statistics for the existence of a common co-integration space against the hypothesis of heterogeneous co-integration spaces are also derived. A bootstrap algorithm to generated pseudo data under the hypothesis of a common co-integration space is proposed, and bootstrap test statistics are derived. Identification of the co-integration vectors of the panel is also discussed.

Key words: Large panels, VAR, VECM, Co-integration, Bootstrap, Homogeneity testing, Identification.
JEL classification: C12, C32, C33

1. Introduction

Large panels have been the focus of much research in the past couple of decades, and they are increasingly used in the empirical literature to take advantage of the wealth of data that is often available. In particular, a lot of attention has been devoted to panels with nonstationary variables and co-integration. Phillips and Moon (1999) develop asymptotic theory for the type of double indexed processes that naturally arise in this literature. Much research has been aimed at testing for co-integration and the number of co-integration vectors in panels (Larsson et al. (2001), Pedroni (2004), Callot (2010)). For reviews of the vast literature on unit root and co-integration testing in non-stationary panels, see Banerjee (1999); Breitung and Pesaran (2008).

The main purpose of this paper is to propose an estimator and a test for a common co-integration space (CCS) in Panels of Co-integrated VAR (PCVAR) models with a large cross section (N) and a large time (T) dimension. A common co-integration space is understood here as meaning the co-integration vectors of each cross-sectional unit span the same space. A bootstrap algorithm to test the common co-integration space hypothesis is proposed and the conditions under which the vectors of the panel's co-integration space are identified are derived.

Most estimation procedures use some form of pooling of the data to increase the precision with which the parameters of interest can be estimated. They rely on the assumption that the dynamics of the panel are similar across units. However, testing whether this assumption holds has received much less attention. The hypothesis of homogeneity of the parameters is itself of interest: it conditions

Date: September 11, 2012.
I would like to thank Niels Haldrup, Søren Johansen, Rolf Larsson, Johan Lyhagen, and seminar participants at the university of Uppsala for valuable comments and discussion. Financial support by the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation, is gratefully acknowledged.

the validity of pooling data and reveals the differences in the dynamics of the cross sectional units under scrutiny. Concerning the possibility of co-integration a useful reference is Westerlund and Hess (2011), who focuses on testing homogeneity of a single co-integration vector in large N, large T panels. In a setting with a small (fixed) N and large T, Larsson and Lyhagen (2007) develop a likelihood based estimator and a test for a common co-integration space and Groen and Kleibergen (2003) uses GMM for the same purpose. However, and in contrast to what is done in the present paper, in their setting co-integration is restricted to each cross section, and it requires that the whole panel can be estimated as a large VAR which severely restricts the combinations of N and T for which this method is applicable.

There are two major sources of difficulty arising when working with large panels.

(1) Panels where the number of units is small relative to the time dimension can be estimated as a single model. However the number of parameters in VAR models grows quadratically with the number of units. This is often referred to as the curse of dimensionality. In order to estimate panels with large N and T one has to reduce the dimension of the parameter space and hence impose constraints on the dynamics of the model.
(2) Many economical series exhibit common patters across countries. When estimating panels without taking this fact into account, the residuals are typically cross sectionally correlated. This cross section dependence of the residuals, in the sense of a non-zero residual covariance matrix between two units of the panel, results in biased inference when pooling estimators and test statistics.

Pesaran (2006) proposes a method to break the curse of dimensionality and model cross section dependence in order to obtain residuals that are uncorrelated across units. He assumes that the residuals are generated by a mix of idiosynchratic shocks and unobserved common factors which are the cause of the cross section dependence. Using weighted cross section averages of the data he constructs a proxy for the unobserved common factors, and show that this method is valid even in the case of heterogeneous panels. Dees et al. (2007) extend this method to vector autoregressions, and Chudik and Pesaran (2011) show that it is robust to an unknown number of unobserved common factors, unit roots in the common factors, and a mild degree of heterogeneity in the parameters.

Another popular approach is to use principal component analysis to estimate the unobserved common factors and augment the model with these estimates. A major contribution in that direction is the panel analysis of nonstationarity in idiosyncratic and common components (PANIC) by Bai and Ng (2005). The data generating process assumed is similar to that of Pesaran (2006) in the sense that residual cross section dependence is generated by an unknown number of (potentially integrated) unobserved common factors. They show that by the method of principal components the common factors, and their number, can be consistently estimated. Westerlund and Larsson (2009) shows that the order of the approximation error is such that it does not vanish when pooling PANIC based test statistics across N. Hence PANIC is not a valid approach for pooling test statistics at the panel level. For this reason this paper will focus on the methodology of Pesaran (2006).

Estimation and testing in this paper is based on the likelihood framework for co-integrated VAR models developed in Johansen (1988, 1991, 1995a). This framework allows for a wide range of economic hypothesis to be tested in the form of likelihood ratio tests of linear restrictions on the parameters of the models. The common co-integration space hypothesis will be tested as a restriction on the co-integration space of each unit of the panel.

Gredenhoff and Jacobson (2001) show that the finite sample size of likelihood ratio tests of linear restrictions in co-integrated VAR can significantly exceed their nominal values. They also show that using the bootstrap greatly improves inference in finite samples, bringing the size of the tests closer to their nominal values. Cavaliere et al. (2010) propose an algorithm to bootstrap co-integration rank tests. The present paper uses a similar bootstrap algorithm to generate pseudo-data under the hypothesis of a common co-integration space, which are then used to construct two panel test statistics based on pooling the individual likelihood ratio statistics and pooling of the individual bootstrap *p-value*.

A well known problem in co-integration analysis is that the parameters of the unrestricted estimated co-integration space is not identified. Identification can be obtained by imposing linear restrictions on the parameters of the co-integration vectors that ensure the unicity of the co-integration vectors and adjustment matrix up to a normalization. Johansen (2010) provides an algebraic condition to verify that a set of linear restrictions is identifying. In the panel case we have to ensure that the co-integration vectors are identified within an individual co-integration space, and that co-integration spaces are identified across individuals. I show that if the procedure proposed in Dees et al. (2007) to control for cross section dependence is used, identification across individual co-integration spaces is ensured by construction of the panel. It follows that imposing on each individual co-integration space a set of restrictions that ensures identification of its co-integration vectors is sufficient to ensure identification of the panel as a whole. This permits the computation of standard errors for the parameters of the co-integration space, and testing of overidentifying restrictions.

The rest of the paper is organized as follows: the next section discusses the model, section 3 introduces the common co-integration space estimator and the associated test. The following section introduces the bootstrap algorithm. Section 5 discusses identification, section 6 investigates the finite sample properties of the test, and section 7 presents an empirical application.

A word on notation: the number of variables in the panel is noted p, the number of lags k, units are indexed by $i \in 1, \cdots, N$ and observations by $t \in 1, \cdots, T$. $X_{NT} \xrightarrow[(N,T)_j]{w} X$ means that X_{NT} converges in distribution to X when N and T increase jointly. Identity matrices of dimension p are noted $\mathtt{I_p}$ and 0 matrices of dimension $p \times p$ are noted $\mathtt{0_p}$. The orthogonal complement of a $p \times s$ matrix A is noted $A_\perp$ and has dimension $p \times p - s$ and satisfies $A'_\perp A = 0$.

2. The Model

Consider a panel vector autoregression in error correction form:

$$\Delta Y_t = \widetilde{\alpha}\rho d_t + \psi D_t + \widetilde{\alpha}\widetilde{\beta}' Y_{t-1} + \sum_{l=1}^{k} \widetilde{\Gamma}_l \Delta Y_{t-l} + \widetilde{\epsilon}_t \tag{1}$$

The variables are stacked $Y_t := (Y'_{1,t}, \dots, Y'_{N,t})'$ in a $[Np \times 1]$ vector. $Y_{i,t}$ is the $p \times 1$ vector of variables for unit i. If there is co-integration among the variables of the models, it is useful to write the model in error correction form to highlight the matrix of co-integration vectors $\widetilde{\beta}$, of dimension $Np \times \tilde{r}$, and the matrix of adjustment parameters $\widetilde{\alpha}$ of dimension $Np \times \tilde{r}$. The co-integration rank of the system, *i.e.* the number of co-integration vectors is $\tilde{r}$. The parameter matrices $\widetilde{\Gamma}_l$ and $\widetilde{\Pi}$ are of dimension $Np \times Np$. d_t is vector of deterministic components restricted to lay in the co-integration space, and D_t is the unrestricted vector of deterministic components. $\widetilde{\epsilon}_t$ is a $Np \times 1$ vector of independent Gaussian shocks to the system, with mean 0 and variance-covariance matrix $\widetilde{\Omega}$.

$$\widetilde{\Omega} = \begin{bmatrix} \widetilde{\Omega}_{11} & \dots & \widetilde{\Omega}_{1N} \\ \vdots & \ddots & \vdots \\ \widetilde{\Omega}_{N1} & \dots & \widetilde{\Omega}_{NN} \end{bmatrix}$$

The log-likelihood function for the model given in equation 1 is:

$$\mathcal{L}\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Gamma}_l, \widetilde{\Omega}, \mu_0, \mu_1\right) = -\frac{T}{2}\ln\left(\det\left(\widetilde{\Omega}\right)\right) - \frac{1}{2}\sum_t \left(\widetilde{\epsilon}'_t \widetilde{\Omega}^{-1} \widetilde{\epsilon}_t\right) \tag{2}$$

The number of parameters in the model depends quadratically on N and p. Even for a system with a moderate number of individuals, the parameters of the model become impossible to estimate by maximum likelihood for conventional sizes of T. To estimate the model some reduction of the dimension of the problem is necessary. One way to obtain such a reduction is to control for the cross section dependence across residuals in order to obtain a likelihood function that is separable and hence a more manageable problem.

As mentioned in the introduction, several methods have been proposed to control for cross-section dependence in order to obtain residuals that are uncorrelated across units. The following section will discuss the approach pioneered by Pesaran (2006) and Dees et al. (2007) which is based on augmenting the model with weighted cross sectional averages. I assume a DGP similar to that of Dees et al. (2007):

Assumption 1. *The DGP of Y_i is given by the common factor model:*

$$\begin{aligned} Y_{it} &= \delta_{i0} + \delta_{i1} t + \gamma_i \mathbf{f}_t + \xi_{it} \\ \Delta \xi_{it} &= \boldsymbol{\Psi}_i (L) \nu_{it}, \ \ \nu_{it} \sim \mathcal{N}(0, I_p) \\ \Delta \mathbf{f}_t &= \boldsymbol{\Lambda}(L) \eta_t, \ \ \eta_t \sim \mathcal{N}(0, I_p) \end{aligned}$$

where $\mathbf{f}_t$ is a $m_f \times 1$ vector of common unobserved factors, with γ_i the associated $p \times m_f$ matrix of individual loadings. $\boldsymbol{\Psi}_i(L) = \sum_{l=0}^{\infty} \Psi_l L^l$ and $\boldsymbol{\Lambda}_i(L) = \sum_{l=0}^{\infty} \Lambda_l L^l$ are lag polynomials composed of absolute summable matrices Ψ_l and Λ_l such that $var(\Delta \xi_{it})$ and $var(\Delta \mathbf{f}_t)$ are positive definite and bounded.

I further assume that $E(\nu_{it}\nu_{jt}) = 0$ that is, the idiosyncratic shocks are cross sectionally uncorrelated.

The data generating process in assumption 1 is a fairly general common factor process allowing for m_f common factors and p individual variables integrated of order at most one, with the possibility of co-integration among the individual variables and between those and the common factors. The cross section dependence in the observed data stems entirely from the unobserved common factors.

I now discuss how the common factors can be approximated by the observed variables in order to control for cross section dependence. Construct weighted averages of the data as:

$$Y_{it}^* = \mathbf{w}_i Y_t = \sum_{j=1}^{N} w_{ij} Y_{jt}$$

where the weights are defined as follows:

$$\begin{aligned} &i) \ \ w_{ii} = 0 && ii) w_{ij} \in]0,1[\quad \forall i \neq j \\ &iii) \sum_{j=1}^{N} w_{ij} = 1 && iv) w_{ij} = O\left(N^{-1}\right) \end{aligned}$$

The three first conditions ensure that the weights for individual i construct a weighted average of $Y_{-i} := \{Y_j | j \neq i\}$. The last condition ensures that the average is not dominated by a single individual, so that idiosyncratic dynamics cancel out when N grows large. Dees et al. (2007) show the following:

$$\begin{aligned} \sum_{j=1}^{N} w_{ij} Y_{jt} &= \sum_{j=1}^{N} w_{ij} \left(\delta_{j0} + \delta_{j1} t + \gamma_j \mathbf{f}_t + \xi_{jt}\right) \\ Y_{it}^* &= \delta_{i0}^* + \delta_{i1}^* t + \gamma_i^* \mathbf{f}_t + \xi_{it}^* \\ \mathbf{f}_t &\xrightarrow[N]{q.m.} \left(\gamma_i^{*\prime} \gamma_i^*\right)^{-1} \gamma_i^* \left(Y_{it}^* - \delta_{i0}^* - \delta_{i1}^* t - \xi_{it}^*\right) \end{aligned}$$

where $\xrightarrow[N]{q.m.}$ stands for convergence in quadratic mean when N grows large. Note that the weights may be time varying as long as they satisfy the conditions above, but for simplicity of exposition the weights won't be indexed by t. When the number of cross section units becomes large, and under assumption 1, the common factors can be approximated by averages of the observed variables. Since deterministic components do not play a crucial role in this paper they will be dropped to simplify the exposition. By augmenting the model with weighted averages, the unobserved factors (and hence the cross section dependence) can be controlled for and we have:

$$\begin{aligned} \Delta Y_{i,t} = \alpha_i \beta_i' \left(Y_{i,t-1}', Y_{i,t-1}^{*'} \right)' \quad &+ \quad \Lambda_{i,0} \Delta Y_{i,t}^* \\ &+ \quad \sum_{l=1}^{k} \Gamma_{i,l} \Delta \left(Y_{i,t-l}', Y_{i,t-l}^{*'} \right) + \epsilon_{i,t} \end{aligned} \tag{3}$$

where:

$$\mathrm{Cov}(\epsilon_{it}\epsilon_{jt}') = \mathtt{0_p} \quad \text{for } i \neq j$$

By this transformation we obtain a model with cross sectionally independent innovations. The model for individual i given in equation 3 is not subject to the curse of dimensionality in the sense that the number of parameters is not a function of N. All the while, it still relates Y_{it} to every other variables in the panel through the weighted averages. This provides some interesting properties to the model and deserves further inspection. Consider:

$$Z_{it} \quad = \quad \left(Y_{i,t-1}', Y_{i,t-1}'^{*} \right)' = W_i Y_t$$

The W_i matrix is composed of a first block of p rows with a unit matrix of dimension p between the columns $(i-1)p+1$ and ip and zeros elsewhere. The second block of p rows is composed of unit matrices multiplied by a weight scalar as defined above, except between column $(i-1)p+1$ and ip where it is equal to zero:

$$W_i = \begin{bmatrix} \mathtt{0_p} & \dots & \mathtt{0_p} & \mathtt{I_p} & \mathtt{0_p} & \dots & \mathtt{0_p} \\ w_{i1}\mathtt{I_p} & \dots & w_{ii-1}\mathtt{I_p} & \mathtt{0_p} & w_{ii+1}\mathtt{I_p} & \dots & w_{iN}\mathtt{I_p} \end{bmatrix} \tag{4}$$

This matrix multiplied on Y_t returns a vector of length $2p$ with Y_{it} and the corresponding weighted average Y_{it}^* stacked. Similarly, define the matrix W_{i0}:

$$W_{i0} = \begin{bmatrix} w_{i1}\mathtt{I_p} & \dots & w_{ii-1}\mathtt{I_p} & \mathtt{0_p} & w_{ii+1}\mathtt{I_p} & \dots & w_{iN}\mathtt{I_p} \end{bmatrix} \tag{5}$$

This matrix applied to Y_t returns the ith weighted average Y_i^* such that: $Y_t^* = W_0 Y_t$. Define also $W = [W_1', \dots, W_N']'$ and similarly $W_0 = [W_{10}', \dots, W_{N0}']'$. These matrices are crucial in the formulation of the model, since they provide a link between every unit in the panel while reducing the dimension of the parameter space.

From the individual model given in equation 3 the model for the full panel can be recovered by stacking:

$$\Delta Y_t = \alpha_i \beta_i' W Y_{t-1}, + \Lambda_0 W_0 \Delta Y_t + \sum_{l=1}^{k} \Gamma_l \Delta W Y_{t-l} + \epsilon_t \tag{6}$$

where by construction:

$$\alpha = \begin{bmatrix} \alpha_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha_N \end{bmatrix} \beta = \begin{bmatrix} \beta_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \beta_N \end{bmatrix} \tag{7}$$

Similarly, the lag matrices Λ_0 and Γ_l as well as the covariance matrix for the full panel are block diagonal:

$$\Omega \quad = \quad E[\epsilon_t \epsilon_t'] = \begin{bmatrix} \Omega_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Omega_{NN} \end{bmatrix} \tag{8}$$

The log likelihood of the full panel given in equation 6 is the standard log likelihood function of the gaussian VAR:

$$\begin{aligned}\mathcal{L}(\Pi, \Lambda_0, \Gamma_l, \Omega) &= -\frac{T}{2}\ln\det(\Omega) - \frac{1}{2}\sum_t\left(\epsilon_t'\Omega^{-1}\epsilon_t\right)\\ &= -\frac{T}{2}\sum_{i=1}^{N}\ln\det(\Omega_{ii}) - \frac{1}{2}\sum_{i=1}^{N}\sum_t\left(\epsilon_{it}'\Omega_{ii}^{-1}\epsilon_{it}\right) = \sum_{i=1}^{N}\mathcal{L}_i(\Pi_i, \Lambda_{0i}, \Gamma_{li}, \Omega_i)\end{aligned}$$

The dimensions of the parameter matrices of the panel given in equation 6 are functions of N, but are sparse with a known sparsity pattern. Since the transformed panel by construction is not subject to residual cross section dependence, its likelihood function is the sum of the likelihood functions of the individual models. This permits independent estimation of the parameters of the individual models by maximum likelihood. The parameters of the full panel can be recovered by manipulation of the estimated parameters of the individual models. In each individual model the weighted averages Y^* are treated as weakly exogenous. However every variable is endogenous in the full panel. Furthermore it allows for immediate feedback between the variables of different units through the Λ_0 matrix.

This section described how a panel VAR subject to residual cross section dependence and where the number of cross section units N is allowed to be large can be transformed into a set of independent VAR models. Using this model, the following section will discuss an estimator and a test for a common co-integration space.

3. Common Co-integration Space Estimator

The model henceforth is the PCVAR constructed as discussed above:

$$\Delta Y_t = \alpha\beta' W Y_{t-1}, + \Lambda_0 W_0 \Delta Y_t + \sum_{l=1}^{k}\Gamma_l \Delta W Y_{t-l} + \epsilon_t \tag{9}$$

Before discussing the CCS estimator and the associated test, the following assumptions are necessary:

Assumption 2. *Assumptions on the residuals:*

i) $\epsilon_t \sim \mathcal{N}(0, \Omega)$

ii) $E(\epsilon_t, \epsilon_{t-j}) = 0 \forall j \neq 0$

iii) Ω *is block diagonal*

In assumption 2, *i*) and *ii*) ensures that the residuals are independent normally distributed with covariance Ω while *iii*) states that Ω is block diagonal. The latter condition ensures that there is no cross section dependence left in the residuals. This assumption on the structure of the covariance matrix can be tested using the test proposed in Callot (2012).

Assumption 3. *Homogeneous co-integration rank:*

The co-integration rank is the same for every individual model of the panel, that is:

$$rk(\alpha_i\beta_i') = r \ \ \forall i$$

If the number of co-integration relations isn't identical across individuals, there can be no common co-integration space. Assumption 3 ensures that this is not the case. To the best of my knowledge, this assumption cannot be tested. But since every individual model contains the same set of variables it seems to be a reasonable assumption. The co-integration rank of a panel of independent CVAR models can be estimated using the tests by Larsson et al. (2001); Pedroni (2004); Callot (2010) for example.

Assumption 4. *Stable process: define* $\Gamma := I_{Np} - \sum_{l=1}^{k}\Gamma_l W$.

i) The matrix $\alpha_\perp'\Gamma\left(W'\beta\right)_\perp$ *has full rank .*

ii) The roots of the characteristic polynomial:

$$A(z) = (1-z)I_{Np} - \alpha\beta' W z - \sum_{l=1}^{k} \Gamma_l W (1-z) z^l$$

are on or outside the unit circle.

$i)$ and $ii)$ in assumption 4 ensure that the process considered is integrated of order at most one, thus ruling out explosive processes. This is a standard assumption in co-integration analysis, it can be found for example in Johansen (1988).

The following procedure permits the estimation of a likelihood maximizing common co-integration space (CCS) estimator β_{CCS}. That is, a constrained estimator such that:

$$\hat{\beta} = \left(I_N \otimes \hat{\beta}_{CCS} \right) \tag{10}$$

Equation 10 indicates that the co-integration space of the full panel is block diagonal with every block being equal to the common co-integration space β_{CCS}, the corresponding hypothesis is noted $\mathcal{H}_{CCS}$. The CCS estimator is the solution to the following constrained optimization problem:

$$\begin{aligned} \hat{\beta}_{CCS} &= \max_{\beta_{CCS}} \sum_i \mathcal{L}_i(\beta_{CCS}) \\ &= \min_{\beta_{CCS}} \sum_i \left(\mathcal{L}_i(\hat{\beta}_i) - \mathcal{L}_i(\beta_{CCS}) \right) \end{aligned} \tag{11}$$

Where $\hat{\beta}_i$ is the unrestricted maximum likelihood estimator obtained by maximizing equation 9. Since $\mathcal{L}_i(\hat{\beta}_i)$ is the upper bound for the likelihood of the individual CVAR and $\sum_{i=1}^{N} \mathcal{L}_i(\hat{\beta}_i)$ is the upper bound for the likelihood of the panel given by equation 9, it follows that $\mathcal{L}_i(\hat{\beta}_i) \geq \mathcal{L}_i(\hat{\beta}_{CCS})$.

The optimization problem in equation 11 can be solved with standard optimization algorithm. Each co-integration vector has to be normalized with respect to a given parameter (for example the first one). We can write $\hat{\beta}_i^* = [\iota, \hat{b}_i]$, with ι being a vector of ones of length r. The probability of the estimated parameter used for normalization being exactly equal to zero is zero. It is clear that $sp(\hat{\beta}_i^*) = sp(\hat{\beta}_i)$ and therefor the optimization problem can be equivalently stated as:

$$\hat{b}_{CCS} = \min_{b_{CCS}} \sum_i \left(\mathcal{L}_i(\hat{\beta}_i^*) - \mathcal{L}_i(\beta_{CCS}^*) \right)$$

where $\hat{\beta}_{CCS}^* = [\iota, \hat{b}_{CCS}]$. This optimization problem is numerically simpler since it has $r(p-1)$ parameters against rp for (11).

An alternative method is to use the switching algorithm proposed in Larsson and Lyhagen (2007) to estimate (10). The switching algorithm was found to be slower than solving the optimization problem for large values of N, and often did not converge.

Lemma 1. *Under assumptions 2, 3 and 4, the quantity:*

$$Q_{iT} := T \left(\mathcal{L}_i(\hat{\beta}_i) - \mathcal{L}_i(\hat{\beta}_{CCS}) \right) \tag{12}$$

is the likelihood ratio test statistic of the hypothesis $\mathcal{H}_0 : \beta_i = \hat{\beta}_{CCS}$ against $\mathcal{H}_1 : \beta_i = \hat{\beta}_i$. This quantity is asymptotically (as $T \to \infty$) χ^2 distributed with degrees of freedom $m = r(2p-1)$.

The proof of this lemma can be found in the appendix. Lemma 1 gives the asymptotic distribution of the likelihood ratio test statistics for the hypothesis that the co-integration space of a given unit of the panel is equal to the estimated common co-integration space.

Remark 1. *Notice that neither the vectors of $\hat{\beta}_i$ nor those of $\hat{\beta}_{CCS}$ are identified. However the hypothesis tested here is one that involves one co-integration space against another. This hypothesis* *is independent of any particular mapping of the spaces, thus we do not need to identify the co-integration vector under the null or under the alternative.*

Lemma 1 implies that after estimating the common co-integration space we have, under the null hypothesis, a set of N independent identically distributed tests statistics. Summing these χ^2 distributed statistics is not feasible in the setting considered in the present paper since the two first moment of the sum of the individual statistics would be function of N and hence tend to infinity with N. Using a central limit theorem we can construct a panel test statistic for $\mathcal{H}_{CCS}$:

$$\bar{Q}_T = \frac{\sum_{i=1}^{N}(Q_{iT} - m)}{\sqrt{2mN}} \tag{13}$$

Equation 13 gives a first panel test statistics for a common co-integration space, its asymptotic distribution is discussed in theorem 1 below.

Theorem 1. *Under assumption 2, 3 and 4, and under the null hypothesis $\mathcal{H}_{CCS} : \beta_i = \hat{\beta}_{CCS}\ \forall i$ with $\frac{\sqrt{N}}{T} \to 0$, we have:*

$$\bar{Q}_T = \frac{\sum_{i=1}^{N}(Q_{iT} - m)}{\sqrt{2mN}} \xrightarrow[(N,T)_j]{d} \mathcal{N}(0,1)$$

Proof of theorem 1 can be found in the appendix. Theorem 1 states that the pooled likelihood ratio test statistics tends jointly in distribution to a standard normal, under the condition that $\frac{\sqrt{N}}{T} \to 0$. This condition on the joint rate of convergence indicates that N is allowed to grow fast relative to T, which means that the asymptotic distribution is likely to be a more accurate approximation of the finite sample behavior of the statistics when the number of cross section units is large. The next section introduces the bootstrap algorithm.

4. Bootstrap algorithm

Likelihood ratio test for linear restrictions on co-integration vectors are known to suffer severe size distortion in finite sample as documented by Gredenhoff and Jacobson (2001). I propose an algorithm to obtain bootstrap *p-values* for the CCS test. As noted in section 2, while each individual model is augmented with weak exogenous variables, the full panel is not conditional on any variables. Thus the full panel is bootstrapped using the following recursion:

$$\begin{aligned}
\Delta Y_t &= \alpha(I_N \otimes \beta'_{CCS})WY_{t-1} + \Lambda_0 W_0 \Delta Y_t + \sum_{l=1}^{k}\Gamma_l \Delta W Y_{t-l} + \epsilon_t^{\dagger,b} \\
(I_{Np} - \Lambda_0 W_0)Y_t &= \left(\alpha(I_N \otimes \beta'_{CCS})W - \Lambda_0 W_0 + I_{Np} + \Gamma_1 W\right) Y_{t-1} \\
&\quad + \left(\Gamma_2 W - \Gamma_1 W\right) Y_{t-2} + \cdots - \Gamma_k W Y_{t-k} + \epsilon_t^{\dagger,b} \\
Y_t &= A^{-1}\left(\alpha(I_N \otimes \beta'_{CCS})W - \Lambda_0 W_0 + I_{Np} + \Gamma_1 W\right) Y_{t-1} \\
&\quad + A^{-1}\left(\Gamma_2 W - \Gamma_1 W\right) Y_{t-2} + \cdots - A^{-1}\Gamma_k W Y_{t-k} + A^{-1}\epsilon_t^{\dagger,b}
\end{aligned} \tag{14}$$

where $A = (I_{Np} - \Lambda_0 W_0)$ is a full rank matrix. This model gives the PCVAR in levels, which will be used to generate the pseudo-data:

Algorithm 1. *Bootstraping the PCVAR.*

(1) Initialization step:
- *Estimate model 9 for all $i \in [1, \cdots, N]$ and compute the likelihood ratio test statistic Q_{iT} for $\mathcal{H}_0 : \beta_i = \hat{\beta}_{CCS}$*

- *Compute the short run parameters of the model under* $\mathcal{H}_0 : \beta_i = \widehat{\beta}_{CCS}$*:* $\widehat{\theta}_i^c = \{\widehat{\alpha}_i^c, \widehat{\lambda}_{i,0}^c, \widehat{\Gamma}_{i,l}^c\}$ *and the residuals* $\widehat{\epsilon}_{i,t}^c$*. Construct the panel parameter matrices as in equation 7 and* **??**. *Note these matrices* $\widehat{\theta}^c = \{\widehat{\alpha}^c, \widehat{\Lambda}_0^c, \widehat{\Gamma}_l^c\}$.

(2) *Construct the bootstrap residuals by resampling from the centered residuals. Note these residuals* $\widehat{\epsilon}_t^{\dagger,b}$*, where b is used to denote the bth iteration of the bootstrap algorithm.*

(3) *Use* $\widehat{\theta}^c$ *and* $\widehat{\epsilon}_t^{\dagger,b}$ *in recursion 14 to generate a sample of pseudo-data noted* Y^b.

(4) *Compute the individual bootstrap p-value for* $\mathcal{H}_0$ *by comparing the original likelihood ratio test statistics to its the empirical distribution computed under* $\mathcal{H}_{CCS}$*:*

$$p_{i,CCS}^{\dagger} = \frac{1}{B}\sum_{b=1}^{B}\mathbf{1}_{\{Q_{iT}<Q_{iT}^{\dagger,b}\}}$$

(5) *Repeat steps 2 and 3 B times.*

This algorithm yields N individual $p - values$ for the hypothesis $\beta_i = \hat{\beta}_{CCS}$. These individual *p-values* can be pooled yielding a test for $\mathcal{H}_{CCS}$: $\bar{P}_T = \frac{\sum_{i=1}^{N}\left(-2\log(p_{i,CCS}^{\dagger})-2\right)}{\sqrt{4N}}$ $\bar{P}_T$ relies on Fisher's method using the result $-2\log(p_{i,CCS}^{\dagger}) \sim \chi^2(2)$. The theorem below shows that both statistics are asymptotically normal.

Theorem 2. *Under assumption 2, 3 and 4 and under* $\mathcal{H}_{CCS}$*, the asymptotic distribution of the* $\bar{P}_T$ *statistic is given by:* $\bar{P}_T \xrightarrow[(T,N)_j]{d} \mathcal{N}(0,1)$

The statistics in theorem 2 converges to a standard normal distribution when N and T are allowed to increase jointly without conditions on the relative rate of convergence.

5. Identification

In the co-integration model, the parameters of the matrix $\alpha\beta'$ are not identified in the sense that one can find a full rank $r \times r$ matrix A such that:

$$\Delta Y_t \quad = \quad \alpha\beta' Y_{t-1} + \epsilon_t = ab' Y_{t-1} + \epsilon_t$$

where $a = \alpha A^{-1'}$ and $b = \beta A$. In this case the two models above are observationally equivalent, in particular $\mathcal{L}(\alpha, \beta) = \mathcal{L}(a, b)$. To identify α and β, one need to find restrictions such that if α, a, β and b satisfy these restrictions, then $A = \mathtt{I_r}$. The identification problem in the standard co-integrated VAR is discussed in much details in Johansen (1995a, 2010). One simple method to obtain an exactly identified co-integration space is to orthogonalize the vectors of the co-integration space with respect to the first r variables of Y_t, so that $\beta' = (I_r, b)$. However such an identification scheme is not always economically meaningful. A wide range of linear restrictions can be imposed in order to obtain an identified model. Johansen (1995a) derives a set of algebraic conditions that are necessary and sufficient to ensure that a set of linear restrictions is identifying.

Lemma 2. *Let* $[H_1, \cdots, H_r]$ *be a set of linear restrictions such that* $H_{k\perp}\beta_k = 0 \ \ \forall k \in 1, \cdots, r$. β_k *is said to be identified if and only if for any* $h = 1, \cdots, r-1$ *and any g indices* $1 \leq k_1 < k_2 < \cdots < k_g \leq r$ *so that* $k_l \neq k \ \forall l \in (1, \cdots, g)$ *we have:*

$$rank \quad \left(H'_{\perp k}\left(H_{k_1}, \cdots, H_{k_g}\right) \geq g \tag{15}$$

The condition above is called the deficient rank condition. This lemma was first stated as Theorem 3 in Johansen (1995a), and a proof can be found therein.

In the case of the PCVAR considered in this paper, the co-integration space is very large (Nr vectors) which would require finding Nr linear restrictions that all satisfy the deficient rank condition. The problem can however be simplified by considering two distinct identification criteria.

i) *within* identification. This criterion is concerned with identifying the vectors of a given individual co-integration space β_i. This is the classical identification problem in the sense of Johansen (2010), which requires r linear restriction matrices satisfying the deficient rank condition so that:

$$\mathcal{L}_i(\beta_i) = \mathcal{L}_i(\beta_i^*) \Rightarrow \beta_i = \beta_i^*$$

ii) *between* identification. This criterion is concerned with identifying the individual co-integration spaces with respect to one another. This requires by finding N linear restriction matrices that satisfy the deficient rank condition so that:

$$\mathcal{L}(\beta_1, \cdots, \beta_N) = \mathcal{L}(\beta_1^*, \cdots, \beta_N^*) \Rightarrow sp(\beta_i) = sp(\beta_i^*)\, \forall i$$

These two identification criteria need to be satisfied in order to obtain an identified co-integration space for the full panel.

In this section I derive the conditions for identification of the co-integration space of the panel as a whole. I show that if the rank for each individual model is 1, the linear restrictions implied by the W_i matrices are sufficient to ensure identification of the panel's co-integration space for any $N \geq 3$. If the individual rank is greater than 1 and if a set of identifying restrictions is imposed on the co-integration vectors of each unit, then the co-integration space of the full panel is identified. It becomes possible to compute standard errors for the parameters of the identified co-integration.

Recall the individual model given in equation 3 without short run dynamics nor deterministic components:

$$\Delta Y_{it} = \alpha_i \beta_i' W_i Y_{t-1} + \epsilon_{it}$$

This notation highlights the fact that the weighting matrix W_i is a matrix of linear restrictions on the co-integration space. It reduces the number of free parameters of each co-integration vector from Np to $2p$. If further linear restrictions are imposed on the co-integration vectors the model can be reformulated as:

$$\begin{aligned}
\Delta Y_{it} &= \alpha_{i1}\phi_{i1}' H_1' W_i Y_{t-1} + \cdots + \alpha_{ir}\phi_{ir}' H_r' W_i Y_{t-1} + \epsilon_{it} \\
\Delta Y_{it} &= \alpha_{i1}\phi_{i1}' \tilde{H}_{i1} + \cdots + \alpha_{ir}\phi_{ir}' \tilde{H}_{ir} Y_{t-1} + \epsilon_{it}
\end{aligned}$$

where $\widetilde{H}_{ik}$ is the restriction imposed on the k^{th} co-integration vector of unit i and ϕ_{ik} the corresponding free parameters.

Now consider the co-integration space of the full panel β and the set of restriction $\left[\tilde{H}_{11}, \cdots, \tilde{H}_{Nr}\right]$. For the first co-integration vector of the first individual of the full panel to be identified, the following conditions (derived from lemma 2) must hold:

$$\begin{aligned}
rank \quad & (\tilde{H}_{\perp ik}' \tilde{H}_{jh}) \geq 1 \text{ for } ik \neq jh \\
rank \quad & (\tilde{H}_{\perp ik}' (\tilde{H}_{j_1 h_1}, \tilde{H}_{j_2 h_2})) \geq 2 \text{ for } ik \neq j_1 h_1 \neq j_2 h_2 \\
\vdots \\
rank \quad & (\tilde{H}_{\perp ik}' (\tilde{H}_{11}, \cdots, \tilde{H}_{ik-1}, \tilde{H}_{ik+1}, \cdots, \tilde{H}_{Nr}) \geq Nr - 1
\end{aligned}$$

with $i, j \in 1, \cdots, N$ and $k, h \in 1, \cdots, r$. These conditions do not depend on the unknown parameter and are only algebraic conditions for generic identification. If these conditions are satisfied the model is said to be generally identified in the sense that it is identified except for a set of parameters of Lebesgue measure zero. However, finding Nr linear restrictions satisfying the deficient rank condition can be a tedious exercise. Theorem 3 simplifies the problem by showing that the *between* identification criterion is satisfied by construction of the panel for $N \geq 3$ and $p \geq 2$.

Theorem 3. *Let $N \geq 3$ and $p \geq 2$, then the linear restrictions imposed by the W_i matrices as defined in equation 4 are sufficient to ensure identification* between *the individual co-integration spaces so that:*

$$\mathcal{L}(\beta_1, \cdots, \beta_N) = \mathcal{L}(\beta_1^*, \cdots, \beta_N^*) \Rightarrow sp(\beta_i) = sp(\beta_i^*) \, \forall i$$

The proof can be found in the appendix. Theorem 3 states that the conditions for *between* identification are satisfied by construction of the PCVAR model. This simplifies considerably the task of identifying the co-integration space of the panel since only r linear restrictions must be imposed on each β_i in order for the Nr co-integration vectors of the panel to be identified. As a special case, notice that if $r = 1$ the N co-integration vectors of the panel are identified up to a normalization by construction. Theorem 6 in Johansen (2010) can then be applied to compute the asymptotic distribution of $Tvec\left(\hat{\beta} - \beta\right)$, which is mixed Gaussian and of $\sqrt{T}vec(\hat{\alpha} - \alpha)$, which is Gaussian.

6. Monte Carlo

To investigate how the tests for the common co-integration space hypothesis ($\mathcal{H}_{CCS}$) perform in finite samples, and assess how much improvement (if any) the use of the bootstrap provides, a Monte Carlo experiment is conducted. In this section the performances of the tests in terms of size are investigated for different combination of cross section dimension N, number of observations T and co-integration rank r. Experiments using heterogenous co-integration spaces are also considered to evaluate the power of the tests. The Y_t^* variables are constructed using uniform weights, that is, $w_{ij} = 1/(N-1) \;\; i \neq j$. Simulations are carried out using `Ox 6.1`. The DGP used are of the form

$$Y_t \;\; = \;\; A^{-1}\left(\alpha(I_N \otimes \beta_{CCS}')W - \Lambda_0 W_0 + I_{Np} + \Gamma_1 W\right) Y_{t-1} - A^{-1}\Gamma_1 W Y_{t-2} + A^{-1}\epsilon_t^{\dagger,b}$$

Experiment A is the simplest possible scenario with 2 variables per individual model ($p = 2$), and the two corresponding weighted averages, and a single co-integration vector $\beta_{CCS} = [1, -1, 0, 0] \;\; \forall i$ that only allows for co-integration among variables of the same unit. The adjustment parameters are set to $\alpha_i = [-0.4, 0.4]$, and the covariance matrix of the panel residuals is $\Omega = \mathtt{I_{Np}}$. In this first experiment, all short run dynamics and deterministic components are left out. This DGP does not allow for any interaction between the variables of different units of the panel.

Experiment B still considers a 2 variable system, but this time allowing for co-integration between Y_i and Y_i^*. The rank of each individual model is equal to two, with the co-integration vectors and adjustment matrix being:

$$\beta_{CCS} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix} \alpha_i = \begin{bmatrix} -0.4 & 0.4 \\ -0.4 & 0 \end{bmatrix}$$

The covariance matrix of the residuals is as before. In this DGP the first variable of the system co-integrates with the second one, and also with the first weighted average Y^*, and adjust to disequilibrium in both equations while the second variable only respond to disequilibriums in the first co-integration vector.

The individual test statistics in Experiment A are both very close to the nominal size of 5%, while being undersized in some instances where T is large. The asymtoptic test based on the pooled Q_i statistics is also undersized when N is small relative to T, but less so than the $\bar{Q}$ statistics, which is strongly undersized when N is small. The $p - value$ based statistics have sizes close to 0.05, the logarithmique statistic performing better when N is large. The central limit theorem using bootstrap moment estimators statistics isn't strongly undersized in very small samples, and outperforms the other panel statistics when N increases.

A slight modification of the DGP of experiment A is introduced to explore the power properties of the tests. The co-integration matrix for each unit is generated with some idiosynchratic uniformly distributed noise on the parameters. The uniform distribution is chosen to ensure that the signs

		Individual tests		Panel tests	
N	T	Bootstrap	Asymptotic	$\bar{Q}_T$	$\bar{P}_T$
10	100	0.0404	0.0472	0.0570	0.0620
10	200	0.0377	0.0403	0.0450	0.0450
10	1000	0.0371	0.0351	0.0310	0.0340
20	100	0.0499	0.0587	0.0920	0.0950
20	200	0.0456	0.0487	0.0520	0.0570
20	1000	0.0428	0.0409	0.0320	0.0320
50	100	0.0554	0.0657	0.2000	0.1970
50	200	0.0523	0.0544	0.0960	0.0930
100	100	0.0568	0.0693	0.3550	0.3620
100	200	0.0538	0.0580	0.1440	0.1470

TABLE 1. Experiment A, 1000 Monte Carlo replications, 199 Bootstrap iterations.

of the parameters affected by noise will not revert, thus ensuring that the resulting DGP is error correcting. More precisely:

$$\beta_i = [1, -1, 0, 0] + \delta\nu_i$$

Where ν_i is a matrix of random numbers uniformly distributed on $[0, 1]$ and δ a scaling parameter such that $\mathcal{H}_{CCS}$ is true when $\delta = 0$. The power curves of the test statistics are plotted in figure 1 for different value of the disturbance scale δ, using $T = 100$, $N = 10$ and $N = 50$, and separating the individual statistics (left panels) and the pooled statistics (right panels). Notice the first the difference in scale between the left and right panels, the panel statistics have much better power than their individual counterparts. However in this setting the bootstrap doesn't seem to improve upon tests based on the asymptotic distribution of the statisitics considered.

		Individual tests		Panel tests	
N	T	Bootstrap	Asymptotic	$\bar{Q}_T$	$\bar{P}_T$
10	100	0.0167	0.0466	0.0490	0.0500
10	200	0.0160	0.0388	0.0430	0.0430
10	1000	0.0157	0.0307	0.0240	0.0260
20	100	0.0289	0.0626	0.1070	0.1140
20	200	0.0276	0.0513	0.0740	0.0770
20	1000	0.0249	0.0414	0.0300	0.0300
50	100	0.0361	0.0692	0.2340	0.2420
50	200	0.0352	0.0570	0.1160	0.1170
100	100	0.0401	0.0728	0.4500	0.4560
100	200	0.0379	0.0604	0.2060	0.2000

TABLE 2. Experiment B, 1000 Monte Carlo replications, 199 Bootstrap iterations.

Experiment B displays some stricking features: The bootstrap individual test is strongly undersized as are for the most part the panel statistics. The logarithmic pooled $p - value$ and the asymptotic normal statistics seems to be the less undersized, while being mildly oversized in large N sample.

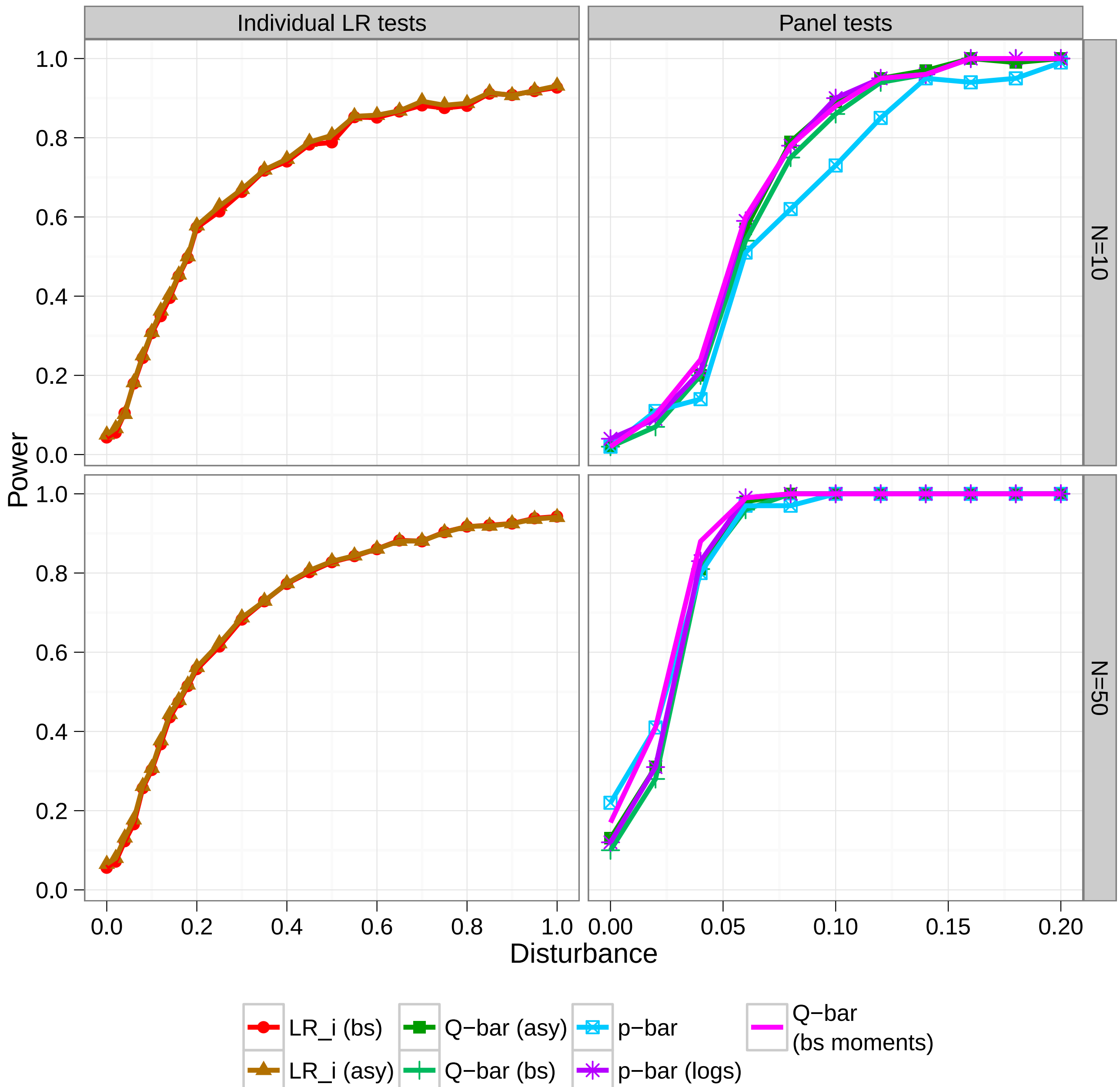

FIGURE 1. Power, experiment A with uniform noise on the parameters

7. EMPIRICAL APPLICATION

The empirical application presented in this section is concerned with interest rate dynamics in the Euro zone. The data set used here is identical to that used in Callot (2010), it consists of series for the 3-month interest rate (noted r_i) and 10-years treasury bond rates (b_i) as well as inflation (changes in the harmonized consumer price index, noted π_i) for 10 founding Euro zone countries. Both interest rates are expressed in monthly returns. The data is observed monthly from January 1995 to December 2011, and was retrieved from the OECD statistical database in August 2012. The weights used to construct the cross section averages are based on the ECB's *weight of the countries within the euro area* series.

After the creation of a currency union was enacted, interest rates in the countries that would constitute the Euro zone started to converge. However inflation differentials remained, and thus real interest rates did not converge to the same extent nominal rates did. Towards the end of the sample, and as a results of the turmoil affecting the Euro zone, long term interests rates started diverging again. In this application we are interested in testing whether despite the different dynamics observed in the data the long run relations between interest rates of different maturities and inflation were identical among those 10 countries.

The first system considered is composed of the long and short interest rate as well as the inflation rate, augmented by the weighted averages of each of these variables. The model is estimated with 1 lagged first difference and a restricted constant. The cointegration rank of this model was investigated in Callot (2010) and found to be 2.

The estimated common cointegration space for this system is:

$$\hat{\beta}_{CCS,1} : r_i - 1.24b_i + 117\pi_i - 0.969r_i^* + 1.28b_i^* - 108\pi_i^* - 0.0249\rho$$
$$\hat{\beta}_{CCS,2} : -11.5r_i + b_i + 4749\pi_i + 11.3r_i^* - 3.51b_i^* - 3592\pi_i^* + 0.0575\rho$$

In both equations the parameters on any given domestic variable and its corresponding weighted average are strickingly similar, with opposite signs, indicating that some equilibrium relation exists among these variables. The second equation seems to be dominated by a relation between domestic and foreign inflation, while the first one seems to describe an equilibrium relation between domestic and foreign interests rates (long and short). However since these parameters are not identified it is not possible to give a structural interpretation to these vectors.

	Q_{iT}	$p(Q_{iT})$	$\frac{1}{B}\sum_{b=1}^{B} Q_{iT}^{\dagger,b}$	$p^{\dagger}(Q_{iT})$
Austria	5.822	0.925	21.885	0.996
Belgium	34.141	0.001	23.730	0.140
Finland	18.935	0.090	33.269	0.742
France	31.947	0.001	38.459	0.586
Germany	21.093	0.049	22.656	0.516
Ireland	21.009	0.050	34.404	0.884
Italy	25.044	0.015	59.400	0.998
Netherlands	6.773	0.872	19.379	0.962
Portugal	19.004	0.088	8.4491	0.026
Spain	21.465	0.044	18.613	0.306

	Panel tests	
	Statistic	$p-value$
$\bar{Q}_T$	5.502	0.000
$\bar{P}_T$	-0.47	0.687

TABLE 3. Individual and pooled Likelihood ratio statistics, with asymptotic and bootstrap $p-values$, 499 bootstrap replications

The upper panel of Table 3 reports the results for the individual likelihood ratio tests. The first column reports the likelihood ratio statistics and the third the average bootstrap likelihood ratio statistics. Column 2 reports the asymptotic $p - value$ corresponding to Q_{iT} and column 4 the bootstrap $p - value$. At a 5% significance level, the asymptotic test rejects the hypothesis $\beta_i = \hat{\beta}_{CCS}$ in 5 out of 10 instances, but only once for the bootstrap test. Indeed, the likelihood ratio statistics are often found to be substantially higher than their expected value of 12. However the average bootstrap statistics are also found to be higher than 12 indicating that the asymptotic test is distorted in finite samples.

The lower panel of Table 3 reports the results of both pooled tests for the common cointegration space hypothesis. The asymptotic test rejects the hypothesis, which is not surprising given the results from the individual tests showing that the likelihood ratio statistics are often larger than their expected values. However the bootstrap version of the test does not reject the hypothesis of a common cointegration space.

We now turn to a second system where both nominal interest rates have been transformed to monthly real interest rates by subtracting monthly inflation to the monthly interest rate. Short and long interest rates are noted rr_i and rb_i, respectively. This system was investigated in Callot (2010) and its cointegration rank was found to be 1. Using the same 10 countries as previously, the common cointegration space hypothesis is strongly rejected as shown in Table 4.

	Q_{iT}	$p(Q_{iT})$	$\frac{1}{B}\sum_{b=1}^{B} Q_{iT}^{\dagger,b}$	$p^{\dagger}(Q_{iT})$
Austria	3.5078	0.476	12.66	1.00
Belgium	4.6507	0.325	11.16	0.996
Finland	7.7835	0.099	8.547	0.550
France	4.5098	0.341	26.23	1.00
Germany	21.104	0.000	14.14	0.122
Ireland	23.856	0.000	6.311	0.004
Italy	6.8447	0.144	11.73	0.724
Netherlands	10.427	0.034	3.563	0.010
Portugal	37.787	0.000	3.716	0.002
Spain	36.151	0.000	5.796	0.002
		Panel tests		
		Statistic	$p - value$	
$\bar{Q}_T$		13.039	0.000	
$\bar{P}_T$		4.928	0.000	

TABLE 4. Individual and panel Likelihood ratio statistics, with asymptotic and bootstrap $p - values$, second system, 499 bootstrap replications

The CCS procedure is iterated removing from the panel the unit with the highest likelihood ratio statistics (in the first step Portugal) until a panel supporting a common cointegration space is found. After 3 steps, removing Portugal, Spain, and Ireland, in that order, such a panel is uncovered. The results are reported in Table 5. The common cointegration space corresponding to this panel is:

$$\hat{\beta}_{CCS} : rr_i + 3.07rb_i - 1.34rr_i^* - 2.21rb_i^* - 0.17\rho$$

The parameters of this space are identified by the normalization on the domestic short real interest rate variable, allowing for a structural interpretation. This vector seems to describes a stationary relation between the spread between domestic and foreign short real interest rates, and the spread between domestic and foreign long real interest rates. Further restriction testing would be required to confirm this interpretation.

	Q_{iT}	$p(Q_{iT})$	$\frac{1}{B}\sum_{b=1}^{B} Q_{iT}^{\dagger,b}$	$p^{\dagger}(Q_{iT})$
Austria	2.179	0.702	6.568	1.00
Belgium	0.2442	0.993	6.458	1.00
Finland	16.08	0.003	5.127	0.010
France	2.202	0.698	9.209	1.00
Germany	14.65	0.005	6.661	0.030
Italy	4.237	0.374	9.774	0.838
Netherlands	3.532	0.472	2.541	0.226
		Panel tests		
		Statistic	$p-value$	
$\bar{Q}_T$		2.023	0.022	
$\bar{P}_T$		1.049	0.147	

TABLE 5. Individual and panel Likelihood ratio statistics, with asymptotic and bootstrap $p-values$, second system reduced, 499 bootstrap replications

8. Conclusion

This paper builds on the work of Pesaran (2006) and Dees et al. (2007) to reduce the dimension of the parameter space of a large N large T panel of vector autoregressions and ensure that the residuals are cross-sectionally independent. The model so constructed has some interesting properties: each individual model can be estimated independently and the number of parameters in the individual model is not a function of N, which enables the estimation of very large panels. The full panel model can be reconstructed from the individual models and it allows for a dependency among all the variables in the panel, both in the long and short run, and contemporaneous as well as lagged. Test statistics derived from the individual models are independent so that panel test statistics can be constructed by pooling of the individual tests.

Using this model I proposes and estimator and a test for a common co-integration space. Asymptotic distribution of the panel test statistics is derived, and a bootstrap algorithm is proposed to obtain bootstrap $p-values$ for the hypothesis of a common co-integration space. I also show that identification of the co-integration vectors of the panel is quite simple is this model since the restrictions imposed by construction on the dynamics of the model ensure that the individual co-integration spaces are identified with respect to one another.

A Monte Carlo simulation is carried to document the finite sample properties of the common co-integration space tests, and it reveals that the test statistics proposed have sizes close to their nominal value of 5%.

Appendix A. Proofs

Proof of lemma 1. Write $\beta_{CCS} = b_{CCS}\phi$ where b_{CCS} is the normalized form of β_{CCS} and ϕ is a vector of length r of free parameters. The distribution of likelihood ratio tests for linear restrictions is discussed in Johansen (1991). The asymptotic distribution of such tests is shown to be χ^2 with a number of degrees of freedom equal to the difference in the number of free parameters under both hypothesis. Since $\mathcal{H}_A\beta_i = \hat{\beta}_i$ doesn't impose any restrictions, there are $2pr$ free parameters, while under $\mathcal{H}_0\beta_i = \hat{\beta}_{CCS}$ there are only r free parameters. Hence

$$LR_{CCS} = T\left(\sum_i \left(\mathcal{L}_i^{max} - \mathcal{L}_{CCS,i}^{max}(\hat{\beta}_{CCS})\right)\right) \xrightarrow[T]{w} \chi^2\left(r\left(2p-1\right)\right)$$

□

Before proving theorem 1 the following lemma is required:

Lemma 3. *Under assumptions 2, 3 and 4, the following holds:*

$$Q_{iT} = Q_{i\infty} + O_p(T^{-1})$$

Where $Q_{i\infty}$ is the asymptotic test statistics which is $\chi^2(m)$ distributed.

Proof of theorem 1. Using Lemma 3, the $\bar{Q}_T$ statistics can be rewritten as:

$$\begin{aligned}
\bar{Q}_T &= \frac{\sum_{i=1}^{N}(Q_{iT} - m)}{\sqrt{2mN}} \\
&= \frac{\sum_{i=1}^{N}(Q_{i\infty} - m) + NO_p\left(T^{-1}\right)}{\sqrt{2mN}} \\
&= \frac{\sum_{i=1}^{N}(Q_{i\infty} - m)}{\sqrt{2mN}} + \frac{\sqrt{N}O_p\left(T^{-1}\right)}{\sqrt{2m}}
\end{aligned}$$

It is clear that if $\frac{\sqrt{N}}{T} \to 0$ the second term of the right hand side of the above equation disappears. To show that the first term of the right hand side converges jointly in (N, T) to a standard normal distribution, the Lyapunov conditions have to be verified (see Davidson (1994) page 372):

i) $\frac{1}{N}\sum_{i=1}^{N} var(Q_{i\infty}) > 0$ uniformly in N. Since $var(Q_{i\infty}) = 2m$, this condition is verified.

ii) $E\left(|Q_{i\infty}|^{2+\delta}\right) < \infty$ for some $\delta > 0$. Notice that $Q_{i\infty} > 0$, and choose $\delta = 1$: $E\left(Q_{i\infty}^3\right) = \sqrt{8/m} < \infty$

Since the Lyapunov conditions are verified, the Lindenberg central limit theorem applies and:

$$\bar{Q}_T \xrightarrow[d]{(N,T)_j} \mathcal{N}(0,1)$$

□

Proof of theorem 2. Under $\mathcal{H}_{\prime} : \beta_i = \hat{\beta}_{CCS}$ we have:

a) $p^{\dagger}_{i,CCS} \sim \mathcal{U}[0, 1]$

b) $-2\log p^{\dagger}_{i,CCS} \sim \chi^2(2)$

Hence under $\mathcal{H}_{CCS}$:

$$\bar{P}_{CCS} \xrightarrow[N]{d} \mathcal{N}(0,1)$$

Since the convergence results above holds for all T, the Lyapounov conditions are verified hance the Lindenberg central limit theorem applies.

□

Proof of theorem 3. I will start showing that the W_i matrices are sufficient for identification between the co-integration spaces first in the case where $N = 3$, and then the general case. The proof uses the algebraic conditions by Johansen (1995b) discussed above.

- $N = 3$: The weigthing matrix W_i is imposed β_i.

$$\Delta Y_{it} = \alpha_i\beta_i' W_i Y_{t-1}$$

For simplicity of notations, and without loss of generality, I will assume the weights are the same for every individual ($w_{ij} = w_{kj} \forall i, k$)

$$W_1 = \begin{bmatrix} \mathtt{I}_p & \mathtt{0}_p & \mathtt{0}_p \\ \mathtt{0}_p & w_2\mathtt{I}_p & w_3\mathtt{I}_p \end{bmatrix}$$

The orthogonal complement is:

$$W_{1\perp} = \begin{bmatrix} \mathtt{0}_p & -w_3\mathtt{I}_p & w_2\mathtt{I}_p \end{bmatrix}$$

$$\begin{aligned} rank\left(W_{1\perp}W_2'\right) &= rk\left(\begin{bmatrix} -w_3\mathtt{I}_p & w_3w_2\mathtt{I}_p \end{bmatrix}\right) \\ &= p \\ rank\left(W_{1\perp}\left(W_2'W_3'\right)\right) &= rk\left(\begin{bmatrix} -w_3\mathtt{I}_p & w_3w_2\mathtt{I}_p & w_2\mathtt{I}_p & -w_3w_2\mathtt{I}_p \end{bmatrix}\right) \\ &= p \end{aligned}$$

Thus if $p \geq 2$ identification is ensured in this case simply by the restrictions imposed by the W_i matrices.

- $N > 3$: $W_{i\perp}$ is a matrix of dimension $[(N-2)p \times Np]$ composed of blocks of $\mathtt{0}_p$ and blocks proportional to $\mathtt{I}_p$:

$$W_{1\perp} = \begin{bmatrix} \mathtt{0}_p & -w_3\mathtt{I}_p & w_2\mathtt{I}_p & \mathtt{0}_p & \cdots & \mathtt{0}_p \\ \mathtt{0}_p & -w_4\mathtt{I}_p & \mathtt{0}_p & w_2\mathtt{I}_p & \cdots & \mathtt{0}_p \\ \vdots & & & & \ddots & \vdots \\ \mathtt{0}_p & -w_N\mathtt{I}_p & \mathtt{0}_p & \mathtt{0}_p & \cdots & w_2\mathtt{I}_p \end{bmatrix}$$

It follows that $W_{i\perp}W_j$ is a matrix of dimension $[(N-2)p \times 2p]$ also composed of identity blocks multiplied by a scalar. For $n < N$, the product matrix $W_{i\perp}(W_{j1}', \cdots, W_{jn}')$, $j_g \neq i \ \ \forall g \in 1, \cdots, n$, is a matrix of dimension $[(N-2)p \times 2np]$ with scaled identity blocks and zero blocks. As an illustration consider $W_{1\perp}W_2'$, $W_{1\perp}W_3'$ and $W_{1\perp}W_N'$:

$$W_{1\perp}W_2' = \begin{bmatrix} -w_3\mathtt{I}_p & w_2w_3\mathtt{I}_p \\ -w_4\mathtt{I}_p & w_2w_4\mathtt{I}_p \\ \vdots & \vdots \\ -w_N\mathtt{I}_p & w_2w_N\mathtt{I}_p \end{bmatrix}$$

$$W_{1\perp}W_3' = \begin{bmatrix} -w_2\mathtt{I}_p & -w_2w_3\mathtt{I}_p \\ \mathtt{0}_p & \mathtt{0}_p \\ \vdots & \vdots \\ \mathtt{0}_p & \mathtt{0}_p \end{bmatrix} W_{1\perp}W_N' = \begin{bmatrix} \mathtt{0}_p & \mathtt{0}_p \\ \vdots & \vdots \\ \mathtt{0}_p & \mathtt{0}_p \\ -w_2\mathtt{I}_p & -w_2w_N\mathtt{I}_p \end{bmatrix}$$

The structure of the product matrices are similar for any combination of i and j_g, $i \neq j_g$. It follows that:

$$rank\left(W_{i\perp}(W_{j1}', \cdots, W_{jn}')\right) = min\left((N-2)p, np\right)$$

The rank condition for identification is satisfied if

$$min\left((N-2)\,p, np\right) \geq n$$

This condition is satisfied if $N \geq 3$ and $p \geq 2$

□

References

Bai, J. and S. Ng (2005): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72, 1127–1177.

Banerjee, A. (1999): "Panel Data Unit Roots and Cointegration: An Overview," *Oxford Bulletin of Economics and Statistics*, 61, 607–629.

Breitung, J. and M. Pesaran (2008): "Unit roots and cointegration in panels," *The Econometrics of Panel Data*, 279–322.

Callot, L. (2010): "A Bootstrap Cointegration Rank Test for Panels of VAR Models," *CREATES Working Papers 2010-75.*

——— (2012): "A simple test for block diagonality of the covariance matrix in Panel VAR models," *Unpublished manuscript.*

Cavaliere, G., A. Rahbek, and A. Taylor (2010): "Bootstrap Sequential Determination of the Co-Integration Rank in VAR Models," *CREATES working paper.*

Chudik, A. and M. Pesaran (2011): "Infinite-dimensional VARs and factor models," *Journal of Econometrics*, 163, 4–22.

Davidson, J. (1994): *Stochastic limit theory: An introduction for econometricians*, Oxford University Press, USA.

Dees, S., F. Mauro, M. Pesaran, and L. Smith (2007): "Exploring the international linkages of the euro area: a global VAR analysis," *Journal of Applied Econometrics*, 22, 1–38.

Gredenhoff, M. and T. Jacobson (2001): "Bootstrap Testing Linear Restrictions on Cointergrating Vectors," *Journal of Business and Economic Statistics*, 19, 63–72.

Groen, J. and F. Kleibergen (2003): "Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models," *Journal of Business & Economic Statistics*, 21, 295–318.

Johansen, S. (1988): "Statistical analysis of cointegration vectors," *Journal of economic dynamics and control*, 12, 231–254.

——— (1991): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica: Journal of the Econometric Society*, 59, 1551–1580.

——— (1995a): "Identifying restrictions of linear equations with applications to simultaneous equations and cointegration* 1," *Journal of econometrics*, 69, 111–132.

——— (1995b): *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford University Press, USA.

——— (2010): "Some identification problems in the cointegrated vector autoregressive model," *Journal of Econometrics*, 158, 262–273.

Larsson, R. and J. Lyhagen (2007): "Inference in panel cointegration models with long panels," *Journal of Business and Economic Statistics*, 25, 473–483.

Larsson, R., J. Lyhagen, and M. Lothgren (2001): "Likelihood-based cointegration tests in heterogeneous panels," *The Econometrics Journal*, 4, 109–142.

Pedroni, P. (2004): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests With an Application to the PPP Hypothesis," *Econometric Theory*, 20, 597–625.

Pesaran, M. (2006): "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica*, 74, 967–1012.

Phillips, P. and H. Moon (1999): "Linear regression limit theory for nonstationary panel data," *Econometrica*, 67, 1057–1111.

Westerlund, J. and W. Hess (2011): "A new poolability test for cointegrated panels," *Journal of Applied Econometrics.*

Westerlund, J. and R. Larsson (2009): "A note on the pooling of individual PANIC unit root tests," *Econometric Theory*, 25, 1851–1868.