ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap

Thorsten Dickhaus *, Natalia Sirotko-Sibirskaya

Institute for Statistics, University of Bremen, Germany

ARTICLE INFO

Article history:
Received 5 September 2017
Received in revised form 22 February 2018
Accepted 14 August 2018
Available online xxxx

Keywords:
Bootstrap
Empirical Fourier transform
Family-wise error rate
Multiple hypothesis testing
Multivariate chi-square distribution
Wald-type statistic

ABSTRACT

Statistical inference methodology in dynamic factor models (DFMs) is extended to the multiple testing context based on a central limit theorem for empirical Fourier transforms of multivariate time series. This theoretical result allows for employing a vector of Wald-type test statistics which asymptotically follows a multivariate chi-square distribution under the global null hypothesis when the observation horizon tends to infinity. Multiplicity-adjusted asymptotic multiple test procedures based on Wald statistics are compared with a model-based bootstrap procedure proposed in recent previous work. Monte Carlo simulations demonstrate that both the asymptotic multiple chi-square test with an appropriate multiplicity adjustment and the bootstrap-based multiple test procedure keep the family-wise error rate approximately at the predefined significance level. The estimation algorithm as well as the implementation of the testing procedures is described in detail and a real-life application is performed on European commodity data.

1. Introduction and motivation

A dynamic factor model (DFM) is a multivariate time series model, where it is assumed that the observational process can be decomposed into the sum of latent common and idiosyncratic factors. The dynamic nature of the process is captured either by autocorrelation in common or idiosyncratic components, or both, or the dynamic influence of the common components on the observational process. The common factors are assumed to capture the significant part of the cross-correlation of the original time series, whereas the dynamics pertaining only to the individual series are contained in the idiosyncratic factors. Due to these characteristics, DFMs can be utilized as a dimension reduction tool as well as to provide meaningful interpretations of the dynamics driving certain observational processes. Because of their interpretability and modeling flexibility, DFMs have been widely employed in economics and finance; see, for example, Sargent and Sims (1977), Forni et al. (2000) and Stock and Watson (2011).

The parameters of a DFM can be estimated both parametrically and non-parametrically in the time as well as in the frequency domain. Classical time-domain estimation procedures employ maximum-likelihood-based methods such as the expectation maximization (EM) algorithm, see, e.g., Watson and Engle (1983), or non-parametric methods based on extracting principal components; see, for instance, Stock and Watson (2002). Recently, the frequency domain analog of the EM-based method has been proposed by Fiorentini et al. (2016) and principal components-based procedures have been extended to the frequency domain by Forni et al. (2000). An alternative parametric estimation method was suggested by Geweke (1977), Geweke and Singleton (1981) and represents an adaptation of the method originally developed for estimating the parameters of the covariance matrix of a static factor model by Lawley (1940) and Jöreskog (1967). Whereas

* Correspondence to: Institute for Statistics, University of Bremen, P. O. Box 330 440, 28344 Bremen, Germany.
E-mail address: dickhaus@uni-bremen.de (T. Dickhaus).

https://doi.org/10.1016/j.csda.2018.08.012
0167-9473/

methods based on the EM algorithm and principal components are traditionally used to estimate larger-scale DFMs, smaller-scale DFMs can be efficiently estimated via direct optimization of the likelihood function. In the present paper we consider small-scale DFMs, thus, we employ the method of direct optimization of the likelihood and provide a detailed description of its step-by-step implementation.

With the introduction of DFMs an important question has been raised as of deciding on the presence of dynamics in the common as well as in the idiosyncratic components. Correct model specification is especially crucial when the cross-sectional dimension is small as, e.g., neglecting the dynamics in the idiosyncratic factors may lead to erroneous model selection and to a subsequent misinterpretation of the model; cf. Maravall (1999) and Fiorentini et al. (2013). Testing procedures for model specification in the DFM context include likelihood-ratios (LR) tests, see Geweke and Singleton (1981), Lagrange multiplier (LM) test, see Watson and Engle (1983), Fernández (1990) and Fiorentini et al. (2013), as well as Wald tests, see Geweke and Singleton (1981). Fiorentini et al. (2013) offer an alternative LM testing approach to check the factors for autocorrelation. However, their method is initially developed for a single common factor case and has to be extended to the multiple common factor context first.

Whereas these methods allow testing each single factor for autocorrelation separately, in the present work we address the question of testing for autocorrelation of the factors simultaneously, thus, accounting for the multiplicity of the problem. To this end we extend the Wald test for the parameters of the spectral density matrix of the exact stationary DFM as in Geweke and Singleton (1981) to the multiple testing context. This extension is based on a multivariate central limit theorem in sequence space for empirical Fourier transforms of the observational process, see Dickhaus and Pauly (2016). Asymptotic normality of the Fourier transforms leads to asymptotic multivariate chi-square distributions for vectors of Wald statistics which can be used as test statistics in multiple test problems regarding the parameters of the spectral density of the observational process. Moreover, we compare the performance of such asymptotic tests based on Wald statistics with tests which are based on a bootstrap approximation of the finite-sample distribution of such vectors of test statistics, as outlined in Dickhaus and Pauly (2016). The idea is to contrast the generic approach which does not use the actual dependence structure with the bootstrap procedure which is based on replicating the dependence structure present in the data.

Thus, we address two important open problems of Dickhaus and Pauly (2016), namely, (i) the implementation of the proposed estimation and testing methodology, and (ii) the numerical comparison of the multivariate chi-square and the bootstrap approximations of the null distribution of the vector of test statistics. From the point of view of data analysis, our methodology can be used to address, among others, the following two problems.

Problem 1. Do the idiosyncratic factors have a non-trivial autocorrelation structure?

Problem 2. Do the common factors have a lagged influence on the observational process?

We will exemplify the proposed methodology by means of these two problems. The paper is organized as follows. Section 2 summarizes the statistical methodology underlying our work. For technical details, we refer to (Dickhaus and Pauly, 2016). We explain how vectors of Wald statistics arise in the context of DFMs when several linear hypotheses have to be tested simultaneously, as it is the case for Problems 1 and 2. Furthermore, the two approximation methods for the null distribution of such vectors (chi-square and bootstrap) are discussed. Section 3 describes the estimation of DFM parameters, Section 4 presents numerical results from simulation studies, and Section 5 is devoted to the analysis of real data. We conclude with a discussion in Section 6.

2. Statistical methodology

In this section, we summarize the statistical concepts underlying the work.

2.1. Dynamic factor model

We consider DFMs of the form

$$\mathbf{X}(t) = \sum_{s=-\infty}^{\infty} \Lambda(s)\,\mathbf{f}(t-s) + \boldsymbol{\varepsilon}(t),\ 1 \le t \le T, \tag{1}$$

where $\mathbf{X} = (\mathbf{X}(t) : 1 \le t \le T)$ denotes a p-dimensional, covariance-stationary stochastic process in discrete time with mean zero, $\mathbf{f}(t) = (f_1(t), \ldots, f_k(t))^\top$ with $k < p$ denotes a k-dimensional vector of so-called common factors and $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \ldots, \varepsilon_p(t))^\top$ denotes a p-dimensional vector of "specific" or "idiosyncratic" factors. We assume that the model dimensions p and k are fixed, while the observation horizon (i.e., sample size) T tends to infinity. As mentioned before, the underlying interpretation of (1) is that the dynamic behavior of the process $\mathbf{X}$ can be approximated by a lower-dimensional "latent" process $\mathbf{f}$. The entry (i, j) of the matrix $\Lambda(s)$ quantitatively reflects the influence of the jth common factor at lead or lag s, respectively, on the ith component of $\mathbf{X}(t)$, where $1 \le i \le p$ and $1 \le j \le k$. In particular, we consider predictable DFMs with a finite number S of lags, which are of the form

$$\mathbf{X}(t) = \sum_{s=0}^{S} \Lambda(s)\,\mathbf{f}(t-s) + \boldsymbol{\varepsilon}(t),\ 1 \le t \le T. \tag{2}$$

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

DFMs can be classified into "exact" and "approximate": in an exact DFM the idiosyncratic factors are assumed to be cross-sectionally uncorrelated, whereas in an approximate DFM one allows for weak cross-sectional correlation between idiosyncratic factors. In the present paper we restrict our attention to exact DFMs.

2.2. Multiple testing

The general setup of multiple testing theory assumes a statistical model $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ parametrized by $\vartheta \in \Theta$ and is concerned with testing a family $\mathcal{H} = (H_i : i \in I)$ of hypotheses regarding the parameter ϑ with corresponding alternatives $K_i = \Theta \setminus H_i$, where I denotes an arbitrary index set. We identify hypotheses with subsets of the parameter space throughout the paper. The intersection $H_0 = \bigcap_{i\in I} H_i$ is called the global hypothesis (in $\mathcal{H}$). Let $\varphi = (\varphi_i : i \in I)$ be a multiple test procedure for $\mathcal{H}$, meaning that each component φ_i, $i \in I$, is a (marginal) test for the test problem H_i versus K_i in the classical sense. Moreover, let $I_0 \equiv I_0(\vartheta) \subseteq I$ denote the index set of true hypotheses in $\mathcal{H}$ and $V(\varphi)$ the number of false rejections (type I errors) of φ, i.e., $V(\varphi) = \sum_{i\in I_0} \varphi_i$. The classical multiple type I error measure in multiple hypothesis testing is the family-wise error rate, FWER for short, and can (for a given $\vartheta \in \Theta$) be expressed as $\mathrm{FWER}_\vartheta(\varphi) = \mathbb{P}_\vartheta(V(\varphi) > 0)$. The multiple test φ is said to (strongly) control the FWER at a pre-defined significance level α, if $\sup_{\vartheta\in\Theta}\mathrm{FWER}_\vartheta(\varphi) \leq \alpha$. A simple, but often conservative method for FWER control is based on the union bound and is referred to as Bonferroni correction in the multiple testing literature. Assuming that $|I| = m$, the Bonferroni correction carries out each individual test φ_i, $i \in I$, at (local) level α/m. The "Bonferroni test" $\varphi = (\varphi_i : i \in I)$ then controls the FWER. Improvements of the Bonferroni procedure, which take dependencies among test statistics into account in order to define a multivariate multiple test, are major topics of modern multiple testing theory; see, for example, Dickhaus (2014). Two possibilities, which will be pursued in the remainder, are the exploitation of multivariate central limit theorems and the utilization of appropriate resampling schemes.

For the comparison of concurring multiple tests (keeping the same type I error rate), also a notion of power is required. To this end, denote by $S(\varphi) = \sum_{i\in I_1} \varphi_i$, where $I_1 = I \setminus I_0$, the (random) number of correctly rejected, false null hypotheses, and let $m_1 = |I_1|$ denote the total number of false null hypotheses in $\mathcal{H}$. One popular definition of a multiple power of φ, which we will use throughout the remainder, is then given by $\mathrm{power}_\vartheta(\varphi) = \mathbb{E}_\vartheta[S(\varphi)/\max(m_1, 1)]$; see Definition 1.4 of Dickhaus (2014).

2.3. Likelihood-based inference in dynamic factor models

In order to maintain a self-contained presentation, let us briefly summarize some essential techniques and results from previous literature.

Lemma 2.1 (*Lemma 2 of Dickhaus and Pauly, 2016*)**.** *Under a DFM, the spectral density matrix $S_{\mathbf{X}}$ of the observable process $\mathbf{X}$ can be decomposed as*

$$S_{\mathbf{X}}(\omega) = \tilde{\Lambda}(\omega) S_{\mathbf{f}}(\omega) \tilde{\Lambda}(\omega)' + S_{\boldsymbol{\varepsilon}}(\omega), \quad -\pi \leq \omega \leq \pi, \tag{3}$$

where $S_{\mathbf{f}}$ and $S_{\boldsymbol{\varepsilon}}$ are the spectral density matrices of the common and the idiosyncratic factors, respectively, $\tilde{\Lambda}(\omega) = \sum_{s=-\infty}^{\infty} \Lambda(s) \exp(-i\omega s)$, and the prime stands for transposition and conjugation.

The identifiability conditions mentioned in Section 1 and further discussed in Section 3 can be plainly phrased by postulating that the representation in (3) is unique (up to scaling). All further methods in this section rely on the assumption of an identified model and on asymptotic considerations as $T \to \infty$. To this end, we consider a scaled version of the empirical (finite) Fourier transform of $\mathbf{X}$. Evaluated at harmonic frequencies, it is given by

$$\tilde{\mathbf{X}}(\omega_j) = (2\pi T)^{-1/2} \sum_{t=1}^{T} \mathbf{X}(t) \exp(it\omega_j), \tag{4}$$

where $\omega_j = 2\pi j/T$, $-T/2 < j \leq \lfloor T/2 \rfloor$. For asymptotic inference with respect to T, we recall from Dickhaus and Pauly (2016) the following additional assumption.

Assumption 2.1. There exist B disjoint frequency bands $\Omega_1, \ldots, \Omega_B$, such that $S_{\mathbf{X}}$ can be assumed approximately constant and different from zero within each of these bands. Let $\omega^{(b)} \notin \{0, \pi\}$ denote the center of the band Ω_b, $1 \leq b \leq B$. As in Hannan (1973) and Geweke and Singleton (1981), we will denote by $n_b = n_b(T)$ a number of harmonic frequencies $(\omega_{j,b})_{1\leq j\leq n_b}$ of the form $2\pi j_u/T$ which are as near as possible to $\omega^{(b)}$, $1 \leq b \leq B$. In this, the integers j_u, $1 \leq u \leq n_b$, in $\omega_{j,b} = 2\pi j_u/T$ are chosen in successive order of closeness to the center.

Exploiting Assumption 2.1, the unknown model parameters in (3) are given by the $d = 2pk + k^2 + p$ distinct parameters in $\tilde{\Lambda}(\omega^{(b)})$, $S_{\mathbf{f}}(\omega^{(b)})$ and $S_{\varepsilon}(\omega^{(b)})$, for all $1 \leq b \leq B$. We denote by $\boldsymbol{\vartheta}_b$, $1 \leq b \leq B$, a vector of dimension d containing all these parameters. The parameter vector $\boldsymbol{\vartheta}_b$ in band Ω_b can be estimated by maximizing the asymptotic (complex Gaussian) likelihood function pertaining to the empirical Fourier transforms of $\mathbf{X}$, see Section 3. The estimation algorithm described in Section 3 delivers not only the numerical value of the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\vartheta}}_b$, but additionally an estimate

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

of the asymptotic covariance matrix of $\sqrt{n_b}\hat{\boldsymbol{\vartheta}}_b$. In view of standard results from likelihood theory (cf., e.g., Section 12.4 in Lehmann and Romano, 2005) concerning asymptotic normality of MLEs, it is assumed that

$$\sqrt{n_b}(\hat{\boldsymbol{\vartheta}}_b - \boldsymbol{\vartheta}_b) \xrightarrow{\mathcal{D}} \mathbf{T}_b \sim \mathcal{N}_d(\mathbf{0}, V_b), \quad 1 \leq b \leq B, \tag{5}$$

as $\min(n_b(T), T) \to \infty$, where the multivariate normal limit random vectors $\mathbf{T}_b$ are stochastically independent for $1 \leq b \leq B$, and that $\hat{V}_b$ is a consistent estimator of V_b. This is very helpful for testing linear (point) hypotheses. Such hypotheses are of the form $H : C\boldsymbol{\vartheta} = \boldsymbol{\xi}$ with a contrast matrix $C \in \mathbb{R}^{r \times Bd}$, $\boldsymbol{\xi} \in \mathbb{R}^r$ and $\boldsymbol{\vartheta}$ consisting of all elements of all the vectors $\boldsymbol{\vartheta}_b$. In Geweke and Singleton (1981) the usage of Wald statistics has been proposed in this context. The Wald statistic for testing H is given by

$$W = N(C\hat{\boldsymbol{\vartheta}} - \boldsymbol{\xi})^\top (C\hat{V}C^\top)^+ (C\hat{\boldsymbol{\vartheta}} - \boldsymbol{\xi}), \tag{6}$$

where $N = \sum_{b=1}^{B} n_b$, $\hat{V}$ is the block matrix built up from the band-specific matrices $N\hat{V}_b/n_b$, $1 \leq b \leq B$, and A^+ denotes the Moore–Penrose pseudo inverse of a matrix A.

Theorem 2.1 (*Theorem 2 of Dickhaus and Pauly, 2016*)**.** *Under the above assumptions, W is asymptotically χ^2-distributed with $\mathrm{rank}(C)$ degrees of freedom under the null hypothesis H, provided that V is positive definite and $N/n_b \leq K < \infty$ for all $1 \leq b \leq B$.*

Returning to the two exemplary simultaneous statistical inference problems outlined in Problems 1 and 2, it has been demonstrated by Dickhaus and Pauly (2016) that they can be formalized by families of linear hypotheses regarding (components of) $\boldsymbol{\vartheta}$. Hence, for each individual hypothesis a Wald statistic can be computed, the asymptotic null distribution of which is chi-squared according to Theorem 2.1. Notice, however, that these Wald statistics are typically dependent, leading to an involved multivariate chi-square distribution for the asymptotic joint null distribution of the vector of all these statistics; see Dickhaus and Royen (2015) for an overview of such multivariate distributions and their usage in multiple testing.

Lemma 2.2 (*Problem 1 Revisited; Lemma 3 of Dickhaus and Pauly, 2016*)**.** *In the notational framework of* Section 2.2, *Problem 1 can be formalized by setting $m = p$, and $I = \{1, \ldots, p\}$. For each $i \in I$, consider the linear hypothesis $H_i : C_{Dunnett}\, \mathbf{s}_{\boldsymbol{\varepsilon}_i} = 0$. The contrast matrix $C_{Dunnett}$ is the "multiple comparisons with a control" contrast matrix with $B-1$ rows and B columns, where in each row j the first entry equals $+1$, the $(j+1)$th entry equals -1 and all other entries are equal to zero. The vector $\mathbf{s}_{\boldsymbol{\varepsilon}_i} \in \mathbb{R}^B$ consists of the values of the spectral density matrix $S_{\boldsymbol{\varepsilon}}$ corresponding to the ith noise component, evaluated at the B centers $(\omega^{(b)} : 1 \leq b \leq B)$ of the chosen frequency bins. Denoting the subvector of $\hat{\boldsymbol{\vartheta}}$ that corresponds to $\mathbf{s}_{\boldsymbol{\varepsilon}_i}$ by $\hat{\mathbf{s}}_{\boldsymbol{\varepsilon}_i}$, the ith Wald statistic is given by*

$$W_i = (C_{Dunnett}\, \hat{\mathbf{s}}_{\boldsymbol{\varepsilon}_i})^\top \left[C_{Dunnett}\, \hat{V}_{\boldsymbol{\varepsilon}_i} C_{Dunnett}^\top \right]^+ (C_{Dunnett}\, \hat{\mathbf{s}}_{\boldsymbol{\varepsilon}_i}),$$

where $\hat{V}_{\boldsymbol{\varepsilon}_i} = diag(\hat{\sigma}^2_{\boldsymbol{\varepsilon}_i}(\omega^{(b)}) : 1 \leq b \leq B)$. Then, under H_i, W_i asymptotically follows a χ^2-distribution with $B-1$ degrees of freedom if the corresponding limit matrix $V_{\boldsymbol{\varepsilon}_i}$ is assumed to be positive definite. Considering the vector $\mathbf{W} = (W_1, \ldots, W_p)^\top$ of all p Wald statistics corresponding to the p specific factors in the model, we finally have that $\mathbf{W}$ asymptotically follows a multivariate chi-square distribution of the type considered in Section 5 of Dickhaus and Royen (2015) *with $B-1$ degrees of freedom in each marginal under the intersection H_0 of the p hypotheses $H_1, \ldots, H_p$.*

Lemma 2.3 (*Problem 2 Revisited; Lemma 4 of Dickhaus and Pauly, 2016*)**.** *As done in* Geweke and Singleton (1981)*, we formalize the null hypothesis that common factor j has a purely instantaneous effect on $\mathbf{X}_i$, $1 \leq j \leq k$, $1 \leq i \leq p$, in the spectral domain by*

$$H_{ij} : \left|\tilde{\Lambda}_{ij}\right|^2 \text{ is constant across the } B \text{ frequency bands}.$$

In an analogous manner to the considerations in Lemma 2.2, the contrast matrix $C_{Dunnett}$ can be used as the basis to construct a Wald statistic W_{ij}. The vector $\mathbf{W} = (W_{ij} : 1 \leq i \leq p, 1 \leq j \leq k)$ then asymptotically follows a multivariate chi-square distribution with $B-1$ degrees of freedom in each marginal under the global null hypothesis H_0, in analogy to the situation in Lemma 2.2.

Many other problems of practical relevance can be formalized analogously by making use of linear contrasts and thus, our framework applies to them, too. Furthermore, the hypotheses of interest may also refer to different subsets of $\{1, \ldots, B\}$. In such a case, the marginal degrees of freedom for the test statistics are not balanced.

Remark 2.1.

(a) The dependency structure among the components of $\mathbf{W}$ is very involved. Hence, it is infeasible to utilize multivariate chi-square quantiles as critical values for the Wald statistics. However, the computer simulations reported by Dickhaus (2012) indicate that generic multiple test procedures for positively dependent test statistics perform well in case of vectors of multivariate chi-square distributed test statistics. In particular, the multiplicity-adjustment by Hommel (1988) is appropriate for FWER control in this context. Hence, such multiple tests will be considered in Sections 4 and 5. These tests do not utilize the actual strength of the dependencies, but the qualitative fact that positive dependency among test statistics is present.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

(b) A different way to calibrate multiple tests based on vectors of Wald statistics as considered in Lemmas 2.2 and 2.3 is to approximate the finite-sample null distribution of $\mathbf{W}$ by means of appropriate resampling schemes. Resampling-based multiple tests for FWER control have been worked out for instance by Westfall and Young (1993), Troendle (1995) and Romano and Wolf (2005). In particular, it is well known that the convergence of Wald-type statistics to their asymptotic χ^2-distribution is rather slow, see Pauly et al. (2015) and Konietschke et al. (2015) and references therein. To address this problem and to make use of the actual dependency structure of $\mathbf{W}$ in the multiple test procedure, a model-based bootstrap approximation of the finite-sample distribution of $\mathbf{W}$ has been derived by Dickhaus and Pauly (2016); see Section 4 in their work for a detailed step-by-step description. The general idea behind the proposed bootstrap procedure is very much in the spirit of a classical parametric bootstrap. However, the re-samples are not generated from the (approximate) data distribution, but from the approximate (asymptotic) distribution of the parameter estimators assuming the DFM structure; see (5).

3. Implementation

Maximum-likelihood-based algorithms for estimation of static factor models have been developed in the 1940s by Lawley (1940, 1941, 1942). However, these algorithms of gradient-descent type are either inefficient in terms of the number of iterations needed for convergence, or do not converge at all even after a substantial number of iterations. In the 1960s Jöreskog and Lawley (1968) and Jöreskog (1967, 1969) suggested using a more efficient numerical procedure developed by Fletcher and Powell (1963). The Fletcher and Powell algorithm belongs to the class of quasi-Newton algorithms and is based on updating not only the gradient, but also the inverse of the matrix of second derivatives, which speeds up the algorithm and guarantees convergence in most of the cases. Moreover, the analytical expression for the matrix of second derivatives is not necessary as the approximation to it is built successively in the process of iterations. Likelihood-based estimation of dynamic factor models in the frequency domain, which is being employed in this paper, is performed by adapting the modification of the Fletcher–Powell algorithm to complex-valued estimation, see Geweke (1977) and Geweke and Singleton (1981).

The estimation steps of the DFM in the frequency domain can be summarized as follows: (i) model identification, (ii) estimation of the number of disjoint frequency bands, and (iii) estimation of the free parameters in the model. These tasks are addressed separately in the following subsections.

3.1. Model identification

The model identification issue in the DFM context is twofold. First, define

$$\delta(p,k) = p(p+1)/2 - [pk + k(k+1)/2 + p],$$

where $p(p+1)/2$ is the number of distinct elements in $S_{\mathbf{X}}$ and $pk+k(k+1)/2+p$ is the number of parameters in unrestricted $\tilde{\Lambda}$, $S_{\mathbf{f}}$ and S_{ε} assuming that the parameters are in $\mathbb{C}$. If the parameters are considered to be in $\mathbb{R}$, their dimension is doubled and the restrictions are imposed both on the real and imaginary parts. If $\delta(p,k) = 0$, i.e., one has as many equations as parameters, the model is identified uniquely. If $\delta(p,k) > 0$, i.e., there are more equations than parameters, there are no solutions. If, however, $\delta(p,k) < 0$, i.e., there are fewer equations than parameters, there are infinitely many choices of $\tilde{\Lambda}$, $S_{\mathbf{f}}$ and S_{ε}, see Lawley and Maxwell (1971).

Second, there is rotational indeterminacy in factor loadings. To see this, consider any nonsingular $k \times k$ matrix $M(\omega)$ and let $\Lambda^*(\omega) = \tilde{\Lambda}(\omega)M(\omega)$ and $S^*_{\mathbf{f}} = M(\omega)^{-1}S_{\mathbf{f}}(M(\omega)^{-1})'$. Then

$$\tilde{\Lambda}(\omega)S_{\mathbf{f}}(\omega)\tilde{\Lambda}(\omega)' = \Lambda^*(\omega)S^*_{\mathbf{f}}(\omega)\Lambda^*(\omega)'.$$

Therefore, one needs to impose k^2 restrictions corresponding to the number of elements in $M(\omega)$ to guarantee identification of $\tilde{\Lambda}$ and $S_{\mathbf{f}}$. In the absence of prior assumptions coming from the underlying scientific theory (e.g., economics or psychology, where factor models are widely applied), one may apply the identification schemes suggested in the literature for the case of orthogonal and oblique factors in order to resolve the rotational indeterminacy. When the factors are orthogonal, it is common to assume $S_{\mathbf{f}} = \mathbf{I}$ and $\tilde{\Lambda}S_{\varepsilon}^{-1}\tilde{\Lambda}'$ to be diagonal, whereas in the case of oblique factors one assumes $diag(S_{\mathbf{f}}) = \mathbf{I}$ and fixes at least $k-1$ entries in each column of $\tilde{\Lambda}$ at zero with the pattern of zeros being such that it cannot be destroyed by any non-singular transformation. Such patterns were first suggested by Thurstone (1947) and further developed by Reiersøl (1950), Howe (1955), Anderson and Rubin (1956) and Lawley (1958) for static factor models. Their suggested schemes can as well be adapted to the DFM set-up by requiring the identification restrictions to hold at each frequency ω, see Geweke (1977), Geweke and Singleton (1981), and Heaton and Solo (2004). We will provide details on the identification schemes used in the present work in Section 4.

3.2. Estimation of the number of disjoint frequency bands

Recall from Assumption 2.1 that we assume that B disjoint frequency bands exist such that $S_{\mathbf{X}}$ can be assumed to be approximately constant within each of these bands. Hence, prior to the estimation of the free parameters in the model one has to identify this number B. Since the spectrum is symmetric about the origin, it is sufficient to consider only the interval $[0,\pi)$.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

In order to perform the division of the harmonic frequencies into bands we use the methodology from Lavielle and Ludeña (2000) and Reschenhofer (2008). This methodology assumes that the spectrum can be approximated by a piecewise constant function. To this end, one needs to determine the locations and the number of "change-points", where we refer to the endpoint of a frequency band as a change-point. If the number $B-1$ of change-points is given, then their location is determined in two steps: first, one minimizes the so-called negative Whittle log likelihood function to determine the optimal height of a function for all possible locations of a change-point; second, one solves a combinatorial optimization problem in order to determine the "optimal" locations of the change points. If the number of the change-points is not given, the latter procedure is repeated for all the possible numbers and the optimal number is chosen by evaluating an information criterion such as the Akaike information criterion (AIC) or the Bayesian (Schwarz) information criterion (BIC).

First we demonstrate how the spectrum can be approximated by a piecewise-constant function for the univariate process and then we comment on how this procedure can be adapted to a multivariate case.

Let us assume that the piecewise constant function on the interval $[0, \pi)$ is defined by

$$g(\omega; a_1, \ldots, a_B, b_1, \ldots, b_{B-1}) = a_1 \mathbf{1}_{[0,b_1)}(\omega) + \ldots + a_B \mathbf{1}_{[b_{B-1},\pi)}(\omega),$$

where $b_1, \ldots, b_{B-1}$ are the change-points, resulting in B bands, namely, $\Omega_1 = [0, b_1), \ldots, \Omega_B = [b_{B-1}, \pi)$. The parameters $a_1, \ldots, a_B$ are the heights of the spectrum in each of these bands. Given any arbitrary configuration of change-points $b_1, \ldots, b_{B-1}$, the remaining parameters of this piecewise constant function, namely, $a_1, \ldots, a_B$, are estimated by minimizing the so-called negative Whittle log likelihood, given by

$$W \equiv W(a_1, \ldots, a_B, b_1, \ldots, b_{B-1}) = \int_{[0,\pi)} \log g(\omega; a_1, \ldots, a_B, b_1, \ldots, b_{B-1}) + \frac{I(\omega)}{g(\omega; a_1, \ldots, a_B, b_1, \ldots, b_{B-1})} d\omega,$$

where I denotes the periodogram. The "optimal" configuration of the change-points themselves is then determined in a combinatorial manner by minimizing W over the different possible configurations of $b_1, \ldots, b_{B-1}$.

In order to adapt this strategy to the approximation of a multivariate spectrum, it is possible to substitute the periodogram of the univariate process by the (spectral) norm of the periodogram of the multivariate process. We will pursue this strategy in Sections 4 and 5.

3.3. Estimation of the free parameters in the model

The asymptotic likelihood function of the parameter vector ϑ_b in frequency band $1 \leq b \leq B$ based on observed data $\mathbf{X} = \mathbf{x}$ is given by

$$\ell_b(\vartheta_b, \mathbf{x}) = \pi^{-p \times n_b} \, |\mathrm{ivech}(\vartheta_b)|^{-n_b} \, \exp\left(- \sum_{j=1}^{n_b} \tilde{\mathbf{x}}(\omega_{j,b})' \, \mathrm{ivech}(\vartheta_b)^{-1} \, \tilde{\mathbf{x}}(\omega_{j,b}) \right),$$

where $\vartheta_b = \mathrm{vech}(S_{\mathbf{X}}(\omega^{(b)}))$, and $\mathrm{ivech}(\vartheta_b) = S_{\mathbf{X}}(\omega^{(b)})$, see Goodman (1963). Taking logarithms and dropping the argument $\omega^{(b)}$, we obtain equivalently

$$\log \ell_b(\vartheta_b, \mathbf{x}) = n_b(-p \log \pi - \log|S_{\mathbf{X}}| - tr(\widehat{S} S_{\mathbf{X}}^{-1})), \tag{7}$$

where

$$\widehat{S} = (n_b)^{-1} \sum_{j=1}^{n_b} \tilde{\mathbf{x}}(\omega_{j,b}) \tilde{\mathbf{x}}(\omega_{j,b})'$$

is the unconstrained spectral density matrix estimate in band b, $1 \leq b \leq B$, see Geweke and Singleton (1981). Instead of maximizing (7), it is convenient to (equivalently) minimize the function

$$f_b(\vartheta_b, \mathbf{x}) = \log|S_{\mathbf{X}}| + tr(\widehat{S} S_{\mathbf{X}}^{-1}) - \log|S| - p, \tag{8}$$

as n_b times the minimum value of $f_b(\vartheta_b, \mathbf{x})$ provides the value of the likelihood ratio which can be used for testing purposes later, see (Jöreskog, 1967). Making use of the decomposition of $S_{\mathbf{X}}$ given in (3), maximization of $f_b(\cdot, \mathbf{x})$ requires solving a system of the following non-linear (in the parameters in ϑ_b) normal equations:

$$\begin{aligned}
\frac{\partial f_b(\vartheta_b, \mathbf{x})}{\partial \tilde{\Lambda}} &= 2 S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - \widehat{S}) S_{\mathbf{X}}^{-1} \tilde{\Lambda} S_{\mathbf{f}} = 0, \\
\frac{\partial f_b(\vartheta_b, \mathbf{x})}{\partial S_{\mathbf{f}}} &= 2 \tilde{\Lambda}' S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - \widehat{S}) S_{\mathbf{X}}^{-1} \tilde{\Lambda} = 0, \\
\frac{\partial f_b(\vartheta_b, \mathbf{x})}{\partial S_{\varepsilon}} &= diag(S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - \widehat{S}) S_{\mathbf{X}}^{-1}) = 0.
\end{aligned}$$

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

The latter optimization problem can be solved with the help of numeric solvers such as the nonlinear constrained optimization function *fmincon* in Matlab with sequential quadratic programming method, *sqp*. Specifying only the gradient analytically already leads to plausible estimation results. However, one has the possibility to add analytical expressions for the Hessian as well. Otherwise, the algorithm provides a positive definite quasi-Newton approximation of the Hessian in each iteration, which is constructed in Fletcher–Powell fashion. Since the parameter vector is complex-valued, in the estimation procedure it is decast into real and imaginary parts, see Geweke and Singleton (1981). This is a local optimization problem, hence, to ensure accurate results it is necessary to use appropriate initial estimates, see Jöreskog (1967).

An additional advantage of using this type of quasi-Newton algorithm is that the Hessian built up during the iterations when inverted and multiplied by $2/n_b$ gives an estimate of the variance–covariance matrix of the maximum likelihood estimates of the free parameters, see Lawley and Maxwell (1971), Lockhart (1967) and Jöreskog (1967). This estimate is, however, usually not precise, so it is advisable to recompute the second-order derivative matrix at the minimum and invert it to obtain more precise estimates.

4. Simulation studies

To address Problems 1 and 2 stated in Section 1, we first consider simulation scenarios with dynamics in either of the factors and then dynamics in both type of factors. First, we introduce a "benchmark" static factor model, Model 1, which is defined by

$$\mathbf{X}(t) = \Lambda(0)\,\mathbf{f}(t) + \boldsymbol{\varepsilon}(t),\ 1 \le t \le T, \tag{9}$$

where $\Lambda \equiv \Lambda(0) = (\lambda_{i,j})_{1\le i\le p,\,1\le j\le k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\mathbf{f}(t) \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$ and $\boldsymbol{\varepsilon}(t) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$. This means, we consider the case that $S = 0$ in (2). Next we introduce dynamics in the model by considering first the dynamics solely in the idiosyncratic factors, Model 2, and then dynamics solely in the common factors, Model 3. Factor dynamics are modeled and simulated as vector-autoregressive processes (VARs) defined as follows:

$$\mathbf{f}(t) = \Psi_1\mathbf{f}(t-1) + \cdots + \Psi_{q_\mathbf{f}}\mathbf{f}(t-q_\mathbf{f}) + \boldsymbol{\zeta}(t), \quad \boldsymbol{\zeta}(t) \sim \mathcal{N}_k(\mathbf{0}, \Sigma_\zeta),$$
$$\boldsymbol{\varepsilon}(t) = \Phi_1\boldsymbol{\varepsilon}(t-1) + \cdots + \Phi_{q_\varepsilon}\boldsymbol{\varepsilon}(t-q_\varepsilon) + \mathbf{u}(t), \quad \mathbf{u}(t) \sim \mathcal{N}_p(\mathbf{0}, \Sigma_u),$$

where $\underset{k\times k}{\Psi_i}$ and $\underset{p\times p}{\Phi_i} = diag$ are autoregressive coefficient matrices of common and idiosyncratic factors, respectively. Finally, in Model 4 we allow both common and idiosyncratic factors to be autocorrelated. The coefficients for all VAR processes are chosen in such a way that the roots of the characteristic polynomials are inside the unit circle, i.e., the processes are stationary which is required by model assumptions.

To solve the rotational indeterminacy it is necessary to adopt the following identification schemes for the orthogonal factors: $S_\mathbf{f} = \mathbf{I}$ and $\tilde{\Lambda} S_\varepsilon \tilde{\Lambda}' = diag$, and for the oblique factors: $diag(S_\mathbf{f}) = \mathbf{1}$ and $\tilde{\Lambda} = (\tilde{\Lambda}_1', \tilde{\Lambda}_2')'$, where $\tilde{\Lambda}_1$ is a $k \times k$ diagonal matrix with positive elements on the diagonal.

Moreover, we impose over-identifying restrictions by setting certain entries in the factor loading matrices $\Lambda(s), 0 \le s \le S$, to zero as follows

$$\Lambda(s)_{5\times 2} = \begin{bmatrix} \times & 0 \\ \times & 0 \\ \times & 0 \\ 0 & \times \\ 0 & \times \end{bmatrix}, \qquad \Lambda(s)_{10\times 3} = \begin{bmatrix} \times & 0 & 0 \\ \times & 0 & 0 \\ \times & 0 & 0 \\ 0 & \times & 0 \\ 0 & \times & 0 \\ 0 & \times & 0 \\ 0 & 0 & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix},$$

where $\times$ corresponds to a free parameter.

Here, we present Monte Carlo simulation results referring to Problems 1 and 2 under the four aforementioned models. The estimation results are presented in Figs. 1–4 (corresponding to Models 1 to 4), whereas the testing results are illustrated in Figs. 5–8 (Models 1 to 4), and summarized in Tables 1–4 (Models 1 to 4).

To illustrate the theoretical assumptions as well as the accuracy of the estimation results the left panels of Figs. 1–4 display the theoretical spectra of the common and idiosyncratic factors (dashed line and dotted line, respectively), as well as the overall theoretical spectrum associated with each model (solid line). The right panels of Figs. 1–4 present the estimation accuracy with the solid red line denoting the average norm of the theoretical spectrum and the dashed line denoting the average norm of the estimated spectrum. The solid black line in the right panel of Figs. 1–4 corresponds to the theoretical overall spectrum, and it is the same as in the corresponding left panel. The estimation results are averaged over 10,000 Monte Carlo simulation runs, therefore, the lines corresponding to the average norm of the theoretical spectrum and the average norm of the estimated spectrum in the right panels of Figs. 1–4 may appear indistinguishable. From these graphical

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

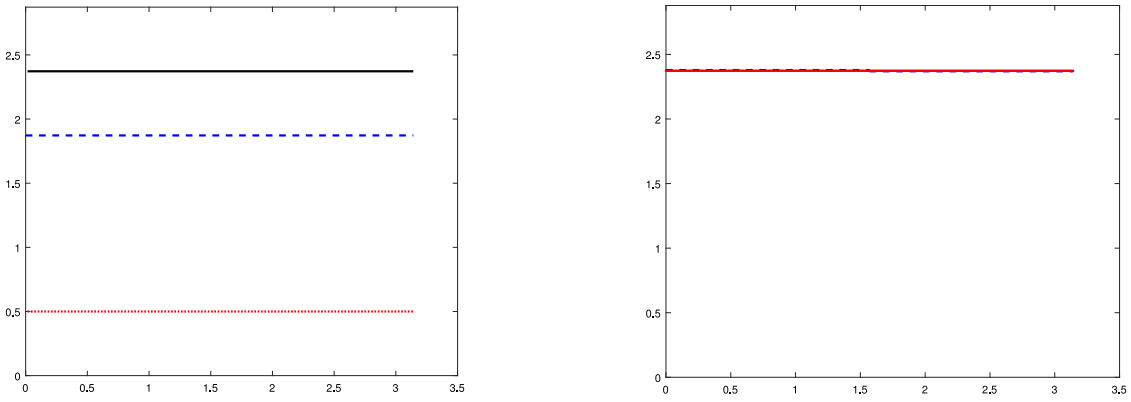

Fig. 1. Estimation results under Model 1. Left: The norms of theoretical spectra of common factors (dashed line), idiosyncratic factors (dotted line), and overall spectrum (solid line). Right: Estimation results for 10,000 simulation runs. Norm of the theoretical spectrum (solid black line), average norm of theoretical spectrum (solid red line), and average norm of estimated spectrum (dashed line). Note: The lines almost coincide in the right-hand side graph. Model parameters: $p = 5, k = 2, T = 1000$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

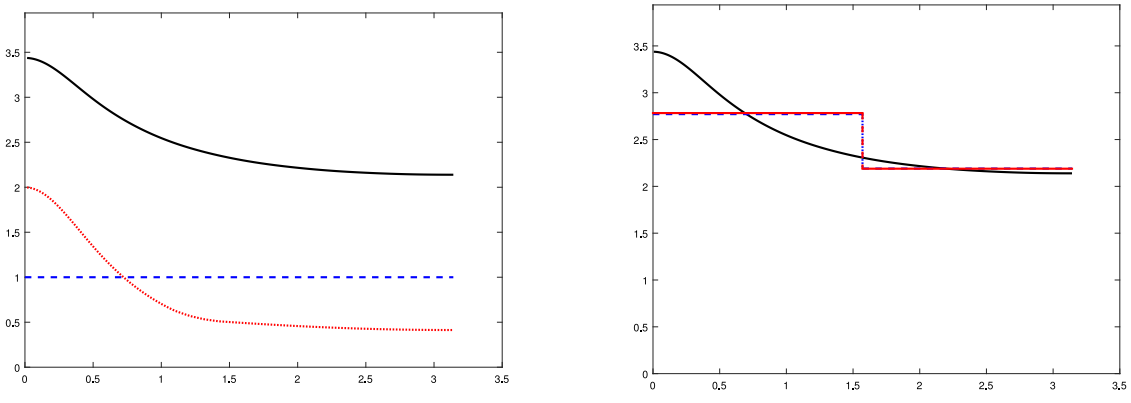

Fig. 2. Estimation results under Model 2. Left: The norms of theoretical spectra of common factors (dashed line), idiosyncratic factors (dotted line), and overall spectrum (solid line). Right: Estimation results for 10,000 simulation runs. Norm of the theoretical spectrum (solid line (smooth)), average norm of theoretical spectrum (solid line (piecewise constant)), and average norm of estimated spectrum (dashed line). Model parameters: $p = 5, k = 2, T = 1000$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

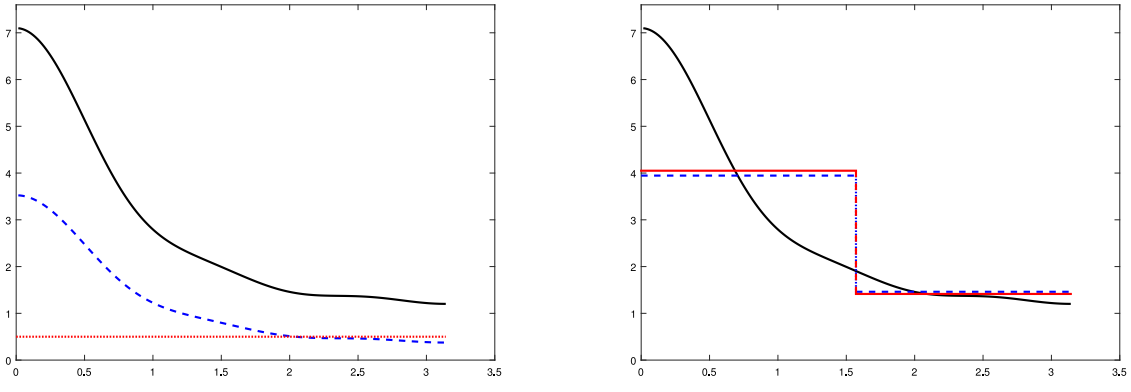

Fig. 3. Estimation results under Model 3. Left: The norms of theoretical spectra of common factors (dashed line), idiosyncratic factors (dotted line), and overall spectrum (solid line). Right: Estimation results for 10,000 simulation runs. Norm of the theoretical spectrum (solid line (smooth)), average norm of theoretical spectrum (solid line (piecewise constant)), and average norm of estimated spectrum (dashed line). Model parameters: $p = 5, k = 2, T = 1000$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

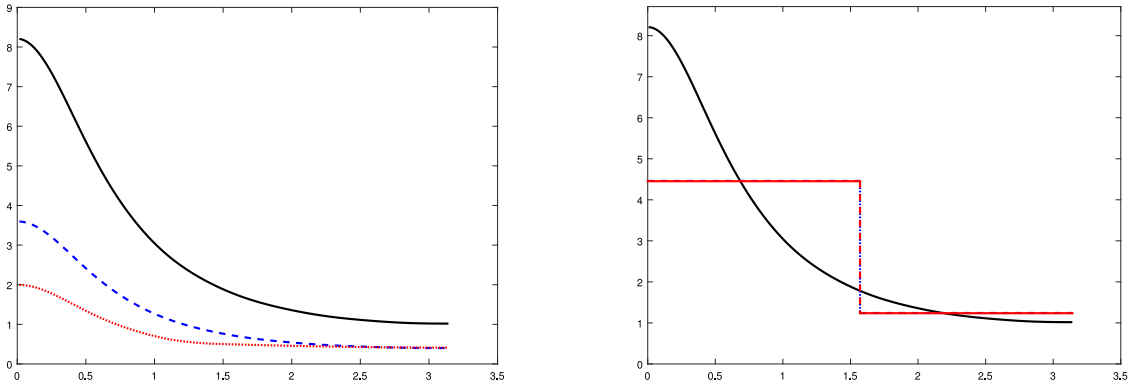

Fig. 4. Estimation results under Model 4. Left: The norms of theoretical spectra of common factors (dashed line), idiosyncratic factors (dotted line), and overall spectrum (solid line). Right: Estimation results for 10,000 simulation runs. Norm of the theoretical spectrum (solid line (smooth)), average norm of theoretical spectrum (solid line (piecewise constant)) and average norm of estimated spectrum (dashed line). Model parameters: $p = 5, k = 2, T = 1000$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Empirical FWERs for Problems 1 and 2 based on unadjusted and adjusted p-values under Model 1 with no autocorrelation in common and idiosyncratic factors based on 10,000 Monte-Carlo repetitions and 100 bootstrap repetitions for varying p, k and T.

	Unadjusted p-values		(Hommel-) Adjusted p-values		Bootstrap	
	Problem 1 FWER	Problem 2 FWER	Problem 1 FWER	Problem 2 FWER	Problem 1 FWER	Problem 2 FWER
$p = 5, k = 2$						
$T = 1000$						
$\alpha = 0.05$	0.2533	0.2333	0.1265	0.0744	0.1279	0.0811
$\alpha = 0.10$	0.3538	0.3932	0.1702	0.1234	0.1719	0.1281
$\alpha = 0.15$	0.4339	0.5262	0.2040	0.1629	0.2050	0.1725
$T = 2000$						
$\alpha = 0.05$	0.2320	0.2264	0.1102	0.0695	0.1100	0.0725
$\alpha = 0.10$	0.3316	0.3867	0.1501	0.1143	0.1527	0.1203
$\alpha = 0.15$	0.4203	0.5171	0.1843	0.1575	0.1863	0.1659
$p = 10, k = 3$						
$T = 2000$						
$\alpha = 0.05$	0.1206	0.1444	0.0404	0.0904	0.0410	0.0921
$\alpha = 0.10$	0.1998	0.2530	0.0638	0.1027	0.0673	0.1067
$\alpha = 0.15$	0.2621	0.3469	0.0858	0.1173	0.0919	0.1217

Table 2
Empirical powers for Problem 1 and empirical FWERs for Problem 2 based on unadjusted and adjusted p-values under Model 2 with no autocorrelation in common and autocorrelation in idiosyncratic factors based on 10,000 Monte-Carlo repetitions and 100 bootstrap repetitions for varying p, k and T.

	Unadjusted p-values		(Hommel-) Adjusted p-values		Bootstrap	
	Problem 1 Power	Problem 2 FWER	Problem 1 Power	Problem 2 FWER	Problem 1 Power	Problem 2 FWER
$p = 5, k = 2$						
$T = 1000$						
$\alpha = 0.05$	0.7038	0.2986	0.6285	0.1291	0.5589	0.1317
$\alpha = 0.10$	0.7808	0.4865	0.7224	0.1753	0.6132	0.1788
$\alpha = 0.15$	0.8219	0.6281	0.7823	0.2193	0.6548	0.2262
$T = 2000$						
$\alpha = 0.05$	0.8484	0.3686	0.8389	0.1288	0.7578	0.1330
$\alpha = 0.10$	0.8969	0.5527	0.8910	0.2027	0.7979	0.2046
$\alpha = 0.15$	0.9243	0.6839	0.9198	0.2715	0.8220	0.2765
$p = 10, k = 3$						
$T = 2000$						
$\alpha = 0.05$	0.9979	0.2473	0.9978	0.1255	0.9969	0.1295
$\alpha = 0.10$	0.9984	0.3711	0.9984	0.1596	0.9973	0.1636
$\alpha = 0.15$	0.9985	0.5040	0.9985	0.1891	0.9977	0.1927

representations of the estimation results we can conclude that the estimation methodology provides fairly good estimates of the models.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

Table 3
Empirical FWERs for Problem 1 and empirical powers for Problem 2 based on unadjusted and adjusted p-values under Model 3 with autocorrelation in common and no autocorrelation in idiosyncratic factors based on 10,000 Monte-Carlo repetitions and 100 bootstrap repetitions for varying p, k and T.

	Unadjusted p-values		(Hommel-) Adjusted p-values		Bootstrap	
	Problem 1 FWER	Problem 2 Power	Problem 1 FWER	Problem 2 Power	Problem 1 FWER	Problem 2 Power
$p = 5, k = 2$						
$T = 1000$						
$\alpha = 0.05$	0.3118	0.4517	0.1556	0.4204	0.1822	0.4299
$\alpha = 0.10$	0.4570	0.4898	0.2056	0.4400	0.2497	0.4505
$\alpha = 0.15$	0.5711	0.5194	0.2488	0.4576	0.3090	0.4680
$T = 2000$						
$\alpha = 0.05$	0.3694	0.5110	0.1460	0.4692	0.1821	0.4762
$\alpha = 0.10$	0.5504	0.5544	0.2177	0.4989	0.2805	0.5053
$\alpha = 0.15$	0.6786	0.5878	0.2795	0.5225	0.3616	0.5273
$p = 10, k = 3$						
$T = 2000$						
$\alpha = 0.05$	0.3702	0.8812	0.2433	0.8464	0.2514	0.8156
$\alpha = 0.10$	0.4603	0.9254	0.2879	0.9056	0.2765	0.8954
$\alpha = 0.15$	0.5230	0.9465	0.3267	0.9337	0.3178	0.9228

Table 4
Empirical powers for Problems 1 and 2 based on unadjusted and adjusted p-values under Model 4 with autocorrelation in common and autocorrelation idiosyncratic factors based on 10,000 Monte-Carlo repetitions and 100 bootstrap repetitions for varying p, k and T.

	Unadjusted p-values		(Hommel-) Adjusted p-values		Bootstrap	
	Problem 1 Power	Problem 2 Power	Problem 1 Power	Problem 2 Power	Problem 1 Power	Problem 2 Power
$p = 5, k = 2$						
$T = 1000$						
$\alpha = 0.05$	0.8137	0.4467	0.7562	0.4104	0.6488	0.4194
$\alpha = 0.10$	0.9011	0.4936	0.8708	0.4343	0.7425	0.4434
$\alpha = 0.15$	0.9350	0.5337	0.9199	0.4572	0.8046	0.4637
$T = 2000$						
$\alpha = 0.05$	0.9745	0.5008	0.9734	0.4571	0.9264	0.4623
$\alpha = 0.10$	0.9835	0.5539	0.9832	0.4898	0.9607	0.4911
$\alpha = 0.15$	0.9875	0.5961	0.9873	0.5207	0.9724	0.5154
$p = 10, k = 3$						
$T = 2000$						
$\alpha = 0.05$	0.9841	0.7917	0.9839	0.7359	0.9751	0.7234
$\alpha = 0.10$	0.9890	0.8415	0.9886	0.7915	0.9857	0.7845
$\alpha = 0.15$	0.9918	0.8725	0.9917	0.8269	0.9889	0.8167

The testing results are presented two-fold: First, we compute empirical family-wise error rates and empirical powers based on unadjusted as well as adjusted p-values in the following manner.

$$\widehat{\text{FWER}} = \frac{1}{MC}\sum_{i=1}^{MC}\mathbf{1}\{\exists j \in I_0 : p_{i,j} < \alpha\},$$

$$\widehat{\text{Power}} = \frac{1}{MC}\sum_{i=1}^{MC}\frac{1}{m_1}\sum_{j\in I_1}\mathbf{1}\{p_{i,j} < \alpha\},$$

where MC denotes the number of Monte Carlo simulation runs, α denotes the target FWER level, and $p_{i,j}$ denotes the p-value for the jth (marginal) test problem in the ith simulation run. Second, we illustrate our testing results by histograms of the minimum adjusted p-values.

As outlined before we consider two competing techniques, namely, asymptotic tests based on Lemmas 2.2 and 2.3 and bootstrap-based tests, and we compare them with respect to their type I and type II error behavior. For the asymptotic tests based on Lemmas 2.2 and 2.3, we compute marginal unadjusted p-values based on the univariate chi-square approximation of each Wald test statistic, as well as multiplicity-adjusted p-values according to Hommel (1988); cf Remark 2.1. The empirical family-wise error rates and powers of the considered tests, based on the unadjusted as well as on adjusted p-values, are presented in Tables 1–4. Notice that all null hypotheses are true in Model 1 (Problems 1 and 2). In Model 2, all null hypotheses are false for Problem 1, and all null hypotheses are true for Problem 2. In Model 3, all null hypotheses are true for Problem 1, and all null hypotheses are false for Problem 2. Finally, all null hypotheses are false in Model 4 (Problems 1 and 2). Besides the model type we vary the model dimensions p, k, and T to judge the sensitivity of the suggested procedures with respect to these parameters. As can be seen from Table 1 presenting the empirical FWERs for Problems 1 and 2 under

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

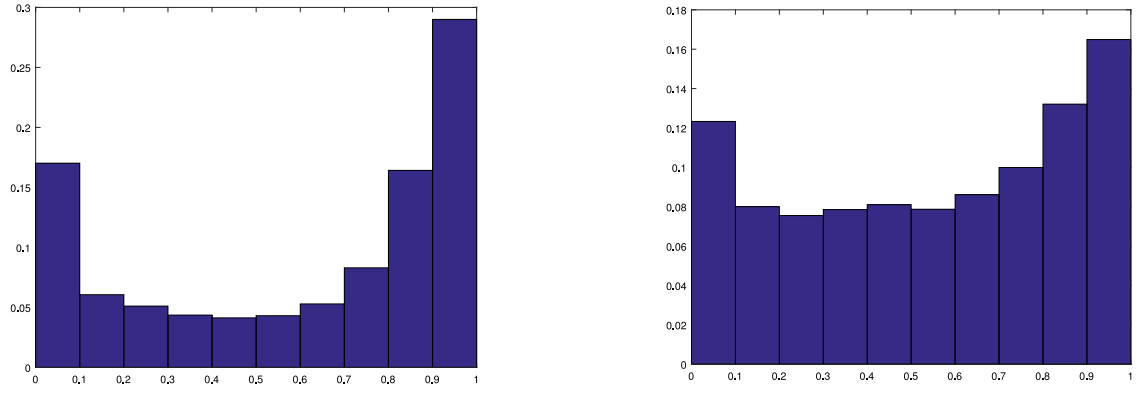

Fig. 5. Histograms of minimal (Hommel-) adjusted p-values under Model 1. Left: Problem 1 on idiosyncratic factors, right: Problem 2 on factor loadings. Model parameters: $p = 5, k = 2, T = 1000$.

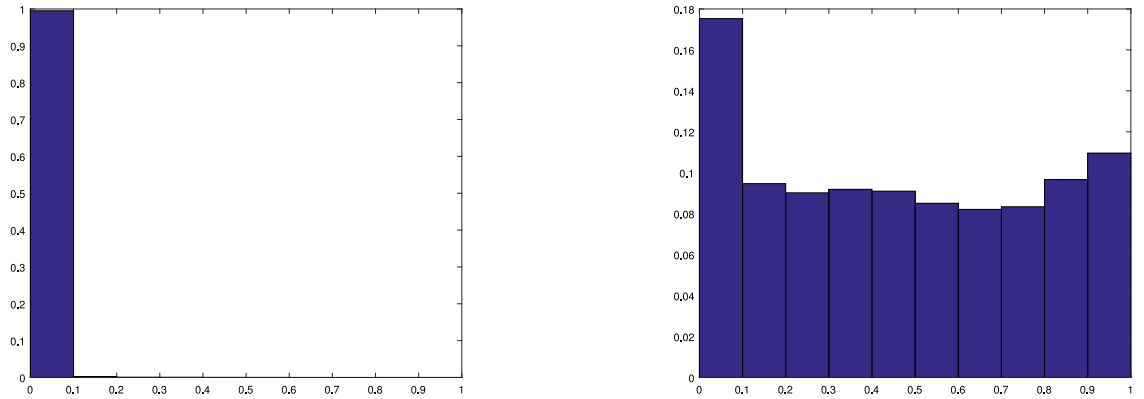

Fig. 6. Histograms of minimal (Hommel-) adjusted p-values under Model 2. Left: Problem 1 on idiosyncratic factors, right: Problem 2 on factor loadings. Model parameters: $p = 5, k = 2, T = 1000$.

Model 1, the empirical FWER decreases when the observational horizon increases (compare $T = 1000$ versus $T = 2000$), and it generally increases when the dimensionality of the model increases, see the results for $p = 5, k = 2$ versus $p = 10, k = 3$. In Models 2 to 4, where common and/or idiosyncratic factors are autocorrelated, we observe an increase in empirical power when increasing the time horizon T, as expected. The behavior of the empirical FWER with increasing T is less pronounced under these models. Finally, the results for Models 1 to 4 demonstrate that the multiplicity adjustment is necessary, because the empirical FWERs based on the unadjusted p-values substantially exceed the significance level α.

In Figs. 5–8 we present the histograms of minimum (Hommel-) adjusted p-values for Problems 1 and 2. Under the respective global hypothesis, these p-values should asymptotically ($T \to \infty$) follow a uniform distribution on $[0, 1]$, while their distribution should be concentrated around zero in the presence of false hypotheses. This theoretical statement is illustrated in the results of the Monte Carlo simulations, as can be seen from the shape of the histograms in the figures. For example, in Model 2 (corresponding to Fig. 6) the idiosyncratic factors are assumed to be autocorrelated, whereas the common factors follow a white noise process. This means that the global null hypothesis for Problem 1 is false under Model 2, whereas the global null hypothesis for Problem 2 is true under Model 2. The left panel of Fig. 6 presents the histogram of the minimal adjusted p-values for Problem 1, with its mass being concentrated around zero, while the right panel of Fig. 6 presents the histogram of the minimal adjusted p-values for Problem 2, with the minimal p-values being approximately uniformly distributed. The other three figures (corresponding to the other three models) can be interpreted analogously.

In summary, our Monte Carlo simulations demonstrate that both the asymptotic multiple chi-square test with an appropriate multiplicity adjustment as well as the bootstrap-based multiple testing procedures keep the family-wise error rate approximately at the predefined significance level α, especially if the DFM has a simple structure as under our Model 1, and if α is not too small such that non-extreme tails of the null distribution have to be approximated. Also, their estimated powers are rather similar. Bootstrap-based testing requires a separate implementation. However, it is in our context based on simulating from a multivariate normal distribution, which can be performed rather efficiently. This makes its use comparable to the (Hommel-) adjusted multiple testing in terms of computational complexity. Thus, we conclude that the two multiple tests can be used interchangeably.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

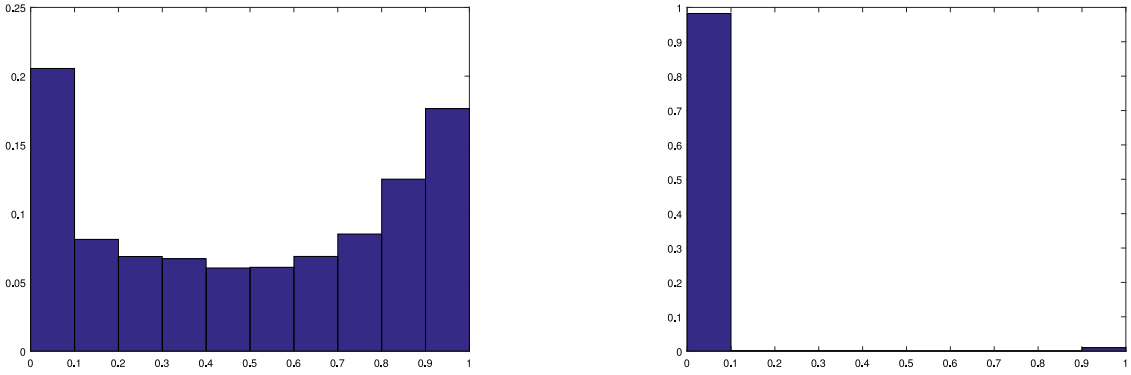

Fig. 7. Histograms of minimal (Hommel-) adjusted p-values under Model 3. Left: Problem 1 on idiosyncratic factors, right: Problem 2 on factor loadings. Model parameters: $p = 5, k = 2, T = 1000$.

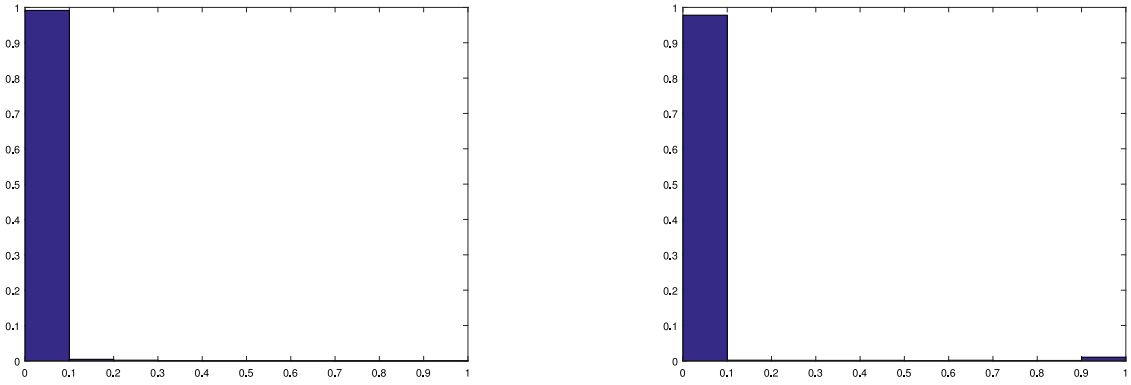

Fig. 8. Histograms of minimal (Hommel-) adjusted p-values under Model 4. Left: Problem 1 on idiosyncratic factors, right: Problem 2 on factor loadings. Model parameters: $p = 5, k = 2, T = 1000$.

We also performed a sensitivity analysis with respect to non-Gaussian distributions and simulated observations from Student's t-distribution with various degrees of freedom. With minor deviations from normality, i.e., in the case of t-distributions with degrees of freedom larger than 10, the testing results did not deviate strongly from the simulation results with observations generated from the normal distribution. However, larger deviations from Gaussianity such as t-distributions with 5 or 3 degrees of freedom lead to a FWER notably larger than the target level. It is possible to reduce it by making the sample sizes larger, but not substantially. Therefore, we do not recommend to use the proposed method if the data are clearly heavy-tailed. In the case that the model-based bootstrap is employed, one may incorporate the information of heavy-tailedness into the resampling scheme, by changing the probability model from which the re-samples are drawn. Similar work in this direction has been done by Creal et al. (2013), Harvey (2013) and Cortes et al. (2017). Extensions of the bootstrap methodology to the t-distribution is straightforward: estimate the degrees of the hypothesized t-distribution from the data and bootstrap from the obtained distribution. A more thorough theoretical analysis of the inference methods employing non-Gaussian distributions is, however, beyond the scope of the present work, and left for future research.

5. Application

In this section we present an application of the proposed methodology to European agricultural data. Our assumption is that such type of data can appropriately be modeled by a DFM. In particular, the considered time series may possibly be driven by several factors such as, e.g., weather conditions as well as the overall economic situation. For our analysis we choose the monthly soft wheat prices which have been obtained from the Eurostat database. The requirement of time series of reasonable length restricts our attention to the five European countries Belgium, France, Italy, the Netherlands and the United Kingdom for the time span January, 1969–April, 1998, resulting in $T = 352$ months. Original Eurostat data contain many missing values which have been linearly interpolated. Fig. 9 displays the five time series of monthly soft wheat prices which appear to be non-stationary by visual inspection. This observation is also confirmed by standard testing on the presence of unit roots in time series such as the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test, and an augmented Dickey–Fuller (ADF) test. One can observe that the time series follow similar dynamics with Belgium and the Netherlands

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

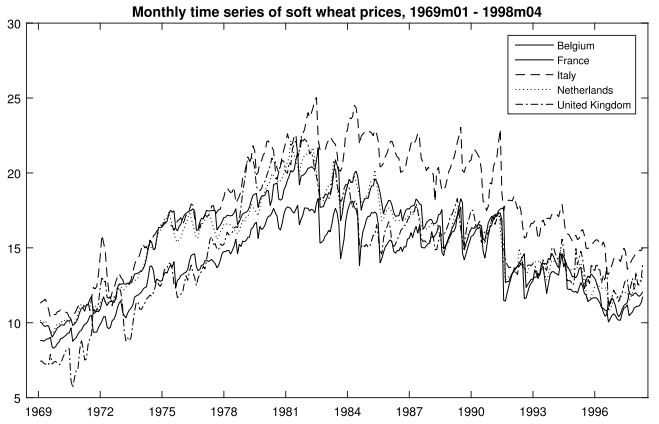

Fig. 9. Monthly time series of soft wheat prices of Belgium, France, Italy, Netherlands, and the United Kingdom from January, 1969 to April, 1998, resulting in $T = 352$ months.

Table 5
P-values of the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test for a unit root in the univariate time series and an augmented Dickey–Fuller (ADF) test for a unit root in a univariate time series for monthly log returns of Belgium, France, Italy, Netherlands, and the United Kingdom from February, 1969 to April, 1998.

Test	Belgium	France	Italy	Netherlands	UK
KPSS	0.100	0.100	0.100	0.100	0.100
ADF	0.010	0.010	0.010	0.010	0.010

having almost coinciding lines, whereas France and Italy have the same pattern with that of Italy being shifted upwards. The dynamics of the UK time series is close to that of Italy before the 1980s and to that of Belgium and France thereafter which possibly implies a structural break in these time series.

Our model assumptions require the data to be stationary. To meet this requirement at least approximately, the original time series have been transformed into logarithmic returns leading to $T = 351$, see Fig. 10 and Table 5. Whereas all log returns appear to be more volatile after the 1980s, the UK log returns seem to follow a very specific pattern: they are particularly volatile in the beginning as well as towards the end of the series.

The likelihood-ratio test suggests that a single-factor model is sufficient for this particular data set with the value $\chi^2(10) = 2.9428$ of its test statistic being far below the corresponding critical value, see Geweke (1977). This result, however, cannot fully be confirmed by the screeplot based on the principal component analysis of the observational (log transformed) matrix $\mathbf{X}$ as according to the screeplot there are at least three factors explaining approximately 47%, 22%, and 16% of the overall variance, respectively. Additional common factors may be caused by the strong dynamics in some of the idiosyncratic factors, see the discussion of Boivin et al. (2008) and Uhlig (2008). Therefore, we estimated the model with $1 \leq k \leq 3$ common factors subsequently to judge whether additional common factors add substantial new information to the model. We adopted the identification scheme for the model as in Geweke (1977).

We present our estimation results for the DFM with $k = 2$ common factors, as we observe quantitatively significant factor loadings for the first factor as well as weaker but still non-negligible loadings for the second factor. The possible presence of the third factor might be due to the strong idiosyncracy in one particular of the time series, possibly the time series of the log returns of the soft wheat of the UK, and the extension of the model to $k = 3$ does not provide remarkably different results. As far as the interpretation of the factors is concerned we hypothesize that the first factor is the overall market trend and the second one is the rainfall which is likely to be random, see Matalas (1963).

The estimated spectrum is illustrated in Fig. 11 and the parameter estimates are listed in Table 6. The spectrum in Fig. 11 has the so-called typical shape of economic variables with spikes in the norm of the periodogram corresponding to the lower frequencies, see Granger (1966). This phenomenon is explained by the dominance of long-term trends, longer cycles, in the behavior of the wheat market in the five aforementioned countries. The parameter estimates provide further insights on the structure of the commodity market for the soft wheat in these countries. The model seems to explain the variations in log returns in Belgium and the Netherlands quite well with significant factor loadings and idiosyncratic variations being quite low. To a lesser extent the log returns of France and Italy can be explained by the model as one can observe quite high idiosyncratic variations and lower factor loadings. Finally, the log returns of the soft wheat in the UK seem to follow a distinctive process with very high factor loadings as well as significant idiosyncratic variances unobserved in the countries of continental Europe. Thus one might conclude that although the log returns are responsive to certain common factors such

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

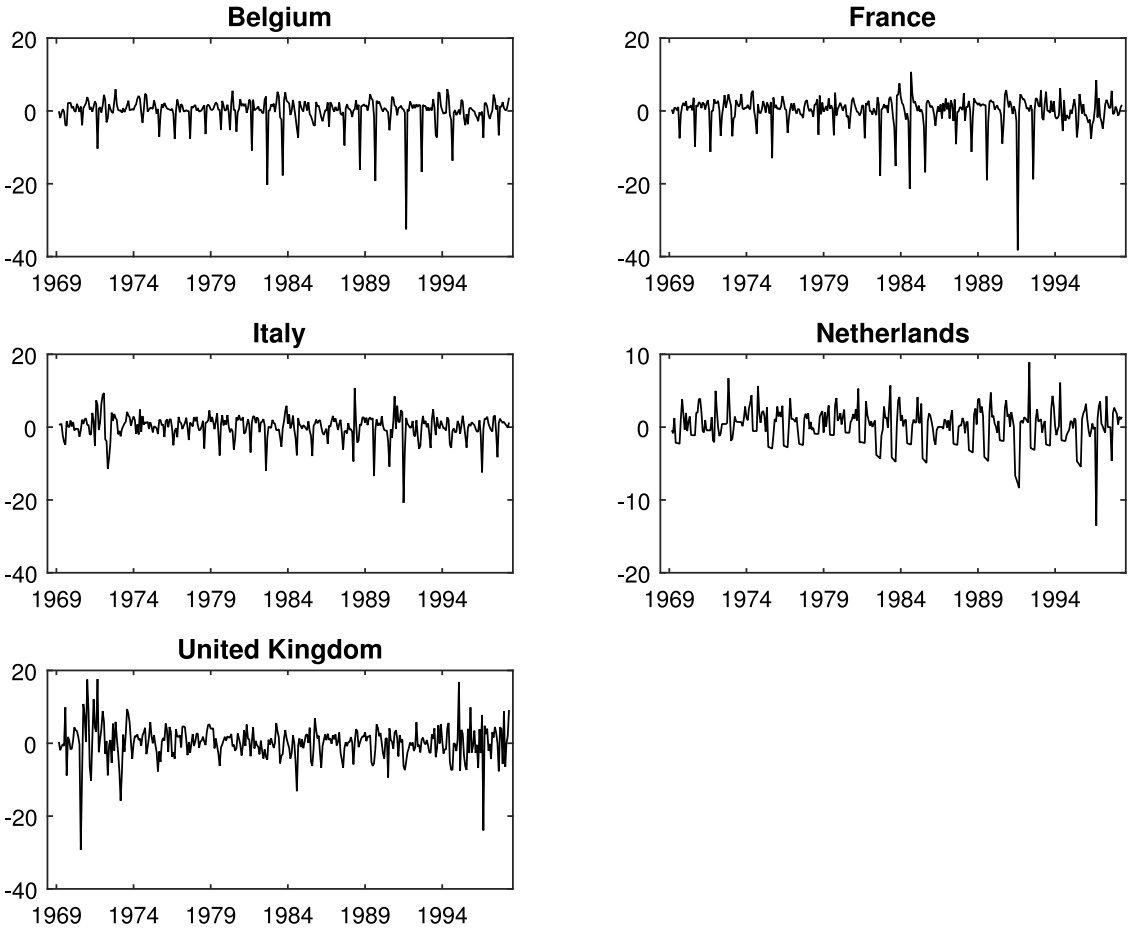

Fig. 10. Monthly time series of log returns of soft wheat prices in percentages of Belgium, France, Italy, Netherlands, and the United Kingdom from February, 1969 to April, 1998, resulting in $T = 351$ months.

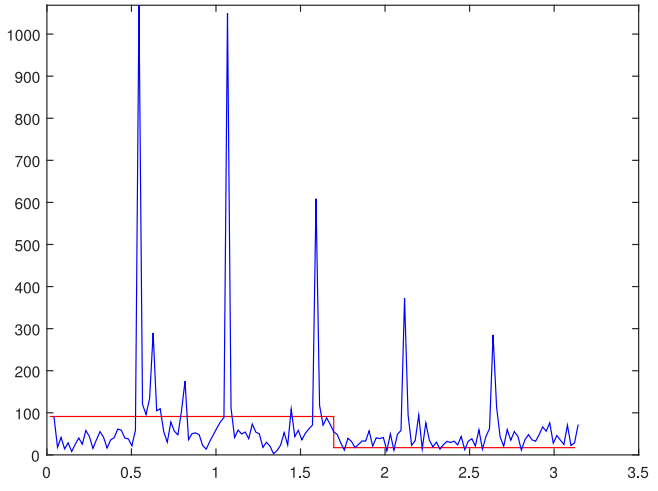

Fig. 11. DFM estimation results with $p = 5$, $k = 2$, $T = 351$ for monthly log returns in percentages of Belgium, France, Italy, Netherlands, and the United Kingdom; February, 1969–April, 1998. The norms of the estimated piecewise constant spectrum as well as the estimated periodograms are plotted against the frequency.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

Table 6
Parameter estimates of the DFM with $p = 5$, $k = 2$, $T = 351$ for monthly log returns in percentages of Belgium, France, Italy, Netherlands, and the United Kingdom; February, 1969–April, 1998. The superscript refers to the bands 1 and 2.

Country	$\tilde{\Lambda}^{(1)}$		$\tilde{\Lambda}^{(2)}$		$s_\varepsilon^{(1)}$	$s_\varepsilon^{(2)}$
Belgium	4.1704	0.0820	2.6546	0.6417	0.0739	4.0783
France	3.1689	1.2306	1.0918	1.1962	9.2086	11.5473
Italy	2.0275	1.9937	0.9417	0.5938	6.2550	4.5313
Netherlands	2.2350	1.5268	1.5821	0.8409	2.3796	0.0000
UK	2.3473	2.6239	0.4150	0.2141	14.5718	13.0928

Table 7
Testing results for autocorrelation in $\tilde{\Lambda}$ and s_ε in the DFM with $p = 5$, $k = 2$, and $T = 351$ for monthly log returns in percentages of Belgium, France, Italy, Netherlands, and the United Kingdom from February, 1969 to April, 1998, relying on $M = 1000$ bootstrap repetitions with $\alpha = 0.05$. The first five rows of the testing results for $\tilde{\Lambda}$ refer to the first common factor and the last five rows refer to the second common factor.

Country	$\tilde{\Lambda}$		s_ε	
	Adjusted p-value	Bootstrap decision	Adjusted p-value	Bootstrap decision
Belgium	< 0.0001	1	< 0.0001	1
France	< 0.0001	1	< 0.0001	1
Italy	< 0.0001	1	< 0.0001	1
Netherlands	0.0034	1	< 0.0001	1
UK	0.0003	1	< 0.0001	1
Belgium	0.2476	0		
France	0.9385	0		
Italy	0.0026	1		
Netherlands	0.0267	0		
UK	0.0178	0		

as, e.g., the overall market trend and the weather conditions, the geographical proximity, e.g., the amount of precipitation in a particular region, plays an equally important role in modeling the corresponding market.

The autocorrelation plots of the log returns of the five considered European countries suggest that the data are autocorrelated, see Fig. 12. It can be seen as well from Fig. 11 that the factors are likely to be autocorrelated as the averaged norm of the periodogram of the observational process exhibits high fluctuations in the first band and notably milder fluctuations in the second band. To judge whether this hypothesis is true as well as to indicate which factors are autocorrelated we refer to Table 7, where we present Hommel-adjusted p-values as well as the test decision by bootstrap for Problems 1 and 2. The testing results at $\alpha = 5\%$ indicate that the first common factor and the idiosyncratic factors are strongly autocorrelated with very small p-values. For the second common factor the testing results are in disagreement for the Netherlands and the UK: whereas the bootstrap fails to reject the null, the asymptotic tests reject the hypothesis of no autocorrelation with quite low p-values.

Based on this analysis we recommend modeling the log returns for the soft wheat in the five aforementioned European countries taking into the account the following characteristics: (i) the geographical proximity plays an important role in the modeling of the respective agricultural log returns, (ii) there is one leading common factor for all the five countries which is autocorrelated and possibly a second one which is nearly random, (iii) the idiosyncratic factors are strongly autocorrelated. These findings can be utilized in constructing forecasts for the respective market. The frequency-domain estimates are typically not used in constructing forecasts in the DFM context. However, this topic is quite well elaborated upon in the time domain; see, e.g., Bai and Ng (2008) and Forni et al. (2005).

6. Discussion

We have comprehensively described a parametric likelihood-based statistical inference approach in small-scale (dynamic) factor models. In particular, details of the implementation of estimation and testing methods have been elucidated in a coherent and unified manner. Furthermore, ready-to-use `MATLAB` programs with which all results of the present manuscript can be reproduced are available from the second author upon request.

As far as the multiple testing results are concerned, we have demonstrated that both the asymptotic chi-square tests and the model-based bootstrap tests approximately keep the FWER level, albeit their type I error behavior remained liberal even for sample sizes of $T = \mathcal{O}(10^3)$. The fact that this behavior was exhibited in a very similar manner by both the asymptotic chi-square tests and the model-based bootstrap tests indicates that the normal approximation of the null distribution of the MLE of the vector of DFM parameters is the most crucial part in the presented methodology. This is in line with the respective comments of Dickhaus and Pauly (2016). Hence, future work may consider nonparametric bootstrap approaches circumventing the assumption of (asymptotic) normality of the MLE in the DFM context. Further possible extensions of this work are to consider other multiple type I error criteria like, for instance, control of the false discovery rate, and to work out multiple testing methodology for larger-scale DFMs, where both of these extensions are interrelated with each other.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

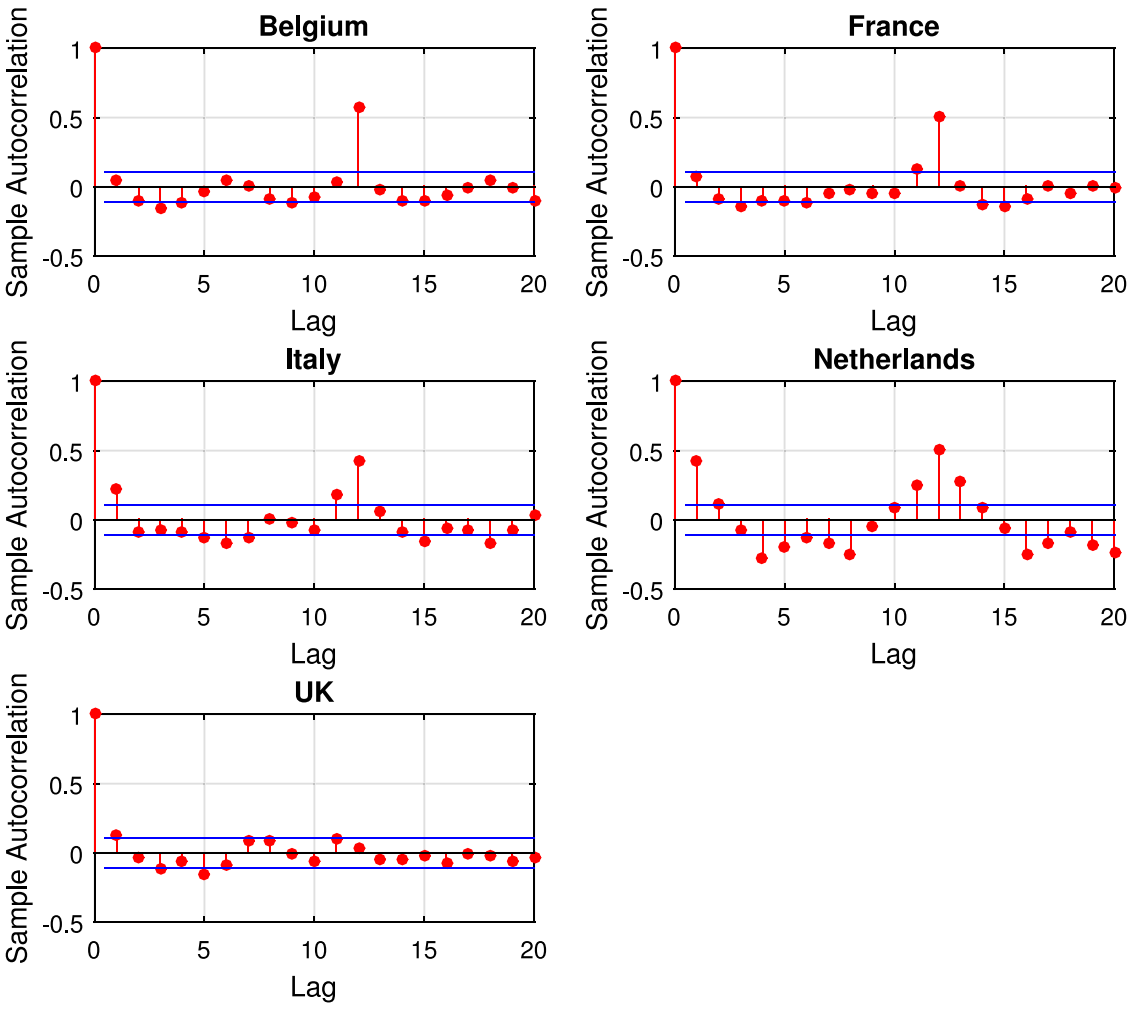

Fig. 12. Autocorrelation plots of monthly time series of log returns of soft wheat prices in percentages of Belgium, France, Italy, Netherlands, and the United Kingdom from February, 1969 to April, 1998, resulting in $T = 351$ months.

Acknowledgments

We are grateful to Markus Pauly for inspiring discussions. We thank two anonymous reviewers for constructive comments which helped to improve the presentation.

References

Anderson, T.W., Rubin, H., 1956. Statistical inference in factor analysis. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 5. pp. 111–150.

Bai, J., Ng, S., 2008. Forecasting economic time series using targeted predictors. J. Econometrics 146 (2), 304–317.

Boivin, J., Giannoni, M.P., Mojon, B., 2008. How has the Euro changed the monetary transmission? Tech. rep.. National Bureau of Economic Research.

Cortes, R., Martins, T., Prates, M., Silva, B., et al., 2017. Inference on dynamic models for non-Gaussian random fields using inla. Braz. J. Probab. Stat. 31 (1), 1–23.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. J. Appl. Econometrics 28 (5), 777–795.

Dickhaus, T., 2012. Simultaneous Statistical Inference in Dynamic Factor Models.SFB 649 Discussion Paper 2012-033. Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany, Available athttp://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2012-033.pdf.

Dickhaus, T., 2014. Simultaneous Statistical Inference with Applications in the Life Sciences. Springer-Verlag Berlin Heidelberg.

Dickhaus, T., Pauly, M., 2016. Simultaneous statistical inference in dynamic factor models. In: Rojas, I., Pomares, H. (Eds.), Time Series Analysis and Forecasting. Springer, pp. 27–45.

Dickhaus, T., Royen, T., 2015. A survey on multivariate chi-square distributions and their applications in testing multiple hypotheses. Statistics 49 (2), 427–454.

Fernández, F.J., 1990. Estimation and testing of a multivariate exponential smoothing model. J. Time Series Anal. 11 (2), 89–105.

Fiorentini, G., Galesi, A., Sentana, E., A spectral EM algorithm for dynamic factor models Banco de Espana Working Paper No. 1619, 2016.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.

Fiorentini, G., Sentana, E., et al., Dynamic specification tests for dynamic factor models CEMFI Working Paper No. 1306, 2013.
Fletcher, R., Powell, M.J., 1963. A rapidly convergent descent method for minimization. Comput. J. 6 (2), 163–168.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. Rev. Econ. Stat. 82 (4), 540–554.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: one-sided estimation and forecasting. J. Amer. Statist. Assoc. 100 (471), 830–840.
Geweke, J.F., 1977. The Dynamic Factor Analysis of Economic Time-Series Models. In: Contributions to Economic Analysis, vol. 103. North-Holland Publishing Company, pp. 365–383 , latent variable in socio-economic models 19, pp.
Geweke, J.F., Singleton, K.J., 1981. Maximum likelihood "confirmatory" factor analysis of economic time series.. Internat. Econom. Rev. 22, 37–54.
Goodman, N., 1963. Statistical analysis based on a certain multivariate complex Gaussian distribution. (An introduction.).. Ann. Math. Stat. 34, 152–177.
Granger, C.W., 1966. The typical spectral shape of an economic variable. Econometrica 150–161.
Hannan, E.J., 1973. Central limit theorems for time series regression.. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 26, 157–170.
Harvey, A.C., 2013. Dynamic models for volatility and heavy tails: with applications to financial and economic time series, Vol. 52. Cambridge University Press.
Heaton, C., Solo, V., 2004. Identification of causal factor models of stationary time series. Econom. J. 7 (2), 618–627.
Hommel, G., 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test.. Biometrika 75 (2), 383–386.
Howe, W.G., 1955. Some contributions to factor analysis. Tech. rep.. Oak Ridge National Lab., Tenn..
Jöreskog, K.G., 1967. Some contributions to maximum likelihood factor analysis. Psychometrika 32 (4), 443–482.
Jöreskog, K.G., 1969. A general approach to confirmatory maximum likelihood factor analysis.. Psychometrika 34 (2), 183–202.
Jöreskog, K.G., Lawley, D.N., 1968. New methods in maximum likelihood factor analysis. Br. J. Math. Stat. Psychol..
Konietschke, F., Bathke, A.C., Harrar, S.W., Pauly, M., 2015. Parametric and nonparametric bootstrap methods for general MANOVA. J. Multivariate Anal. 140, 291–301.
Lavielle, M., Ludeña, C., 2000. The multiple change-points problem for the spectral distribution. Bernoulli 845–869.
Lawley, D., 1958. Estimation in factor analysis under various initial assumptions. Br. J. Math. Stat. Psychol. 11 (1), 1–12.
Lawley, D., Maxwell, A., 1971. Factor analysis as a statistical method. American Elsevier Pub. Co., URL https://books.google.de/books?id=DZ0QAQAAIAAJ.
Lawley, D.N., 1940. The Estimation of Factor Loadings by the Method of Maximum Likelihood. Proc. Roy. Soc. Edinburgh 60, 64–82.
Lawley, D.N., 1941. Further Investigations in Factor Estimation. Proc. Roy. Soc. Edinburgh Sect. A Math. Phys. Sci. 61, 176–185.
Lawley, D.N., 1942. The application of the maximum likelihood method to factor analysis.. Br. J. Psychol. 33, 172–175.
Lehmann, E.L., Romano, J.P., 2005. Testing statistical hypotheses, third ed.. Springer Texts in Statistics. New York, NY: Springer..
Lockhart, R.S., 1967. Asymptotic sampling variances for factor analytic models identified by specified zero parameters. Psychometrika 32 (3), 265–277.
Maravall, A., 1999. Unobserved components in economic time series. In: Pesaran, H., Schmidt, P. (Eds.), Handbook of Applied Econometrics Volume 1: Macroeconomics. Blackwell Publishing Ltd, Oxford, UK.
Matalas, N.C., 1963. Autocorrelation of rainfall and streamflow minimums. US Government Printing Office.
Pauly, M., Brunner, E., Konietschke, F., 2015. Asymptotic Permutation Tests in General Factorial Designs.. J. R. Stat. Soc., Ser. B. 77 (2), 461–473.
Reiersøl, O., 1950. On the identifiability of parameters in Thurstone's multiple factor analysis. Psychometrika 15 (2), 121–149.
Reschenhofer, E., 2008. Freqeuncy Domain Modeling with Piecewise Constant Spectra. J. Mod. Appl. Statist. Methods 7 (2), 467–470.
Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping.. Econometrica 73 (4), 1237–1282.
Sargent, T.J., Sims, C., 1977. Business cycle modeling without pretending to have too much a priori economic theory. New Methods Bus. Cycle Res. 1, 145–168.
Stock, J.H., Watson, M., 2011. Dynamic factor models. Oxf. Handb. Econom. Forecast..
Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97 (460), 1167–1179.
Thurstone, L.L., 1947. Multiple-factor analysis. University of Chicago Press.
Troendle, J.F., 1995. A stepwise resampling method of multiple hypothesis testing.. J. Amer. Statist. Assoc. 90 (429), 370–378.
Uhlig, H., 2008. Macroeconomic Dynamics in the Euro Area. Tech. rep.. National Bureau of Economic Research.
Watson, M.W., Engle, R.F., 1983. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. J. Econometrics 23 (3), 385–400.
Westfall, P.H., Young, S.S., 1993. Resampling-based multiple testing: examples and methods for p-value adjustment.. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley, New York.

Please cite this article in press as: Dickhaus T., Sirotko-Sibirskaya N., Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.012.