****
This is a draft
Please do not quote or circulate
****

A Subsampled Penalty Criterion to Estimate the Number of Non-Vanishing Common Factors in Large Panels

Oualid Bada and Dominik Liebl[1]

Abstract

The problem of assessing the number of common factors in large dimensional panel data has been heavily researched and discussed in the recent literature on factor models. The importance of this issue in many econometric and finance applications has triggered the development of a new generation of information criteria, that are, in contrast to the traditional functions such as AIC and BIC, consistent under both large cross-section dimensions n and large time periods T. Most of the existing works focus, however, only on the concept of consistency without caring about the (asymptotic) efficiency. This paper is concerned with the problem of estimating the number of factors that do not vanish when n and T get large in general panel models. The factors are allowed to be extracted directly form observed variables or estimated with other model parameters such is the case in advanced panel data models with unobserved heterogeneous common factors. Our method relies on a subsampling strategy that works with high-performance under minimal assumptions. Finite sample performance of the estimator is examined via Monte Carlo studies. In our application, we estimate the number of efficiency factors in the U.S. largest banks over the period from 1999 to 2009.

Keywords: Large Panel Data, Factor Structure Model, Model Selection Criteria, Subsampling

1. Introduction

The problem of assessing the number of common factors in large dimensional panel data has been heavily researched and discussed in the recent literature on factor models. The importance of this issue in many econometric and finance applications has triggered the development of a new generation of information criteria, that are, in contrast to the traditional functions such as AIC and BIC, consistent under both large cross-section dimensions n and large time periods T; see, e.g., Bai and Ng (2002), Onatski (2009), Kapetanios (2010), Kneip et al. (2012), etc. Most of the existing works focus, however, only on the concept of consistency without caring about the (asymptotic) efficiency; a reason why many ongoing researches aspire towards finding high-performance and precise criteria.

We consider the following general model presentation:

$$Y_{it} = h(Z_{it}, \Theta) + \sum_{k=1}^{r} \lambda_{ik} f_{kt} + e_{it}, \quad \text{for } i = 1, \ldots, n \text{ and } t = 1, \ldots, T, \qquad (1)$$

where $h(Z_{it}, \Theta)$ contains observed regressors or instruments that interact with Y_{it} (linearly or non-linearly) through a set of unknown parameters, $\Theta = \{\theta_1, \ldots, \theta_P\}$, λ_{ik} are unobserved individual loading parameters, f_{kt} are unobserved common factors, and d is the unknown factor dimension that has to be estimated jointly with the model parameters.

This paper is concerned with the problem of estimating d, the number of factors that do not vanish when n and T get large, independently of the method established to consistently estimate the parameters Θ, λ_{kl} and f_{lk}. In the special case, where $E[h(Z_{it}, \Theta)|Z_{it}] = 0$, Model (1) can be reverted to a classical static factor model. According to Forni et al. (2014), Stock and Watson (1999), and Bai (2003), consistent estimates of λ_{ik} and f_{kt} can be obtained, in this case, by using the method of asymptotic principal components. If $h(Z_{it}, \Theta)$ models a linear interaction between P regressors and a $(P \times 1)$ vector of non-zero homogeneous parameters, then (1) can be interpreted as a panel data model with time-varying individual effects (or with interactive fixed effects) as described in Bai (2009), Kneip et al. (2012), Ahn et al. (2013), and Bada and Kneip (2013). In addition, if the parameters are individual specific such that $h(Z_{it}, \Theta) = \sum_{i=1}^{n} Z_{it}^{'}\theta_i$, then Model (1) can be treated as an heterogeneous panel data model with multifactor error structures as in Pesaran (2006). In the case where some factors and/or some regressors are non-stationary whereas the idiosyncratic errors are generated by stationary processes, Model (1) can be classified into the category of panel cointegration models; see, e.g., Bai et al. (2009) and Kapetanios et al. (2011). Note that this class of panel models, composed of observed regressors and unobserved common factors, has attracted an increasing attention in the recent literature on applied econometrics because it provides a generalization of the restrictive standard two-way panel data models, in which $r = 2$, $f_{1t} = 1$, for all $t = 1, \ldots, T$, and $\lambda_{i2} = 1$, for all $i = 1, \ldots, n$. Moreover, model (1) can include dynamic factor models in static form as discussed in Stock and Watson (2005). In the latter case, $Z_{it} = (Y_{i,t-1}, \ldots, Y_{i,t-P})^{'}$ and $h(Z_{it}, \Theta) = \sum_{p=1}^{P} \theta_p Y_{i,t-p}$.

It is important to emphasize that the goal of estimating the factor dimension is not only in order to reduce the space dimension of the examined variable but also to assess the true number of factors that really exist and do not vanish when n and T get large. Such an analysis can not be obtained from a simple scree-plot or by naively interpreting the cumulated variance proportions of the factors, especially when the factor structure contains a mixture of weak and strong factors. As an example, consider the case of a response variable that is influenced by the effect of stationary random shocks as well as non-stationary global stochastic trends. The variance of the non-stationary factors will diverge, in this case, whereas the variance of the well behaved factors remain bounded and small. A scree-plot and/or a descriptive statistic based on the cumulated variance proportions will, hence, lead to overlook the weak factors when T is large.

Besides the risk of getting spurious interpretations, the misspecification of the true number of factors can also lead to inconsistent estimation of the remaining model parameters when one or more of the missing factors are correlated with the elements of $h(Z_{it}, \Theta)$. This is a reason why it is sometimes crucial to estimate the static number of factors r even when the factor structure arise from a smaller number of dynamic factors (also called primitive factors). Indeed, q-dimensional dynamic factor models of order m can be re-expressed in terms of r-dimensional static factors that include both the leads and the lags of the dynamic factors such that $r = q(m+1)$. As argued in Bada and Kneip (2013), the number of dynamic factors can be estimated a posteriori once the model parameters are consistently estimated with sufficiently fast convergence rates.

In the recent literature on applied econometrics, the most frequently used method for estimating the number of static factors is the penalized panel information criteria developed by Bai and Ng (2002). The authors consider approximate factor models in the sens of Chamberlain and Rothschild (1983) and extend the idea of the standard penalty criteria used in the time series analysis to provide consistent criteria for large panels under weak forms of heteroskedasticity and dependency in the errors. However, the performance of this criteria depends by construction on the choice of an a-priori maximum number of factors to be considered in the analysis. Onatski (2009) assumes Gaussian idiosyncratic errors and proposes a sequential testing procedure based on the ratios of adjacent eigenvalues obtained form the factor structure covariance matrix. Onatski (2010) allows for the errors to be either serially correlated or cross-sectionally dependent and proposes a threshold approach based on the empirical distribution properties of the largest eigenvalue. In order to robustify the criteria of Bai and Ng (2002), Alessi et al. (2010) propose to adjust the height of the penalty term by introducing a scaling parameter and using the calibration strategy of Hallin and Liška (2007), who consider similar criteria in context of dynamic factor models. Kapetanios (2010) proposes a sequential testing procedure based on a subsampling method. The author assumes, however, the factors and the loading parameters to be stationary and weakly dependent over time and across individuals. Note that all this methods assume the factors to be extracted directly from observed variables and not estimated with other model parameters such is required in Model (1).

Whereas the most common way to estimate f_{lt} and λ_{il} is factor analysis based on the principle component method, various techniques have been developed to estimate the set of parameters $\Theta = \{\theta_1, \ldots, \theta_P\}$. In the context of panel models with homogeneous slope parameters, Kneip et al. (2012) propose a semi-parametric method based on spline theory. Alternatively, Bai et al. (2009) uses an iterative least squares method. Bada and Kneip (2013) extend the method of Bai et al. (2009) and propose a parameter cascading strategy to calibrate the penalty term iteratively and estimate the factor dimension and the model parameters simultaneously. Pesaran (2006) considers the case of heterogeneous slope parameters and attempts to control for the unobserved factor structure by augmenting the model with the cross-section averages of the dependent variables and the cross-section averages of the observed explanatory variables.

Our method relies on a subsampling strategy that works with high-performance under minimal assumptions. This technique is similar to bootstrap, but much more appropriate to deal with the problem of (weakly) dependent idiosyncratic errors. The idea behind our approach is to optimize a panel criterion that is panelized with a multi-layered parameter structure. We show that the parameters of the penalty term can be estimated from the data by using a subsampling strategy independently of the method applied to estimate Θ, λ_{il} and f_{lt}. All we need is that the estimators ensure some convergence conditions that follow basically from the existing literature on large factor and panel model analyses.

The remainder of this paper is organized as follows: Section 2 presents and discusses the required assumptions. In Section 3, we argue how we can parametrize the panel information criterion to provide asymptotic efficiency. In Section 4, we describe a subsampling strategy to estimate the parameters of the penalty term in the information criterion. Conclusions and remarks are provided in Section 5.

2. Estimating the Number of Factors

In vector notation, model (1) can be presented as following:

$$Y_i = h(Z_i, \Theta) + \sum_{k=1}^{r} F_k \lambda_{ik} + \epsilon_i, \tag{2}$$

where $Y_i = (Y_{i1}, \ldots, Y_{iT})^{'}$, $h(Z_i, \Theta) = \big(h(Z_{i1}, \Theta), \ldots, h(Z_{iT}, \Theta)\big)^{'}$, and $F_k = (f_{k1}, \ldots, f_{kT})^{'}$. Let

$$\hat{Y}_i(k) = \begin{cases} h(Z_i, \hat{\Theta}), & \text{for} \quad k = 0, \text{ and} \\ h(Z_i, \hat{\Theta}) + \sum_{l=1}^{k} \hat{F}_l \hat{\lambda}_{il}, & \text{for} \quad k = 1, 2, \ldots < \infty, \end{cases}$$

be an estimator of

$$X_i(k) = \begin{cases} h(Z_i, \Theta) & \text{if} \quad k = 0, \\ h(Z_i, \Theta) + \sum_{l=1}^{k} F_l \lambda_{il} & \text{if} \quad 1 \leq k < r, \text{ and} \\ h(Z_i, \Theta) + \sum_{l=1}^{r} F_l \lambda_{il} & \text{if} \quad k \geq r, \end{cases}$$

and let

$$V_{n,T}(k) = \frac{1}{nT}\sum_{i=1}^{n}\big(Y_i - \hat{Y}_i(k)\big)^{'}\big(Y_i - \hat{Y}_i(k)\big)$$

be ordered such that

$$V_{n,T}(0) - V_{n,T}(1) \geq V_{n,T}(1) - V_{n,T}(2) \geq \cdots \geq 0.$$

Note that in the standard factor models (i.e., without the presence of $h(Z_{it}, \Theta)$), the ordered differences $V_{n,T}(0) - V_{n,T}(1), V_{n,T}(1) - V_{n,T}(2), \ldots$ will conform to the ordered eigenvalues of the covariance matrix $\frac{1}{nT}\sum_{i=1}^{n} Y_i Y_i^{'}$ if the asymptotic principal component method is used to fit the model; see e.g., Stock and Watson (2005), Bai and Ng (2002), and Bai (2003).

First, we follow Bai and Ng (2002) and construct a panel dimensionality criterion of the form:

$$PC_{\alpha}(k) = V_{n,T}(k) + k g_{n,T}(\alpha), \tag{3}$$

with the difference that the penalty $g_{n,T}(\alpha)$ is not a simple function of n and T but depends also on a set of parameters that can be specified through a control parameter α. The specification of these parameters will be discussed later. If $g_{n,T}(\alpha)$ is well specified, a consistent estimator of r can be obtained by minimizing numerically $PC_{\alpha}(k)$ with respect to k such that

$$\hat{r} = \arg\min_{k=1,\ldots,k_{max}} PC_{\alpha}(k). \tag{4}$$

2.1. Assumptions and Inherited Results

Let $(\Omega, \mathcal{F}, P)$, define the probability space of the random idiosyncratic errors $\{e_{it}\}_{i,t\in\mathbb{N}^2}$. We denote by P_c the conditional probability generating $\{e_{it}\}_{i,t\in\mathbb{N}^2}$ given $\{X_i(r)\}_{i\in\mathbb{N}}$. Throughout, we use the Landau-symbols $O_p(.)$, $o_p(.)$ and $\Omega_p(.)$ and the probability limit operator plim to describe the asymptotic behavior of the random variables in the probability space $(\Omega, \mathcal{F}, P_c)$.

Our method requires an estimator $\hat{Y}_{it}(k)$ that satisfies the following general result:

$$V_{n,T}(k) - \mathcal{V}_{n,T}(r) = \quad o_p(1), \quad \text{for all finite } k \geq r \text{ and} \tag{5}$$

$$V_{n,T}(k) - \mathcal{V}_{n,T}(r) = \quad \Omega_p(1), \quad \text{for all } k < r, \tag{6}$$

where

$$\mathcal{V}_{n,T}(k) = \frac{1}{nT}\sum_{i=1}^{n}\big(Y_i - X_i(k)\big)^{'}\big(Y_i - X_i(k)\big).$$

Condition (6) can be written as $\operatorname{plim}\inf_{n,T\to\infty} V_{n,T}(k) - \mathcal{V}_{n,T}(r) > 0$. It reflects (in an informal way) the asymptotic bias arising from under-parameterizing the factor structure by using $k < r$ factors instead of r. Condition (5) can be split up into $V_{n,T}(k) - \mathcal{V}_{n,T}(k) = o_p(1)$, for all $k \leq r$, and $V_{n,T}(k) - \mathcal{V}_{n,T}(r) = O_p\big(\pi(n,T)\big)$, for all finite $k > r$, where $\pi(n,T)$ is a positive function strictly decreasing in n and T. The quotient $1/\pi(n,T)$, in this case, will reflect the ability of the estimator $\hat{Y}_i(k)$ to detect the fact that the common information generated by the factor structure is accumulated stochastically faster than the unit specific information in the idiosyncratic errors, as $n, T \to \infty$.

We want to emphasize that, in the most of the existing works on large panel models of form (1), (5) and (6) follow immediately from the asymptotic properties of the estimators.

Example 1. Straightforward computations show that the semi-parametric estimator of Kneip et al. (2012) satisfy Requirements (5)-(6). More precisely, for all finite $k \geq r$, $V_{n,T}(k) - \mathcal{V}_{n,T}(r) = O_p\big(\max\{T^{-\frac{4}{5}}, n^{-1}\}\big)$, if the goal is to optimize the smoothness of the individual curves $\hat{v}_i(t) = \sum_{k=1}^{r} \hat{\lambda}_{ik}\hat{f}_{kt}$, and $V_{n,T}(k) - \mathcal{V}_{n,T}(r) = O_p\big(\max\{T^{-1}, (n,T)^{-\frac{4}{5}}, \frac{\max\{1,(n,T)^{-\frac{4}{5}}T^4\}}{T^5 n}, n^{-1}\}\big)$, if the goal is to optimize the smoothness for the estimated common factors $\hat{f}_{1t}, \ldots, \hat{f}_{rt}$.

Example 2. In Bai and Ng (2002), Bai (2003), and Bai (2009), $V_{n,T}(k) - \mathcal{V}_{n,T}(r) = O_p(\max\{n^{-1}, T^{-1}\})$ if $k \geq r$ and $\operatorname{plim\,inf}_{n,T\to\infty} V_{n,T}(k) - \mathcal{V}_{n,T}(r) = \tau_k$ if $k < r$, where $\tau_k > 0$.

Example 3. In the nonstationary panel data model of Bai (2004), we have $\operatorname{plim\,inf}_{n,T\to\infty} \frac{T}{\log\log T}\big(V_{n,T}(k) - \mathcal{V}_{n,T}(r)\big) > 0$, for $k < r$ and $V_{n,T}(k) - \mathcal{V}_{n,T}(r) = O_p(\max\{n^{-1}, T^{-2}\})$ for all finite $k \geq r$.

Whatever the method used for estimating Θ, λ_{il} and f_{lt}, the substantial requirement for our approach is that (5)-(6) hold true under the following additional assumptions:

Assumption 1. [Weak Dependence in the Cross-Section and Time Dimensions]

(i) The time series $\{e_{it}\}_{i,t\in\mathbb{N}^2}$ are stationary and independent of $X_i(r) = (X_{i1}(r), \ldots, X_{iT}(r))'$ for all $i, t \in \mathbb{N}^2$.

(ii) There is a known blockordering of the cross-sections such that the sequence $\{e_{it}\}_{i,t\in\mathbb{N}^2}$ is homogeneous in the sense of Politis and Romano (1994).

(iii) For each fraction $\lfloor n^s \rfloor$ and $\lfloor T^s \rfloor$ of n and T, where $0 < s < 1$ and $\lfloor x \rfloor$ denotes the largest integer $\leq x$, the sequence $\{e_{it}\}_{i,t\in\mathbb{N}^2}$ satisfies the following mixing condition:

$$\lim_{n,T\to\infty} \frac{1}{S_{n,T}} \sum_{\xi=1}^{S^*_{n,T}} \xi \alpha_e\big(\xi, s(n), s(T)\big) = 0, \tag{7}$$

where $S_{n,T} = (n - \lfloor n^s \rfloor + 1)(T - \lfloor T^s \rfloor) + 1)$, $S^*_{n,T} = \max\{(n - \lfloor n^s \rfloor + 1), (T - \lfloor T^s \rfloor + 1)\}$ and $\alpha_e\big(\xi, s(n), s(T)\big)$ is the mixing function as defined in Politis and Romano (1994) with the distance ξ and the underlying probability measure P_c.

Assumption 2. [Existence of Non-Degenerate Asymptotic Distributions]

(i) For each $k \geq r$, there exists a non-degenerate asymptotic distribution $K_k(x|P_c)$, continuous in x, such that

$$\lim_{n,T\to\infty} P_c\left[\pi(n,T)^{-1}\big(V_{n,T}(r) - V_{n,T}(k)\big) \leq x\right] = K_k(x|P_c).$$

(ii) For each finite $k < r$, there exists a strictly positive function $\tau_k(n,T)$, non-decreasing in n, T, and a strictly positive coefficient $\mu(k)$ such that

$$\text{plim} \inf_{n,T\to\infty} \tau_k(n,T)^{-1}\big(V(k) - \mathcal{V}(r)\big) \geq \mu(k).$$

Assumption 3. [Simultaneous Sample Divergence] Let $n, T \to \infty$ simultaneously, such that $T \sim n^c$ for some $0 < c < \infty$. The function $\pi(n,T)$ is of the form

$$\pi(n,T) = \max\{n^{\beta_1}, T^{\beta_2}, n^{\beta_3}T^{\beta_4}\},$$

for some $\infty < \beta_i < 0$, $i = 1, \ldots, 4$, so that we can write $\pi(n,T) = (\min\{n,T\})^\beta$ with $\beta = \min\{\beta_1, c\beta_2, (\beta_3 + c\beta_4)\}$ if $n = \min\{n,T\}$ and $\beta = \min\{\beta_1/c, \beta_2, (\beta_3/c + \beta_4)\}$ if $T = \min\{n,T\}$.

To save on nation, we denote, hereafter, the asymptotic distribution $K_k(x|P_c)$ by $K_k(x)$.

Assumptions 1 (i) and (ii) follow basically the setup of Kapetanios (2010), Connor and Korajczyk (1993) and Politis and Romano (1994). They allow for some dependence in both time and cross-section dimensions. Condition (7) in Assumption 1 generalizes the classical (unidimensional) strong mixing concept to the case of a two-parameter space (n,T) as proposed by Politis and Romano (1994). For more intuitions and details about strong mixing processes, we refer the reader to, e.g., Doukhan (1994), Politis and Romano (1994), and Ibragimov and Rozanov (1978). Note that Assumption 1 (ii) rules out simple random sampling and imposes on the individuals that are correlated to be located in the same neighborhood and sampled as blocks.

In Assumption 2 (i), we only assume that the statistic $\pi(n,T)^{-1}\big(V_{n,T}(r) - V_{n,T}(k)\big)$ has asymptotically a non-degenerate distribution function that depends only on P_e. No additional assumptions on the type of $K_k(x)$ or on their properties are placed. Moreover, we let $\tau_k(n,T), \mu(k)$ and $\pi(n,T)$ to be unknown. Assumption 2 (ii) allows for the term in Equation (6) to diverge with n and T. Such a setup escapes from the restrictive stationarity assumption of the factors and allows for the factor structure to have an infinite variance and/or a time-varying mean.

The first part of Assumption 3 relates T and n is such a way that the asymptotic behavior of each dimension can be functionally substituted by the behavior of the other, i.e., we can write T as $T(n)$ or $n = n(T)$. This assumption is frequently used in the literature on large panels, when n and T diverge simultaneously.

Proposition 1. *Let* $U_{n,T} = \min_{k\leq r}\{\mu(k)\tau_k(n,T)\}$ *and* $L_{n,T}(\alpha) = q(\alpha)\pi(n,T)^{1-\alpha}$, *where* $q(\alpha) = \inf_x\{x : K_{r+1}(x) \geq 1-\alpha\}$, *with* $\alpha \in (0,1)$. *Suppose that, under our additional Assumptions 1-3, the considered estimators of* $\{\Theta, \lambda_{il}, f_{lt} : i \in \{1,\dots,n\},\ t \in \{1,\dots,T\},\ l\{1,\dots,r\}\}$ *satisfy* (6)-(5) *and* $\hat{r}$ *minimizes* $PC_\alpha(k)$, *with respect to* $k = 1,\dots,k_{max} < \infty$, *then if*

$$L_{n,T}(\alpha) \leq g_{n,T}(\alpha) \leq U_{n,T},$$

for sufficiently large n *and* T, *we have*

$$\lim_{n,T\to\infty} P(\hat{r} = r) = 1.$$

2.2. Calibrating the Panel Criterion

Note that Proposition 1 provides sharper consistency conditions on $g_{n,T}(\alpha)$ than those that are given in Theorem 2 and Theorem 1 in Bai and Ng (2002) and Bai (2004) respectively, especially when α is small. Bai's conditions can be derived and generalized to our situation by replacing the double inequality in Proposition 1 by

$$\begin{array}{ll} (i): & \lim_{n,T\to\infty} \tau(n,T)^{-1} g_{n,T}(\alpha) \to 0 \text{ and} \\ (ii): & \lim_{n,T\to\infty} \pi(n,T)^{-1} g_{n,T}(\alpha) \to \infty. \end{array}$$

However, the problem with (i) and (ii) is that the degree of freedom in the choice of the penalty is too large. Any ad-hoc choice of $g_{n,T}(\alpha)$ will be, hence, compensated by a loss of an (asymptotic) efficiency.

From Proposition 1, we can see that an efficient penalty can be obtained by setting $g_{n,T}(\alpha)$ at one of the boundaries, $U_{n,T}$ or $L_{n,T}(\alpha)$, since the consistency condition can be achieved with the smallest possible sample size. We show that, by performing a subsampling strategy, we can estimate the components of $L_{n,T}(\alpha)$, namely the quantile $q(\alpha)$ and the rate $\pi(n,T)$, with sufficiently high convergence rates and plug the estimators in the dimensionality criterion $PC_\alpha(k)$ to estimate r consistently.

The basic idea of our method can be summarized in the following steps:

1. we propose a subsampling strategy to construct consistent estimators, $\tilde{q}(\alpha)$ and $\tilde{\pi}(n,T)$, of $q(\alpha)$, and $\pi(n,T)$, respectively,
2. we use $\tilde{q}(\alpha)$ and $\tilde{\pi}(n,T)$ to calibrate the penalty term such that $g_{n,T}(\alpha) = \tilde{q}(\alpha)\tilde{\pi}(n,T)^{(1-\alpha)}$, and
3. estimate r by minimizing

$$PC_\alpha(k) = V_{n,T}(k) + k\tilde{q}(\alpha)\tilde{\pi}(n,T)^{(1-\alpha)}, \tag{8}$$

with respect to $k \in \{0,1,\dots,<\infty\}$.

Remark 1. The level α can be interpreted as a parameter that controls for the trade-off between the accuracy of $\tilde{q}(\alpha)$ and the (asymptotic) inefficiency of $PC_\alpha(k)$. The presence of $(1-\alpha)$ in the subscript of $\tilde{\pi}(n,T)$ avoids reverting to extreme value theory. In contrast to Kapetanios (2010), the consistency of our dimensionality selection procedure does not require any specific restriction on the asymptotic behavior of α proportionally to n and T.

But because estimating the components $q(\alpha)$, and $\pi(n,T)$ in $L_{n,T}(\alpha)$ from the subsamples requires an a-priori information about r and estimating r requires an estimate of $L_{n,T}(\alpha)$, we propose a stepwise backwards procedure that we describe by the following Algorithm.

Algorithm (Backwards Procedure)

1. Choose a sufficiently large finite number k_{max} as a starting value for r.
2. Use k_{max} to calculate an estimate $\tilde{\pi}(n,T)$ for $\pi(n,T)$.
3. Use $\tilde{\pi}(n,T)$ to calculate an estimate $\tilde{q}(\alpha)$ for $q(\alpha)$.
4. Use $\tilde{\pi}(n,T)$ and $\tilde{q}(\alpha)$ to calculate an estimate $\hat{r}$ for r.

Optionally, one can replace the starting value k_{max} in Step 1 by $\hat{r}$ and repeat 1-4 until stabilization of the estimates.

2.3. A Subsampling Strategy to Estimate the Penalty Components

In order to retain both the temporal and the cross-sectional structure existing in the data, we construct blocks of sizes $s(T) < T$ and $s(n) < n$ in such a way that the orderings of the observations in each block are preserved. To simplify the discussion, we consider the subsampling function to be of the form $s(x) = \lfloor x^s \rfloor$, where $0 < s < 1$ and $\lfloor x \rfloor$ denotes the largest integer $\leq x$. The maximal possible number of blocks that we can construct from the original sample is $S_{n,T} = (T - s(T) + 1)(n - s(n) + 1)$.

To simplify the discussion, we present our estimators, in Sections 2.3.1 and 2.3.2, in a backwards logic, i.e., we construct an estimator of $\pi(n,T)$ as if we know r and an estimator of $q(\alpha)$ as if we know r and $\pi(n,T)$. In Section 2.3.2, we show that $\hat{r}$ is a consistent estimator if the starting value is larger or equal to r with a probability approaching 1, as $n, T \to \infty$.

2.3.1. Estimating the $\boldsymbol{(1-\alpha)}$*-quantile* $\boldsymbol{q(\alpha)}$

For given r and $\pi(.,.)$, define the statistic

$$Q^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor} \quad = \quad \pi(\lfloor n^s \rfloor, \lfloor T^s \rfloor)^{-1} \phi^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(r+1), \tag{9}$$

where $\phi^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(r+1) = V^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(r) - V^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(r+1)$ and $V^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(k)$ are based on the observations of the subsample $j \in \{1, \ldots, S_{n,T}\}$, for $k = r, r+1$.

The empirical distribution of $Q^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}$ is obtained by

$$\tilde{K}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(x) = \frac{1}{S_{n,T}} \sum_{j=1}^{S_{n,T}} \mathbf{I}\big(Q^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor} \leq x\big), \tag{10}$$

where $\mathbf{I}(.)$ is the indicator function.

Theorem 1. *Under Assumptions 1- 3 with* $\lfloor T^s \rfloor / T \to 0$ *and* $\lfloor n^s \rfloor / n \to 0$*, as* $\lfloor T^s \rfloor, \lfloor n^s \rfloor, T, n \to \infty$*, we have, for given* r *and* $\pi(\lfloor n^s \rfloor, \lfloor T^s \rfloor)$*,*

$$\tilde{K}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}\Big(x(\alpha)\Big) - K_{r+1}\Big(x(\alpha)\Big) = o_p(1), \tag{11}$$

where $x(\alpha) = \inf\{x : K_{r+1}(x) = 1 - \alpha\}$*.*

Corollary 1. *Under the Assumptions of Theorem 1 together with the assumptions that* $K_{r+1}(x)$ *is strictly increasing at* $x(\alpha)$*, we have*

$$\tilde{K}^{-1}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(1 - \alpha) = K^{-1}_{r+1}(1 - \alpha) + o_p(1), \tag{12}$$

for each $\alpha \in (0, 1)$*, where* $G^{-1}(.)$ *indicates the quantile function.*

A natural estimator of $q(\alpha)$ can be, hence, obtained by

$$\tilde{q}(\alpha) = \inf\{x : \tilde{K}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(x) \geq 1 - \alpha\}, \tag{13}$$

where $\alpha \in (0, 1)$.

2.3.2. Estimating the Order of Magnitude $\boldsymbol{\pi(n, T)}$

The idea of estimating the order of magnitude is based on the accommodating form of $\pi(n,T) = \big(\min\{n,T\}\big)^{\beta}$ and the result of Corollary 1. In fact, following Bertail et al. (1999) and multiplying both sides in (12) with $\pi(\lfloor n^s \rfloor, \lfloor T^s \rfloor) = \big(\min\{\lfloor n^s \rfloor, \lfloor T^s \rfloor\}\big)^{\beta}$, we get

$$\tilde{q}^*_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(\alpha) = q(\alpha)\Big(\min\{\lfloor n^s \rfloor, \lfloor T^s \rfloor\}\Big)^{\beta} + o_p\Big(\big(\min\{\lfloor n^s \rfloor, \lfloor T^s \rfloor\}\big)^{\beta}\Big), \tag{14}$$

where $\tilde{q}^*_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(\alpha) = \pi(\lfloor n^s \rfloor, \lfloor T^s \rfloor)\tilde{q}(\alpha)$. Note that, for given r, the term $\tilde{q}^*(\alpha)$ does not require $\pi(\lfloor n^s \rfloor, \lfloor T^s \rfloor)$ to be known since it corresponds to the empirical $(1 - \alpha)$- quantile of the statistic $\phi^{(j)}_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(r + 1)$,which is fully observed.

By taking the logarithm on both sides, we get

$$\log\Big(\tilde{q}^*_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(\alpha)\Big) = \log\Big(q(\alpha)\Big) + \beta \log\Big(\min\{\lfloor n^s \rfloor, \lfloor T^s \rfloor\}\Big) + o_p(1). \tag{15}$$

From (15), we can see that by constructing a second estimate $\tilde{q}^*_{\lfloor n^{s'} \rfloor, \lfloor T^{s'} \rfloor}(\alpha) \neq 0$ based on a sample size, say $(\lfloor n^{s'} \rfloor, \lfloor T^{s'} \rfloor)$, smaller than $(\lfloor n^s \rfloor, \lfloor T^s \rfloor)$, we can interpolate an estimate for β. Explicitly,

$$\tilde{\beta} = \frac{\log\Big(\tilde{q}^*_{\lfloor n^s \rfloor, \lfloor T^s \rfloor}(\alpha) / \tilde{q}^*_{\lfloor n^{s'} \rfloor, \lfloor T^{s'} \rfloor}(\alpha)\Big)}{\log\Big(\min\{\lfloor n^s \rfloor, \lfloor T^s \rfloor\} / \min\{\lfloor n^{s'} \rfloor, \lfloor T^{s'} \rfloor\}\Big)}. \tag{16}$$

Replacing β in $\pi(n,T)$ by $\tilde{\beta}$, we get the following estimator for the convergence rate:

$$\tilde{\pi}(n, T) = \big(\min\{n, T\}\big)^{\tilde{\beta}}. \tag{17}$$

Lemma 1. *In addition to the Assumptions of Corollary 1, let* $\lfloor T^{s'} \rfloor$ *and* $\lfloor n^{s'} \rfloor$ *be chosen such that* $\lfloor T^{s'} \rfloor, \lfloor n^{s'} \rfloor \to \infty$ *and* $\lfloor T^{s'} \rfloor / T^s, \lfloor n^{s'} \rfloor / n^s \to 0$*, as* $n, T \to 0$ *and* α *such that* $\tilde{q}^*_{\lfloor n^{s'} \rfloor, \lfloor T^{s'} \rfloor}(\alpha) \neq 0$*. Then*

$$\tilde{\beta} - \beta = o_p\bigg(\log\Big(\frac{\min\{\lfloor n^s \rfloor, \lfloor T^s \rfloor\}}{\min\{\lfloor n^{s'} \rfloor, \lfloor T^{s'} \rfloor\}} \Big)^{-1} \bigg)$$

and

$$\tilde{\pi}(n, T) = \pi(n, T)(1 + o_p(1)).$$

Theorem 2. *Let* $\tilde{\pi}(n, T)$ *and* $\tilde{q}(\alpha)$ *estimated in a Stepwise Backwards Procedure with* $k_{start} \geq r$ *as described in Section 2.3, then under Assumptions 1- 3 together with* $\lfloor T^s \rfloor / T \to 0$ *and* $\lfloor n^s \rfloor / n \to 0$*, as* $\lfloor T^s \rfloor, \lfloor n^s \rfloor, T, n \to \infty$*, we have*

$$\hat{r} - r = o_p(1)$$

.

All proofs are given in the Appendix.

3. Monte Carlo Simulations

The goal of this section is to compare, through Monte Carlo experiments, the performance of our dimensionality criterion with the existing selection procedure. We consider 5 different criteria: the panel criteria PC1 and IC1 of Bai and Ng (2002), the panel cointegration criterion IPC1 proposed by Bai (2004), the threshold criterion ED of Onatski (2010), and the information criterion $IC^T_{1;n}$ of Hallin and Liška (2007). The maximal number of factors used in PC1, IC1, IPC1, ED, and $IC^T_{1;n}$ is also set to 8. The calibration strategy of Hallin and Liška (2007) (second "'stability interval"' procedure) is applied on a grid interval of length 128 with the borders 0.01 and 3 as they have suggested.

Our data generating processes are based on the following panel data model:

$$Y_{it} = h(X_{it}, \beta)(1 - c) + c \sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it},$$

for all $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$, where X_{pit} are the observed regressors, f_{lt} are the factors to be estimated, λ_{il} are the corresponding loading parameters, $c \in [0, 1]$ controls for the weight of the factor structure in the model, and ϵ_{it} is the idiosyncratic error term.

The examined panel sets are the outcomes of 6 different DGPs:

AppendixA. Proofs

Lemma 2. *Let* $U_{n,T} = \min_{k\leq r}\{\mu(k)\tau_k(n,T)\}$ *and* $L_{n,T}(\alpha) = q(\alpha)\pi(n,T)^{1-\alpha}$, *where* $q(\alpha) = \inf_x\{x : K_{r+1}(x) \geq 1-\alpha\}$, *with* $\alpha \in (0,1)$, *and , then there exist twosome* $(n^{\alpha}_{a,b}, T^{\alpha}_{a,b})$ *such that* $U_{n,T} \geq L_{n,T}(\alpha)$, *for all* $n \geq n^{\alpha}_{a,b}$ *and* $T \geq T^{\alpha}_{a,b}$.

Proof of Lemma 2. Let $q(\alpha) = \inf_x\{x : K_{r+1}(x) \geq 1-\alpha\}$, then there exists, for each $\epsilon \in (0,\alpha]$, a finite $q(\epsilon) \geq q(\alpha)$ such that $K_{r+1}(q(\epsilon)) \leq 1-\epsilon$. Because $\pi(n,T)^{1-\epsilon}$ is strictly decreasing in n and T, for any $\epsilon \in (0,\alpha]$ and $\alpha \in (0,1)$, and by assumption $\min_k\{\tau_k(n,T)\mu(k)\}$ is strictly positive and non-decreasing in n and T, then $\lim_{n,T\to\infty} L_{n,T}(\epsilon)/U_{n,T} = 0$. This completes the proof. □

Proof of Proposition 1. In order to prove that $\lim_{n,T\to\infty} P(\hat{r} = r) = 1$, under the conditions of Proposition 1, we need to show that the probability of the complement converges to zero, i.e., $\lim_{n,T\to\infty} P\Big(PC_{L,U}(k) < PC_{L,U}(r)\Big) = 0$, for all $k \in \{\underline{k}, \overline{k}\}$, where $\underline{k} = 0,\dots,r-1$, and $\overline{k} = r+1,\dots,k_{max} < \infty$.

Recall that

$$PC_{L,U}(k) - PC_{L,U}(r) = V_{n,T}(k) - V_{n,T}(r) - (r-k)g_{n,T}(L,U) \tag{A.1}$$

By adding and subtracting the terms $\mathcal{V}_{n,T}(k)$ and $\mathcal{V}_{n,T}(r)$ in (A.1), for $k = \underline{k}$, we get

$$\begin{aligned} PC_{L,U}(\underline{k}) - PC_{L,U}(r) &= V_{n,T}(\underline{k}) - V_{n,T}(r) - (r-\underline{k})g_{n,T}(L,U) \\ &= \Big(V_{n,T}(\underline{k}) - \mathcal{V}_{n,T}(\underline{k})\Big) + \mathcal{V}_{n,T}(\underline{k}) \\ &\quad -\Big(V_{n,T}(r) - \mathcal{V}_{n,T}(r)\Big) - \mathcal{V}_{n,T}(r) \\ &\quad -(r-\underline{k})g_{n,T}(L,U). \end{aligned} \tag{A.2}$$

We known from (5) that $V_{n,T}(\underline{k}) - \mathcal{V}_{n,T}(\underline{k}) = o_p(1)$ and $V_{n,T}(r) - \mathcal{V}_{n,T}(r) = o_p(1)$. Multiplying both sides in (A.1) by $\tau_{\underline{k}}(n,T)^{-1}$, and exploiting the fact that $\tau_{\underline{k}}(n,T)$ is strictly positive and not decreasing in n and T, we obtain

$$\begin{aligned} \tau_{\underline{k}}(n,T)^{-1}\Big(PC_{L,U}(\underline{k}) - PC_{L,U}(r)\Big) &= \tau_{\underline{k}}(n,T)^{-1}\Big(\mathcal{V}_{n,T}(\underline{k}) - \mathcal{V}_{n,T}(r) \\ &\quad -(r-\underline{k})g_{n,T}(L,U)\Big) + o_p(1) \\ &= \tau_{\underline{k}}(n,T)^{-1}\Big(\varphi_{n,T}(\underline{k}+1) \\ &\quad -(r-\underline{k})g_{n,T}(L,U)\Big) + o_p(1). \end{aligned} \tag{A.3}$$

Let's, now, focus on the right-hand side of (A.3) and define

$$\mathcal{Z}_{n,T}(\underline{k}) = \tau_{\underline{k}}(n,T)^{-1}\Big(\varphi_{n,T}(\underline{k}+1) - (r-\underline{k})g_{n,T}(L,U)\Big).$$

If $g_{n,T}(L,U) \leq U_{n,T}$, for sufficiently large n and T, it follows

$$\begin{aligned}
\lim_{n,T\to\infty} P\Big(\mathcal{Z}_{n,T}(\underline{k})<0\Big) &\leq \lim_{n,T\to\infty} P\Big(\tau_{\underline{k}}(n,T)^{-1}\varphi_{n,T}(\underline{k}+1) < (r-\underline{k})\tau_{\underline{k}}(n,T)^{-1}U_{n,T}\Big)\\
&= \lim_{n,T\to\infty} P\Big(\tau_{\underline{k}}(n,T)^{-1}\varphi_{n,T}(\underline{k}+1) < (r-\underline{k})\tau_{\underline{k}}(n,T)^{-1}\min_{k\leq r}\{\mu(k)\tau_k(n,T)\}\Big)\\
&\leq \lim_{n,T\to\infty} P\Big(\tau_{\underline{k}}(n,T)^{-1}\varphi_{n,T}(\underline{k}+1) < (r-\underline{k})\tau_{\underline{k}}(n,T)^{-1}\mu(\underline{k})\tau_{\underline{k}}(n,T)\Big)\\
&= \lim_{n,T\to\infty} P\Big(\tau_{\underline{k}}(n,T)^{-1}\tfrac{1}{(r-\underline{k})}\varphi_{n,T}(\underline{k}+1) < \mu(\underline{k})\Big)\\
&\leq \lim_{n,T\to\infty} P\Big(\tau_{\underline{k}}(n,T)^{-1}\tfrac{1}{(r-\underline{k})}\inf_{m\geq n,\, S\geq T}\varphi_{m,S}(\underline{k}+1) \leq \mu(\underline{k})\Big)\\
&= 0.
\end{aligned} \tag{A.4}$$

The last equality follows directly from Assumption 2.

Because

$$\tau_{\underline{k}}(n,T)^{-1}\Big(PC_{L,U}(\underline{k}) - PC_{L,U}(r)\Big) \xrightarrow{p} \mathcal{Z}_{n,T}(\underline{k}), \tag{A.5}$$

the asymptotic equivalence implies

$$\lim_{n,T\to\infty} P\Big(\tau_{\underline{k}}(n,T)^{-1}\big(PC_{L,U}(\underline{k}) - PC_{L,U}(r)\big) < 0\Big) = 0, \tag{A.6}$$

and hence

$$\lim_{n,T\to\infty} P\Big(PC_{L,U}(\underline{k}) - PC_{L,U}(r) < 0\Big) = 0, \tag{A.7}$$

which completes the first part of the proof. To complete the second part, it thus remains to show that $\lim_{n,T\to\infty} P\Big(PC_{L,U}(\overline{k}) - PC_{L,U}(r) < 0\Big) = 0$, for all $\overline{k} \in \{r+1,\ldots,k_{max}\}$.

Because $V_{n,T}(0) - V_{n,T}(1) \geq V_{n,T}(1) - V_{n,T}(2) \geq \cdots \geq 0$ for all n and T, we have $V_{n,T}(\overline{k}) - V_{n,T}(r) \geq -(\overline{k}-r)(V_{n,T}(r+1) - V_{n,T}(r))$. Hence,

$$PC_{L,U}(\overline{k}) - PC_{L,U}(r) \quad \geq \quad -(\overline{k}-r)\phi_{n,T}(k_{r+1}) + (\overline{k}-r)g_{n,T}(L,U), \tag{A.8}$$

From (A.8), and Lemma 2, we get

$$\begin{aligned}
\lim_{n,T\to\infty} P\Big(PC_{L,U}(\overline{k}) - PC_{L,U}(r) \geq 0\Big) &\geq \lim_{n,T\to\infty} P\Big(-\phi_{n,T}(r+1) + q(\alpha)\pi(n,T)^{1-\alpha} \geq 0\Big)\\
&= \lim_{n,T\to\infty} P\Big(\phi_{n,T}(r+1) \leq q(\alpha)\pi(n,T)^{1-\alpha}\Big)\\
&= \lim_{n,T\to\infty} P\Big(\pi(n,T)^{-1}\phi_{n,T}(r+1) \leq q(\alpha)\pi(n,T)^{-\alpha}\Big).
\end{aligned} \tag{A.9}$$

Since $\pi(n,T)^{-\alpha} \to \infty$, as $n,T\to\infty$ for any $\alpha \in (0,1)$, then

$$\lim_{n,T\to\infty} P\Big(PC_{L,U}(\overline{k}) - PC_{L,U}(r) \geq 0\Big) = 1.$$

This completes the proof. $\square$

Proof of Theorem 1. Define the exact distribution function of $\pi(n,T)^{-1}\vartheta_{n,T}(r+1)$ as

$$K_{n,T,r+1}(x) = P\left[\pi(n,T)^{-1}\vartheta_{n,T}(r+1) \leq x\right]. \tag{A.10}$$

We denote by $p[.]$ the probability measure defined in the probability space of $\{e_{it}\}_{i,\in\mathbb{N},\, t\in\mathbb{Z}}$. The empirical distribution of $\pi(n,T)^{-1}\phi_{n,T}(r+1)$ based on the $S_{n,T}$ subsamples is given by

$$\tilde{K}_{n^s,T^s,r+1}(x) = \frac{1}{S_{n,T}}\sum_{j=1}^{S_{n,T}} \mathbf{I}\big(\pi(n^s,T^s)^{-1}\phi^{(j)}_{n^s,T^s}(r+1) \leq x\big), \tag{A.11}$$

where $\mathbf{I}(.)$ is the indicator.

Let $\vartheta^{(j)}_{n^s,T^s}(r+1)$ denote the random variable $\vartheta_{n,T}(r+1)$ obtained from the same subsample constructed for $\phi^{(j)}_{n^s,T^s}(r+1)$. Since by assumption $n^s, T^s \to \infty$ and $\pi(n^s,T^s)^{-1}\big(\phi^{(j)}_{n^s,T^s}(r+1) - \vartheta^{(j)}_{n^s,T^s}(r+1)\big) = o_p(1)$, then adding and subtracting $\vartheta^{(j)}_{n^s,T^s}(r+1)$ in (A.11) implies

$$\tilde{K}_{n^s,T^s,r+1}(x) = \frac{1}{S_{n,T}}\sum_{j=1}^{S_{n,T}} \mathbf{I}\big(\pi(n^s,T^s)^{-1}\vartheta^{(j)}_{n^s,T^s}(r+1) \leq x\big) + o_p(1). \tag{A.12}$$

Let's, now, focus on the first term on the right-hand side of (A.12) and define

$$\tilde{U}_{n^s,T^s,r+1}(x) = \frac{1}{S_{n,T}}\sum_{j=1}^{S_{n,T}} \mathbf{I}\big(\pi(n^s,T^s)^{-1}\vartheta^{(j)}_{n^s,T^s}(r+1) \leq x\big). \tag{A.13}$$

The principle of the asymptotic equivalence tells us that to prove the assertion of Theorem 1, it is sufficient to prove that $\tilde{U}_{n^s,T^s,r+1}(x)$ converges in probability to $K_{n^s,T^s,r+1}(x)$; since, by Assumption 2, $K_{n^s,T^s,r+1}(x)$ converges to $K_{r+1}(x)$, as $n,T\to\infty$.

Because the subsampling strategy is based on sampling blocks without changing the cross-section and the temporal ordering of the variables, stationarity and cross-section homogeneity imposed in Assumption 1 implies that $E\Big(\tilde{U}_{n^s,T^s,r+1}(x)\Big) = K_{n^s,T^s,r+1}(x)$. Since $K_{n^s,T^s,r+1}(x) \to K_{r+1}(x)$, as $n,T\to\infty$, it remains to prove that $Var\Big(\tilde{U}_{n^s,T^s,r+1}(x)\Big) \to 0$, as $n,T\to\infty$. For this purpose, define

$$I^j_{n^s,T^s,r+1} = \mathbf{I}\big(\pi(n^s,T^s)^{-1}\vartheta^{(j)}_{n^s,T^s}(r+1) \leq x\big) \tag{A.14}$$

and

$$\Gamma_{n^s,T^s,r+1}(h) = \frac{1}{S_{n,T}}\sum_{j=h+1}^{S_{n,T}} Cov(I^j_{n^s,T^s,r+1}, I^{j+h}_{n^s,T^s,r+1}). \tag{A.15}$$

By homogeneity assumption, the total variance of $\tilde{U}_{n^s,T^s,r+1}(x)$ is hence

$$Var\Big(\tilde{U}_{n^s,T^s,r+1}(x)\Big) = \tfrac{1}{S_{n,T}}\Big(\Gamma_{n^s,T^s,r+1}(0) + 2\textstyle\sum_{h=1}^{S_{n,T}-1}\Gamma_{n^s,T^s,r+1}(h)\Big).$$

Let $S^*_{n,T} = \max\{(n-n^s+1), (T-T^s+1)\}$, then we can write $Var\Big(\tilde{U}_{n^s,T^s,r+1}(x)\Big) = A_1 + A_2$, where

$$\begin{aligned} A_1 &= \tfrac{1}{S_{n,T}}\Big(\Gamma_{n^s,T^s,r+1}(0) + 2\sum_{h=1}^{S^*_{n,T}-1}\Gamma_{n^s,T^s,r+1}(h)\Big) \\ &= O\big(\tfrac{1}{S_{n,T}}\big) + O\big(\tfrac{S^*_{n,T}}{S_{n,T}}\big) = o(1) \end{aligned}$$

and

$$A_2 = \frac{1}{S_{n,T}}\sum_{h=S^*_{n,T}}^{S_{n,T}-1}\Gamma_{n^s,T^s,r+1}(h).$$

By the strong-mixing inequality for the covariance between two bounded random variables (see, e.g. Politis and Romano (1994)), a close inspection of A_2 shows that

$$|A_2| \leq Const.\frac{1}{S_{n,T}}\sum_{\xi=1}^{S^*_{n,T}}\xi\alpha_e(\xi)$$

Assumption 1 implies $A_2 = o(1)$, which provides $Var\Big(\tilde{U}_{n^s,T^s,r+1}(x)\Big) \to 0$, as $n, T \to \infty$. This completes the proof. $\square$

Proof of Corollary 1. To prove the assertion of Corollary 1 follow the same argument of Bertail et al. (1999).

Proof of Lemma 1. The first part of Lemma 1 is an immediate result of Corollary 1 and Equation (16) if the subsamples are chosen in such a way that following properties $T^{s'}/T^s, n^{s'}/n^s \to 0$, as $n, T \to 0$, are satisfied. In fact, equation (15) is a consequence of Lemma 2. Using it for $\min\{n^s, T^s\}$ and $\min\{n^{s'}, T^{s'}\}$ and taking the difference, we get

$$\log\Big(\hat{q}^u_{n^s,T^s}(\alpha)\Big) - \log\Big(\hat{q}^u_{n^{s'},T^{s'}}(\alpha)\Big) = \beta_{r+1}\bigg(\log\Big(\min\{n^s,T^s\}\Big) - \log\Big(\min\{n^{s'},T^{s'}\}\Big)\bigg) + o_p(1).$$

Dividing by $\bigg(\log\Big(\min\{n^s,T^s\}\Big) - \log\Big(\min\{n^{s'},T^{s'}\}\Big)\bigg)$ and rearranging, we get

$$\tilde{\beta}_{r+1} = \beta_{r+1} + o_p\bigg(\log\Big(\frac{\min\{n^s,T^s\}}{\min\{n^{s'},T^{s'}\}}\Big)^{-1}\bigg).$$

This completes the first part of the proof.

Dividing $\hat{\kappa}^*_{r+1}(n,T)$ by $\kappa^*_{r+1}(n,T)$, we get

$$\hat{\kappa}^*_{r+1}(n,T)/\kappa^*_{r+1}(n,T) = \big(\min\{n,T\}^{\hat{\beta}_{r+1}}\big)/\big(\min\{n,T\}^{\beta_{r+1}}\big)$$

Replacing $\hat{\beta}_{r+1}$ with β_{r+1}, we get

$$\hat{\kappa}^*_{r+1}(n,T)/\kappa^*_{r+1}(n,T) = 1 + o_p(1)$$

since $\frac{\min\{n^s,T^s\}}{\min\{n^{s'},T^{s'}\}} \to \infty$. This completes the proof. $\square$

PROOF OF THEOREM 2. The proof follows directly from Lemmas 2 and 1 since $\hat{\kappa}^*_{r+1}(n,T)/\kappa^*_{k_{max}}(n,T) = 1 + o_p(1)$. Consistency of $\tilde{q}(\alpha)$ is not required to prove the consistency of $\hat{r}$ since $\tilde{q}(\alpha)\kappa^*_{r+1}(n,T)^{1-\alpha} = o_p(1)$ and $q(\alpha)\kappa^*_{r+1}(n,T)^{-\alpha}$ diverges with n and T for any $\alpha \in (0,1)$.

Ahn, S., Lee, Y. H., Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. Journal of Econometrics 174 (1), 1 – 14.

Alessi, L., Barigozzi, M., Capasso, M., Dec. 2010. Improved penalization for determining the number of factors in approximate factor models. Statistics and Probability Letters 80 (23-24), 1806–1813.

Bada, O., Kneip, A., 2013. Parameter Cascading for Panel Models with Unknown Number of Heterogeneous Factors : An Application on the Credit Spread Puzzle. Forthcoming in Computational Statistics and Data and Analysis (0).

Bai, J., 2003. Inferential Theory for Factor Models of Large Dimensions. Econometrica 71 (1), 135–171.

Bai, J., 2004. Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics 122 (1), 137–183.

Bai, J., 2009. Panel Data Models With Interactive Fixed Effects. Econometrica 77 (4), 1229–1279.

Bai, J., Kao, C., Ng, S., Apr. 2009. Panel cointegration with global stochastic trends. Journal of Econometrics 149 (1), 82–99.

Bai, J., Ng, S., 2002. Determining the Number of Factors in Approximate Factor Models. Econometrica 70 (1), 191–221.

Bertail, P., Politis, D. N., Romano, J. P., 1999. On Subsampling Estimators with Unknown Rate of Convergence On Subsampling Estimators With Unknown Rate of Convergence. Journal of the American Statistical Association (94:446), 569–579.

Chamberlain, G., Rothschild, M., Sep. 1983. Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. Econometrica 51 (5), 1281.

Connor, G., Korajczyk, R. A., 1993. A Test for the Number of Factors in an Approximate Factor Model. The Journal of Finance 48 (4), 1263–1291.

Doukhan, P., 1994. Mixing: Properties and Examples. Springer-Verlag.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2014. The Generalized Dynamic-Factor Model: Identification and Estimation. The Review of Economics and Statistics 82 (4), 540–554.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102, 603–617.

Ibragimov, I. A., Rozanov, Y. A., 1978. Gaussian Random Processes. Springer New York.

Kapetanios, G., 2010. A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets. Journal of Business & Economic StatisticsEconomic Statistics 28 (3).

Kapetanios, G., Pesaran, M. H., Yamagata, T., Feb. 2011. Panels with non-stationary multifactor error structures. Journal of Econometrics 160 (2), 326–348.

Kneip, A., Sickles, R. C., Song, W., 2012. A New Panel Data Treatment for Heterogeneity in Time Trends. Econometric Theory 28, 590–628.

Onatski, A., 2009. Testing Hypotheses About the Number of Factors in Large Factor Models. Econometrica 77 (5), 1447–1479.

Onatski, A., 2010. Determining the Number of Factors from Empirical Distribution of Eigenvalues. The Review of Economics and Statistics 92 (4), 1004–1016.

Pesaran, H. M., 2006. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. Econometrica 74 (4), 967–1012.

Politis, D. N., Romano, J. P., 1994. Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions. The Annals of Statistics 22 (4), 2031–2050.

Stock, J. H., Watson, M. W., 1999. Diffusion Indexes. Revised version of NBER Working Paper #6702, August 1998.

Stock, J. H., Watson, M. W., 2005. Implications of Dynamic Factor Models for VAR Analysis. Working paper.