Evaluating Restricted Common Factor models for non-stationary data

Francesca Di Iorio
University of Naples Federico II

Stefano Fachin
"Sapienza" University of Rome

Abstract

We propose to evaluate exclusion or homogeneity restrictions on the loadings in non-stationary factor models on the basis of the number of factors estimated for the partially data de-factored under the null hypothesis, with the probability of rejecting true null hypothesis estimated by the bootstrap. Simulation results suggest the procedure has good properties and may thus be a valuable tool for applied factor modelling of non-stationary data.

Keywords: Approximate factor model, loadings restrictions, principal components, large data sets, bootstrap.

JEL codes: C12, C33, C55

January 2016

1 Introduction[1]

Consider a panel of N non-stationary time series of length T, for instance observations of the same variable Y over different units. An effective way of synthesizing the information contained in this dataset is postulating that long-run growth is driven in all units by a small number of latent non-stationary factors (F_j, $j = 1, \ldots, r$), while idiosyncratic shocks (ε_{ti},$t = 1, \ldots, T$, $i = 1, \ldots, N$) are weakly dependent over both time and units. We thus obtain the approximate factor model:

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\Lambda} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{Y}$ is the $T \times N$ data matrix, $\mathbf{F}$ is the $T \times r$ matrix of latent factors, $\boldsymbol{\Lambda}$ is the $r \times N$ matrix of factor loadings and $\boldsymbol{\varepsilon}$ the $T \times N$ error matrix[2]. Bai (2004, henceforth simply Bai) showed that under the very general assumptions made above on the orders of integration of factors and disturbances the space spanned by the factors and their number can be estimated consistently using Principal Components (PC).

Interestingly, the already large and rapidly growing literature on this class of models deals essentially with the problems of selecting the number of factors and estimating factors and loadings[3], while hypothesis testing on the loadings has not received much attention. However, restrictions of exclusion (a given unit is not affected by a given factor) and homogeneity (all units in a given group are equally affected by a given factor) may be interesting either because implied by economic theory or empirically relevant. For instance, Reis and Watson (2010) studied 214 US price series using an economic model which includes a "pure inflation" factor, assumed to be transmitted with constant loadings on all prices of the dataset. This restriction was tested by means of t-tests in separate regressions of each series on the estimated factor. As Amengual and Repetto (2014) point out, this solution ignores the consequences of cross-section correlation. To this we may also add that, ignoring the multiple-testing set-up, the resulting p-values are individidually compared with standard critical values, 5% and 1%, thus losing control of the family wise error rate. Amengual and Repetto (2014) proposed instead to use a joint $F-$type test comparing unrestricted and restricted sum of squares, and derived the asymptotic distribution under the fully general assumption that the number of restrictions grows with the sample size. However, their results assume stationarity, which leaves open the problem for the empirically relevant non-stationary case of interest here. For instance, in previous work (Ciccarelli and Fachin, 2016) one of the authors of this paper estimated a factor model for a set of 19th century value added series for the Italian economy disaggregated by industry and region. Since the series are non-stationary the factor analysis has been carried out applying Bai's PC method. The procedure selected two non-stationary factors, which can be essentially described as a growing trend and a Kondratieff cycle with a period of about 25 years. The loadings of the trending factor contain especially interesting information, as they summarise the role played by the various regions and industries in a crucial phase of Italian economic development. Without entering into much detail, the key point is that the leading role of the North-Western regions stands out clearly, and so does the tendency of the South to lag behind. However, within each of the four macroareas typically used to partition the country (NW, NE, Centre, South) the loadings vary considerably across

[1] Correspondance to stefano.fachin@uniroma1.it. All computations have been carried out using programs written in Hansl, the programming language of the free software package Gretl (http://gretl.sourceforge.net/). Many thanks to Umberto Ferraro Petrillo for fundamental computing help.

[2] This model can be easily generalised to allow for a dynamic structure and stationary common factors, but here for convenience of exposition and with no loss of generality we will consider only its simpler form (1).

[3] See Stock and Watson (2011) for a general review and Banerjee, Marcellino and Masten (2015) for developments of non-stationary modelling.

regions and, most importantly, industries. Then, an important question is the following: can the loadings be constrained to be constant for all industries and regions in the same macroarea? The answer may considerably enhance our understanding of the geography of early Italian industrial development, as a positive one would imply that geographic differences mattered more than the industry mix. We propose to evaluate restrictions of this type on the basis of the number of factors estimated for the data partially de-factored under the null hypothesis, with the probability of rejecting true null hypothesis estimated by the bootstrap. We now (section 2) proceed to outline the proposed procedure, then report the results of some simulations (section 3) and an empirical illustration (section 4). Section 5 concludes.

2 Evaluating restrictions on factor loadings using model selection criteria

Suppose that the approximate factor model (1) has been estimated for the data $\mathbf{Y}$ following Bai's procedure. Model selection using the three information criteria (IPC_1, IPC_2, IPC_3, defined in eq. (12), p. 145) with a maximum number of factors $max(k) > r$ suggested $\widehat{k}$ factors. From Bai's theorem 1 we then know that $lim_{N,T\to\infty} P(\widehat{k} = r) = 1$. Suppose also for convenience that we are especially interested in the first factor, $\mathbf{F}_1$ (*i.e.*, that associated to the highest eigenvalue), and more precisely in testing homogeneity of its loadings within M groups of units. Accordingly, we write the model for unit i at time t as

$$y_{ti} = \lambda_{1i} F_{t1} + \sum_{j=2}^{\widehat{k}} \lambda_{ji} F_{tj} + \varepsilon_{it} \tag{2}$$

and the restriction as

$$H_0 : \lambda_{1i} = \lambda_{1G_j},\ \forall i \in G_j,\ j = 1, \ldots, M \tag{3}$$

Then the restricted factor model for unit i at time t is

$$y_{ti} = \lambda_{1G_j} F^0_{t1} + \sum_{j=2}^{\widehat{k}} \lambda_{ji} F_{tj} + \varepsilon_{0ti}, i \in G_j,\ j = 1, \ldots, M$$

In matrix terms:

$$\mathbf{Y} = \mathbf{F}^0_1 \mathbf{\Lambda}^0_1 + \mathbf{F_2 \Lambda_2} + \boldsymbol{\varepsilon}_0 \tag{4}$$

where $\mathbf{F}^0_1$ is the $T \times 1$ vector of the first factor under the constrained loadings structure, $\mathbf{F}_2$ the $T \times (\widehat{k} - 1)$ matrix of the other factors, $\mathbf{\Lambda}^0_1 = [\lambda_{1G_1} \ldots \lambda_{1G_M}]$ the $1 \times N$ vector of loadings of the first factor obeying the null of homogeneity, and $\mathbf{\Lambda_2}$ the $(\widehat{k} - 1) \times N$ matrix of loadings for the remaining factors.

Before examining in detail estimation of model (4) let us see how we propose to evaluate if H_0 is compatible with the data.

The starting point is that by Bai's theorem 4, when H_0 holds the estimate of the restricted common component $\mathbf{F}^0_1 \mathbf{\Lambda}^0_1$ will converge to the true common component $\mathbf{F}_1 \mathbf{\Lambda}_1$ at a rate given by $Min(\sqrt{N}, \sqrt{T})$. Then for N and T large enough the partially de-factored data ($\mathbf{Z}$) obtained subtracting the restricted common component $\mathbf{F}^0_1 \mathbf{\Lambda}^0_1$ from the data are such that

$$\begin{aligned} \mathbf{Z} &= \mathbf{Y} - \mathbf{F}^0_1 \mathbf{\Lambda}^0_1 && (5)\\ &= \mathbf{F}_1 \mathbf{\Lambda}_1 + \mathbf{F_2 \Lambda_2} + \boldsymbol{\varepsilon} - \mathbf{F}^0_1 \mathbf{\Lambda}^0_1 && (6)\\ &= \mathbf{F_2 \Lambda_2} + \boldsymbol{\varepsilon} && (7) \end{aligned}$$

and thus include only the $(\widehat{k}-1)$ unrestricted factors collected in $\mathbf{F_2}$. Hence, when H_0 holds performing Bai's analysis on $\mathbf{Z}$ allowing for a maximum of $\widehat{k}$ factors we will asymptotically select $(\widehat{k}-1)$ factors with probability 1. On the other hand, if H_0 is false $\mathbf{F}_1\mathbf{\Lambda}_1$ and $\mathbf{F}_1^0\mathbf{\Lambda}_1^0$ will not cancel out, so that

$$\mathbf{Z} = \mathbf{Y} - \mathbf{F}_1^0\mathbf{\Lambda}_1^0 + \mathbf{F_2}\mathbf{\Lambda_2} \tag{8}$$

$$= \mathbf{F}_1\mathbf{\Lambda}_1 + \mathbf{F_2}\mathbf{\Lambda_2} - \mathbf{F}_1^0\mathbf{\Lambda}_1^0 + \boldsymbol{\varepsilon}. \tag{9}$$

Thus in principle $\mathbf{Z}$ includes $(\widehat{k}+1)$ factors, although the two common components $\mathbf{F}_1\mathbf{\Lambda}_1$ and $\mathbf{F}_1^0\mathbf{\Lambda}_1^0$ may be empirically difficult to distinguish in moderate samples. In this case $\mathbf{Z}$ may appear to have $\widehat{k}$ factors. At any rate, the consistency property of Bai's procedure implies that performing the analysis with $max(k) = (\widehat{k}+2)$ we will tend to select a number of factors $k_0 > (\widehat{k}-1)$. Summing up, selecting for the partially de-factored data $\mathbf{Z}$ one factor less than for the raw data $\mathbf{Y}$ provides support in favour of H_0, while selecting the same number or even more against it.

Let us now move to the constrained estimation and testing algorithm. Following Amengual and Repetto (2014), we propose to estimate model (4) with a two stages approach. In the first stage the constrained factor and loadings are estimated iteratively. Broadly speaking, the idea is to alternate steps in which the factor is taken as given and the loadings estimated under the desired constraints, and steps in which the factor is estimated taking as given the loadings as estimated in the previous step. In the second stage the uncostrained factors and their loadings are then estimated by PC's applied to the partially de-factored data obtained subtracting the constrained common components from the data. The details are as follows.

Constrained estimation procedure

A *Iterative estimation of the constrained common component* $\mathbf{F}_1^0\mathbf{\Lambda}_1^0$.

1. Estimate using Bai's procedure the uncostrained factor model $\mathbf{Y} = \mathbf{F}_1\mathbf{\Lambda}_1 + \mathbf{F_2}\mathbf{\Lambda_2} + \boldsymbol{\varepsilon}$. Label the unrestricted estimates of the first factor and its loadings as $\mathbf{F}_1^{(1)}, \mathbf{\Lambda}_1^{(1)}$.

2. Given $\mathbf{F}_1^{(1)}$, obtain estimates of the loadings under the homogeneity constraint. In this set-up the easiest way is by time series regressions of Y on F^1 separately for each of the M groups: $Y_{ti} = \boldsymbol{\lambda}_{1G_j} F_{1t}^{(1)}$, $t = 1, \ldots, T$, $i \in G_j$, $j = 1, \ldots, M$. Collect these estimates in the $1 \times N$ vector $\mathbf{\Lambda}_1^{(2)} = \left[\boldsymbol{\lambda}_{11}^{(2)} \ldots \boldsymbol{\lambda}_{1N}^{(2)}\right]$.

3. Given $\mathbf{\Lambda}_1^{(2)}$, obtain a new estimate of the factor by running the T cross-section regressions $Y_{ti} = \boldsymbol{\lambda}_{1i}^{(2)} F_{t1}$, $i = 1, \ldots, N$, with Y_{ti}, $\boldsymbol{\lambda}_{1i}^{(2)}$ as observables and F_{t1} as unknown parameter. Collect these estimates in the $T \times 1$ vector $\mathbf{F}_1^{(2)} = \left[F_{11}^{(2)} \ldots F_{T1}^{(2)}\right]'$.

4. Repeat step 2 using $\mathbf{F}_1^{(2)}$, obtaining a new vector of loadings $\mathbf{\Lambda}_1^{(3)}$.

5. Repeat step 3 using $\mathbf{\Lambda}_1^{(3)}$, obtaining a new estimate of the factor, $\mathbf{F}_1^{(3)}$.

6. Iterate until the maximum difference over all units and periods between the estimates of the common components in two successive iterations, $\underset{t,i}{max}(\mathbf{F}_1^{(n)}\mathbf{\Lambda}_1^{(n)} - \mathbf{F}_1^{(n-1)}\mathbf{\Lambda}_1^{(n-1)})$, meets the chosen convergence criterion. Define the restricted estimates of factor and loadings as $\widehat{\mathbf{F}}_1^0 = \mathbf{F}_1^{(n)}$, $\widehat{\mathbf{\Lambda}}_1^0 = \mathbf{\Lambda}_1^{(n)}$.

Remark 1 When H_0 holds for N and T large enough steps 2-6 will not be necessary, as the preliminary unrestricted estimates $\mathbf{F}_1^{(1)}\mathbf{\Lambda}_1^{(1)}$ will converge to the true common component $\mathbf{F}_1^0\mathbf{\Lambda}_1^0$. When, on the contrary, H_0 does not hold, the constrained estimates of the loadings will be averages of the latent true heterogenous loadings.

Remark 2 Exclusion restrictions are trivially handled setting to zero the relevant elements of the constrained loadings vector $\mathbf{\Lambda}_1^{(2)}$ in step 2 and following.

Remark 3 Restrictions on more than one factor are easily handled defining accordingly the regressions in steps 2-3 and using as a stopping rule the intersection of the convergence criteria.

B *Estimation of the uncostrained common components* $\mathbf{F_2\Lambda_2}$.

This is carried out applying Bai's PC method to the partially de-factored data $\mathbf{Z} = \mathbf{Y} - \widehat{\mathbf{F}}_1^0\widehat{\mathbf{\Lambda}}_1^0$. Since consistent selection requires $max(k) > r$ we must set $max(k) = (\widehat{k} + 2)$, as when H_0 is false $\mathbf{Z}$ will have $(\widehat{k} + 1)$ factors. Define the selected number of factors as k_0 and the estimated factors and loadings as $\widehat{\mathbf{F}}_{\mathbf{2}}$, $\widehat{\mathbf{\Lambda}}_{\mathbf{2}}$.

Remark Bai's different criteria in practice very often suggest different number of factors, with the final choice made on the basis of an informal joint evaluation of the three criteria (see, *e.g.*, Bai's empirical illustration on sectoral employment, p. 159). The standard view of the null hypothesis as holding unless otherwise proved suggests to follow in Stage B a parsimonious approach, selecting as the number of factors for the partially de-factored data the smallest between those suggested by the three criteria.

C *Evaluation of the null hypothesis* $H_0 : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_1^0$

Comparing the number of factors selected for $\mathbf{Z}$, k_0, with that chosen for $\mathbf{Y}$, $\widehat{k}$, we can now assess if H_0 is compatible with the data:

■ if $k_0 = (\widehat{k} - 1)$ do not reject $H_0 : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_1^0$;

■ else, reject H_0.

Clearly, the conclusion reached in this way is based upon asymptotic foundations, as it hinges upon the consistency of the criteria used to select the number of factors. In practice, in finite samples we may choose in Stage B the wrong number of factors, thus rejecting a true null hypothesis or failing to reject a false one. Although there is a strict analogy with the Type I and Type II errors of classical testing, since our procedure is not a test in the classical sense we prefer to define the event $k_0 > (\widehat{k} - 1)$ when H_0 holds, which leads to rejecting a true null hypothesis, a "false positive". Analogously, we define the event $k_0 = (\widehat{k} - 1)$ with H_0 *not* holding, which entails failing to reject a false null hypothesis, as a "false negative". Exactly as in classical testing we need estimates of the probability of the Type I error, here in order to know how reliable our conclusions are we need an estimate of the probability of a false positive. Our proposal is to obtain it using the bootstrap. To this end we need to construct pseudodatasets $\mathbf{Y}^*$ which:

(*i*) have the same time series properties of the originary data $\mathbf{Y}$;

(*ii*) have the same same number of factors of $\mathbf{Y}$;

(*iii*) obey the null hypothesis.

The first requirement, which dictates the choice of the resampling algorithm to be used, will be discussed below. The other two requirements may be met constructing the matrix of common components of $\mathbf{Y}^*$ as the sum of $\widehat{\mathbf{F}}_1^0\widehat{\mathbf{\Lambda}}_1^0$ and a component, $(\mathbf{F_2\Lambda_2})^*$, obtained resampling the matrix of the uncostrained estimates of the common components associated with the remaining $(\widehat{k}-1)$ factors, $\widehat{\mathbf{F}}_\mathbf{2}\widehat{\mathbf{\Lambda}}_\mathbf{2}$. To obtain $\mathbf{Y}^*$ we then simply append to these common components a matrix of pseudo-residuals, $\boldsymbol{\varepsilon}^*$, obtained resampling those of the unconstrained factor model (1), $\widehat{\boldsymbol{\varepsilon}}$:

$$\mathbf{Y}^* = \widehat{\mathbf{F}}_1^0\widehat{\mathbf{\Lambda}}_1^0 + (\mathbf{F_2\Lambda_2})^* + \boldsymbol{\varepsilon}^*$$

Let us now discuss how which resampling schemes should be applied in order to ensure that the first requirement is respected. Resampling the residuals is the easiest task. Since they are known to be weakly dependent over time and cross-sectionally, applying the Stationary Bootstrap (Politis and Romano, 1994) to the entire matrix (*i.e.*, resampling blocks of rows) will yield pseudo-residuals reproducing the dependence structure over both dimensions[4].

Resampling the common components $\mathbf{F_2\Lambda_2}$, known to be $I(1)$, is a more delicate problem, which recalls closely that to be solved in order to construct a bootstrap panel unit root or cointegration test. Parker, Paparoditis and Politis (2006) showed that asymptotically valid unit root tests can be carried out applying the Stationary Bootstrap to the first differences of an $I(1)$ series and cumulating the pseudo-differences to obtain the pseudo-levels. Di Iorio and Fachin (2014) extended this result to cointegration tests and panel cointegration tests for independent units. We thus conjecture that it may be extended further to the selection of the number of factors in Bai's procedure.

The bootstrap procedure we propose for the estimation of the probability of a "false positive" is thus the following:

Bootstrap procedure

1. resample applying the Stationary Bootstrap the first differences of the common components of the uncostrained factors as estimated in Stage B, $\Delta(\widehat{\mathbf{F}}_\mathbf{2}\widehat{\mathbf{\Lambda}}_\mathbf{2})$, obtaining the pseudo-differences $\Delta(\mathbf{F_2\Lambda_2})^*$; cumulate them using as starting values the estimates for $t=1$, $\sum_{j=2}^{k}\widehat{\lambda}_{ji}\widehat{F}_{1j}$, $i=1,\ldots,N$, and obtain $(\mathbf{F_2\Lambda_2})^*$.

2. resample applying the Stationary Bootstrap the residuals $\widehat{\boldsymbol{\varepsilon}}$ of the unconstrained factor model (1) obtaining the pseudo-residuals $\boldsymbol{\varepsilon}^*$.

3. append $(\mathbf{F_2\Lambda_2})^*$ and $\boldsymbol{\varepsilon}^*$ to $\widehat{\mathbf{F}}_1^0\widehat{\mathbf{\Lambda}}_1^0$, obtaining the pseudo-data $\mathbf{Y}^* = \widehat{\mathbf{F}}_1^0\widehat{\mathbf{\Lambda}}_1^0 + (\mathbf{F_2\Lambda_2})^* + \boldsymbol{\varepsilon}^*$ which obey $H_0 : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_1^0$ and have $\widehat{k}$ common factors.

[4]Briefly, the Stationary Bootstrap is a block resampling scheme working as follows. Let $L_1, \ldots, L_T$, be a sequence of values randomly chosen from a geometric distribution of parameter θ. The first L_1 observations of the pseudoseries are given by L_1 observations from the originary series starting at a location randomly chosen from a uniform distribution on $\{2, ..., T\}$, followed by L_2 observations starting at a different randomly chosen location, and so on. The process ends when the pseudoseries reaches the desired length. Such a series will reproduce the weak dependence links between the data.

4. estimate the unrestricted factor model on the pseudodata $\mathbf{Y}^*$ allowing for a maximum of $(\widehat{k}+1)$ factors. Let $\widehat{k}^*$ be the selected number of factors.

5. obtain an estimate of the factor $\widehat{\mathbf{F}}_1^{0*}$ and the associated loadings $\widehat{\mathbf{\Lambda}}_1^{0*}$ respecting the homogeneity constraint using the iterative algorithm described in Stage A.

6. construct the partially de-factored pseudo-data $\widehat{\mathbf{Z}}^* = \mathbf{Y}^* - \widehat{\mathbf{F}}_1^{0*}\widehat{\mathbf{\Lambda}}_1^{0*}$.

7. estimate the unrestricted factor model on $\widehat{\mathbf{Z}}^*$ allowing for a maximum of $(\widehat{k}^*+2)$ factors (Stage B of the constrained estimation algorithm). Let k_0^* be the selected number of factors.

8. if $k_0^* = \widehat{k}^* - 1$ do not reject $H_0 : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_1^0$; else reject it.

9. repeat 1-8 a large number of times and compute the proportion P^* of cases in which H_0 is rejected. This is the bootstrap estimate of the probability under H_0 of a false positive outcome.

3 Monte Carlo experiment

The performances of the proposed procedure have been evaluated by simulation using a Monte Carlo data generating process (DGP) very close to that used by Bai. The permanent component is driven by two $I(1)$ factors ($r = 2$), and the stationary part follows independent ARMA processes:

$$y_{ti} = \lambda_{1i}F_{t1} + \lambda_{2i}F_{t2} + \varepsilon_{ti} \tag{10a}$$
$$\Delta F_{1t} = e_{t1} \tag{10b}$$
$$\Delta F_{2t} = e_{t2} \tag{10c}$$
$$\varepsilon_{ti} = \phi_i\varepsilon_{t1} + \nu_{ti} + \theta_i\nu_{t-1,i} \tag{10d}$$

The noises e_1, e_2 are independent standard gaussians, and the $\nu_i's$ independent gaussians with heterogenous variances:

$$\begin{aligned} e_{t1},\ e_{t2} &\sim NID(0,1) \\ \nu_{ti} &\sim NID(0,s_i),\ s_i \sim unif(0.5,1.5). \end{aligned}$$

The ARMA parameters also allowed to be heterogenous over units:

$$\phi_i \sim unif(0.4,0.6) \tag{11a}$$
$$\theta_i \sim unif(0.3,0.5) \tag{11b}$$

The loadings of the second factor are also generated as indipendent gaussians with mean and standard deviation 1:

$$\lambda_{2i} \sim NID(1,1) \tag{12}$$

Obviously, all these parameters are fixed across simulations.

The null hypothesis is always that the loadings of the first factor are homogenous within two groups, set for simplicity of equal size ($N_1 = N_2 = 0.5N$):

$$H_0 : \lambda_{1i} = \begin{cases} \lambda_{1G_1},\ i \leq 0.5N \\ \lambda_{1G_2},\ i > 0.5N \end{cases} \tag{13}$$

To evaluate the frequency of "false positive" outcomes we fixed without loss of generality the loadings of the first group at 1 and those of the second at 5:

$$\lambda_{1i} = \begin{cases} 1, \, i \leq 0.5N \\ 5, \, i > 0.5N \end{cases} \tag{14}$$

while to assess the ability of the procedure to reject false hypothesis we split the second group in two subgroups of equal dimension with the loadings as follows:

$$\lambda_{1i} = \begin{cases} 1, & i \leq 0.5N \\ 5, & 0.5N < i \leq 0.75N \\ 10, & i > 0.75N \end{cases} \tag{15}$$

The obvious point that, given all the other DGP parameters, the ability of the procedure to detected wrong hypothesis will clearly grow with the difference between the loadings of the second and third group is less important then it may appear at a first sight. In this set-up assessing this ability in absolute terms can be argued to be a virtually impossible task, as the null hypothesis may be violated in a extremely large number of different ways beyond differences in loadings (homogeneity with different partitions of the units, full heterogeneity). Rather, we can aim at examining how this ability changes with the time and cross-section sample sizes, and for this task there is not much loss of generality in considering a given set of loadings.

We considered time and cross-section dimensions ranging from 50 to 150, sample sizes quite representative of empirical studies. In all cases we mimicked actual empirical work by estimating an unrestricted factor model on the raw data with $max(k) = (r + 1)$ and carrying out the two-stages estimation, with the convergence criterion been at 0.0001, only if the selected number of factors was greater than 1 (if a single factor is selected for $\mathbf{Y}$ the factor model data does not make sense for $\mathbf{Z}$). Consistently with the simulation results reported by Bai (Table 1) the estimated number of factors, $\widehat{k}$, turned out to be almost always (from 98% to 99.8% of the cases) equal to the true number r.

In a first set of experiments we evaluated the behaviour of the selection procedure on the de-factored data, with both H_0 true (frequency of "false positive" outcomes, Table 1) and H_0 false (frequency of "true positive" outcomes, Table 2), with 5000 Monte Carlo replications. The results reported here have been obtained using in Stage A the criterion IPC_1 (IPC_2 and IPC_3 gave essentialy similar results, available on request). For the de-factored data we choose the smallest number of factors across all IPC's, an automatic rule very convenient for simulations.

Let us examine first the frequency of "false positive" outcomes reported in Table 1. Recall that these are the Monte Carlo estimates of the probability of selecting for the de-factored data $\mathbf{Z}$ a number of factors $k_0 > (\widehat{k} - 1)$ when H_0 holds. As expected, these frequencies are always negligible, in most of the cases smaller than 1%. The proposed procedure thus appears to be reliable in the sense that it is very unlikely to suggest rejection of true null hypothesis.

Moving to the frequencies of "true positive" outcomes (frequency of the cases in which $k_0 > (\widehat{k} - 1)$ when H_0 is false), in Table 2 we can see that the procedure seems also to have a good ability to detect false null hypothesis. However, as discussed above, more than the measures of its absolute levels our attention is centred on its behaviour as the sample size change. Here the key point seems to be that the length of the time sample matters more than the number of units. Thus the clear message is that the procedure should be applied with care with small or moderate time samples.

Finally, in a second set of experiments we studied the bootstrap estimates of the probability of a "false positive" outcome. Here we report exploratory findings with 250 Monte Carlo replications and bootstrap

redrawings; more extensive experiments are in progress. The block lengths for the Stationary Bootstrap have been generated with a geometric distribution with mean parameter computed as $1.75\sqrt[3]{T}$, as in Palm, Smeekes and Urbain (2011).

From Table 3 we can appreciate that the bootstrap algorithm proposed above delivers rather accurate estimates. To read this table recall that for each Monte Carlo replication we obtained one bootstrap estimate; over all the 250 Monte Carlo replications we thus have a distribution of bootstrap estimates, which in Table 3 is summarised by the mean $\overline{f}^*$. This is to be compared with the estimate of the probability given by the frequency f of "false positives" over these 250 Monte Carlo replications, clearly somehow different from those reported in Table 1 obtained from 5000 replications. Overall we find the bootstrap estimates to be quite close to the Monte Carlo frequencies. For instance, for samples sizes close to that of our empirical illustration, $T = 50, N = 150$, we find that given a Monte Carlo frequency of "false positives" of 0.4% the mean of the distribution of the bootstrap estimates is 0.6%. Clearly, given the very low frequency of "false positive" outcomes a proper evaluation requires a high number of Monte Carlo replications and bootstrap redrawings. As anticipated above, this is object of current work.

Table 1
"False positive"
Frequency×100

		N	
T	50	100	150
50	0.5	0.0	0.1
100	1.2	0.1	0.3
150	1.4	0.4	1.0

DGP: equations (10a)-(12),(14)

Table 2
"True positive"
Frequency×100

		N	
T	50	100	150
50	73.8	61.1	68.9
100	82.4	78.8	87.7
150	85.8	85.5	92.5

DGP: equations (10a)-(12),(15)

Table 3
Bootstrap estimates of the
Frequency (×100) of "false positives"
Monte Carlo statistics

	$T = 50$			$T = 150$		
	N	50	150	N	50	150
f		2.0	0.4		4.0	1.2
$\overline{f}^*$		2.3	0.6		3.2	2.0

DGP: equations (10a)-(12),(14)
f : Monte Carlo frequency
$\overline{f}^*$: Monte Carlo mean of bootstrap estimates

4 Empirical illustration

As a final step, to illustrate an empirical application we apply the proposed procedure to the 19th century Italian annual regional value added data described in the Introduction. The dataset, estimated by Ciccarelli and Fenoaltea (2009, 2014), covers 10 industries in 16 regions. One industry was absent in one region, so that $N = 159$. The four macroareas (NW, NE, Centre, South) include respectively 30, 20, 40 and 69 units (industry/region pairs). The period is 1861-1913, hence $T = 53$. As mentioned in the Introduction, applying Bai's procedure Ciccarelli and Fachin (2016) selected $\widehat{k} = 2$. This choice was supported by the finding that the fully de-factored data are clearly stationary. The first factor is an upward-sloping trend with loadings showing a clear spatial pattern: setting the average across industries and regions in the NW at 100, those of the NE and the Centre are both about 70 and of the South about 60. These large regional differentials suggest testing the hypothesis that the loadings are constant for all industries and regions within each of these four macroareas. This would imply that (*i*) regional differences mattered more than industrial ones, (*ii*) the North-South divide has deep roots in Italian economic history. Formally we can state the null hypothesis as:

$$H_0 : \lambda_{1i} = \lambda_{1G_j},\ \forall i \in G_j,\ j = NW, NE, Centre, South \tag{16}$$

In this case it is quite interesting to examine the results obtained using in Stage B each of the three criteria in turn as a basis for the selection of the restricted factor model. Before entering into the details, it is useful to compare our procedure with a classical test. The outcome of classical tests is given by the value of the test statistic and its p-value, either computed on the basis of the asymptotic distribution of the statistic or by simulation. The decision to reject or not reject is then taken conditionally upon the probability of rejecting a true null hypothesis which is deemed acceptable (nominal level, α, typically fixed at customary levels), which may be fixed at will. As a consequence, the final decision is not univocally determined: we may reject at a nominal level α if $p < \alpha$, while not rejecting at some other level $\alpha' < \alpha$ such that $\alpha' < p$.

On the opposite, in our procedure the outcome is the number k_0 of factors in the de-factored data, which determines univocally the decision. If $k_0 > (\widehat{k} - 1)$ the null hypothesis is rejected, else it is not. Contrary to classical tests, in our procedure we have no control on the probability of rejecting a true null

(what we defined above a "false positive"); we can simply estimate it using the boostrap so to know how reliable the conclusion reached is. In fact, this probability depends on how good the performance of the model selection criteria used is with the sample at hand.

Let us move to the empirical results. First of all, from Table 4 we see that IPC_1 suggests the same number of factors found for the raw data: $k_0 = 2$. Since de-factoring under H_0 does not reduce the number of factors we should reject it. How much can we thrust this conclusion? In other terms, how high is the risk that the null hypothesis we are rejecting is in fact true? This information is provided by the bootstrap estimate of the probability of a false positive in the bottom row of Table 4 (f^*), equal to 14%. We thus estimate that using IPC_1 we have a 14% probability of choosing too many factors and wrongly rejecting a true set of restrictions. This is clearly too high. Before moving further, it is essential to make clear that f^* must *not* be read as the equivalent of the p-value of classical tests: on the opposite, it is the analogous of the nominal level of a classical test leading to rejection. This may appear puzzling at first sight, but it is in fact easily seen. Our procedure suggests rejection, with the probability that H_0 is in fact true estimated as 14%. When would a classical test reject with such a Type I error? Obviously, a test with a nominal level $\alpha = 14\%$.

The second criterion, IPC_2, also suggests $k_0 = 2$, thus rejection. The bootstrap estimate of the probability of a false positive is about 10%. Although slightly lower, this probability would be considered still somehow too high by most applied workers.

Finally, the third criterion, IPC_3 suggests $k_0 = 1$, which is thus the "parsimonious estimate" we claimed above to be a natural choice in Stage B in case of contrasting indications from the three criteria. The implication is that, since de-factoring under H_0 does reduce the number of factors, its restrictions are compatible with the data. Of course, since now we are not rejecting, we are not interested in the probability of "false positives"; rather, to thrust our conclusions we would need to know the probability of "false negatives". In principle we could estimate it using the bootstrap along analogous lines, but the results would inevitably be limited to some specific alternative hypothesis. We thus more simply recover some indications from the results of our simulation experiment. From Table 3 we see that with similar sample sizes, $T = 50$, $N = 150$, the probability of "false negatives" is around 30%. This high value suggests that in our case caution is advisable.

Table 4
Factor Modelling of data de-factored under H_0

	IPC_1	IPC_2	IPC_3
k_0	2	2	1
H_0	*rejected*	*rejected*	*not rejected*
f^*	14.0	9.7	-

k_0 : number of factors
f^* : bootstrap estimate of P("false positive")$\times$100,
5000 bootstrap redrawings

5 Conclusions

Although research in factor modeling is very active to say the least, the problem of evaluating restriction on the loadings has not received much attention yet. In this paper we examined this topic in the special case of non-stationary data. We propose to evaluate exclusion or homogeneity restrictions in non-stationary factor models on the basis of the number of factors estimated for the partially data defactored under the null hypothesis, with the probability of rejecting true null hypothesis estimated by the bootstrap. Simulation results suggest the procedure has good properties and may thus be a valuable tool for applied factor modelling of non-stationary data. As an empirical illustration, we applied the procedure to the factor model estimated by Ciccarelli and Fachin (2016) for the Ciccarelli and Fenoaltea (2009, 2014) 19th century value added series for the Italian economy disaggregated by industry and region. The hypothesis tested is that the loadings of the common trend factor are homogenous for all industries and regions within each of the four macroareas typically used to partition the country. The results are mixed, with model selection on the basis of two criteria rejecting homogeneity but with a probability of a "false positive" outcome higher than the levels traditionally accepted in classical tests, and one criteria supporting it, with zero probability of a "false positive". Finally, clearly much work still needs to be done. Current research includes proving the consistency of Bai's model selection procedure for the boostrap data and more extensive simulations of the bootstrap procedure.

6 References

Amengual, D., L. Repetto (2014) "Testing a large number of hypotheses in approximate factor models" Working Paper, CEMFI.

Bai, J. (2004) "Estimating cross-section common stochastic trends in nonstationary panel data" *Journal of Econometrics,* 122, 137–183.

Banerjee, A., M. Marcellino, I. Masten (2015) "An Overview of the Factor-augmented Error-Correction Model" University of Birmingham, Dept. of Economics Discussion Paper 15-03.

Chang, Y. (2004) "Bootstrap unit root tests in panels with cross-sectional dependency" *Journal of Econometrics*, 120, 263-293.

Ciccarelli, C., S. Fachin (2016) "Common Factors and spatial spillovers in the industrial development of the Italian regions, 1861-1913" WP, "Sapienza" University of Rome.

Ciccarelli, C., S. Fenoaltea (2009) "La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione statistica. 1 Le industrie non manifatturiere" Rome, Bank of Italy.

Ciccarelli, C., S. Fenoaltea (2014) "La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione statistica. 2 Le industrie estrattivo-manifatturiere" Rome, Bank of Italy.

Di Iorio, F., S. Fachin (2014) "A panel cointegration study of the long-run relationship between savings and investments in the OECD economies, 1970-2007" *Empirical Economics*, 46, 1271-1300.

Palm, F.C., S. Smeekes, J.P. Urbain (2011) "Cross-sectional dependence robust block bootstrap panel unit root tests" *Journal of Econometrics*, 163, 85-104.

Parker, C., E. Paparoditis, D.N. Politis (2006) "Unit root testing via the stationary bootstrap" *Journal of Econometrics*, 133, 601-638.

Politis, D.N., J.P. Romano (1994) "The Stationary Bootstrap" *Journal of the American Statistical Association*, 89, 1303-1313.

Reis, R., M.W. Watson "Relative goods' prices, pure inflation, and the Phillips correlation" *American Economic Journal: Macroeconomics*, 2, 128-157.

Stock, J.H., M. W. Watson (2011) "Dynamic Factor Models" M.P. Clements, DF. Hendry (Eds.) *The Oxford Handbook of Economic Forecasting* Oxford University Press, Oxford.