COMSTA: 6673 Model 3G pp. 1–17 (col. fig: NIL)

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

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.