Journal of Econometrics 186 (2015) 388–406

Contents lists available at ScienceDirect

Journal of Econometrics

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

Asymptotic analysis of the squared estimation error in misspecified factor models

Alexei Onatski *

Faculty of Economics, University of Cambridge, Sidgwick Ave, Cambridge, CB3 9DD, United Kingdom

ARTICLE INFO

Article history:
Available online 19 March 2015

JEL classification:
C38
C52

Keywords:
Misspecification
Factor model
Number of factors
Loss efficiency

ABSTRACT

In this paper, we obtain asymptotic approximations to the squared error of the least squares estimator of the common component in large approximate factor models with possibly misspecified number of factors. The approximations are derived under both strong and weak factors asymptotics assuming that the cross-sectional and temporal dimensions of the data are comparable. We develop consistent estimators of these approximations and propose to use them for model comparison and for selection of the number of factors. We show that the estimators of the number of factors that minimize these loss estimators are asymptotically loss efficient in the sense of Shibata (1980), Li (1987), and Shao (1997).

1. Introduction

Empirical analyses of high-dimensional economic data often rely on approximate factor models estimated by the principal components method (see Stock and Watson (2011) for a recent survey of related literature). Many of these analyses intend to accurately estimate a low-dimensional common component of the data. For example, the interest may lie in the part of multinational data that can be attributed to a common business cycle, as in Forni and Reichlin (2001), or in the decomposition of sectoral output growth rates into the common and idiosyncratic parts, as in Foerster et al. (2011). Unfortunately, the estimation problem is complicated by the fact that the number of factors is typically unknown and is likely to be misspecified. This paper studies consequences of the misspecification for the squared error of the estimated common component.

Assuming that the cross-sectional and temporal dimensions of the data, n and T, are comparable, we derive asymptotic approximations to the squared error loss through the order $n^{-1} \sim T^{-1}$. We consider both strong and weak factors asymptotics. Under the latter, the asymptotic loss turns out to be minimized not necessarily at the true number of factors.

We develop estimators of the loss which are consistent under strong and under weak factors asymptotics, and propose to use them for model comparison and for selection of the number of factors. We show that estimators of the number of factors that minimize the proposed loss estimates are asymptotically loss efficient in the sense of Shibata (1980), Li (1987), and Shao (1997). The majority of recently proposed estimators of the number of factors, including the popular Bai and Ng (2002) estimators, are asymptotically loss efficient under the strong factors asymptotics, but not under the weak factors one.

The basic framework of our analysis is standard. We consider an approximate factor model

$$X = \Lambda F' + e, \tag{1}$$

where X is an $n \times T$ matrix of data, Λ is an $n \times r$ matrix of factor loadings, F is a $T \times r$ matrix of factors and e is an $n \times T$ matrix of idiosyncratic terms. Throughout the paper, we will treat Λ and F as unknown parameters. Equivalently, our results can be thought of as conditional on the unobserved realizations of random Λ and F.

Suppose that we estimate the first p of the factors and the corresponding loadings by the least squares, and let us denote the estimates as $\hat{F}_{1:p}$ and $\hat{\Lambda}_{1:p}$, respectively. As is well known, $\hat{F}_{1:p}$ and $\hat{\Lambda}_{1:p}$ can equivalently be obtained by the principal components (PC) method. That is, the columns of $\hat{F}_{1:p}/\sqrt{T}$ are unit-length eigenvectors of $X'X$, and $\hat{\Lambda}_{1:p} = X\hat{F}_{1:p}/T$. In the special case where the idiosyncratic terms are i.i.d. $N(0, 1)$, these are the maximum likelihood estimates subject to the normalization. Since we do not know the true value of r, p may be smaller, equal, or larger than r. We will say that the number of factors is misspecified if $p \neq r$.

We are interested in the effect of the misspecification on the quality of the PC estimate $\hat{\Lambda}_{1:p}\hat{F}'_{1:p}$ of the common component $\Lambda F'$

* Tel.: +44 1223 335240; fax: +44 1223 335475.
E-mail address: ao319@cam.ac.uk.

http://dx.doi.org/10.1016/j.jeconom.2015.02.016
0304-4076/

of the data. This quality is measured by the average (over time and cross-section) squared error

$$L_p = \mathrm{tr}\left[(\hat{\Lambda}_{1:p}\hat{F}_{1:p}' - \Lambda F')(\hat{\Lambda}_{1:p}\hat{F}_{1:p}' - \Lambda F')'\right]/(nT). \quad (2)$$

Our interest in L_p is motivated by several reasons. First, accurate extraction of the common component is important in many applications. Second, in the special case where the idiosyncratic terms are i.i.d. $N(0,1)$, L_p is proportional to the Kullback–Leibler distance between the true model (1) and the factor model with factors $\hat{F}_{1:p}$ and loadings $\hat{\Lambda}_{1:p}$. Recall that the expected value of such a distance is usually approximated by Akaike's (1973) information criterion (AIC). In Section 3, we show that the AIC approximation does not hold in the large factor model setting, and propose a valid alternative.

Finally, loss functions similar to L_p are widely used in the context of linear regression models. For example, Mallows (1973) "measure of adequacy for prediction" of linear regression model $Y = Z_{1:p}\beta_{1:p} + \varepsilon$ when the true model is $Y = Z\beta + u$ is given by $(\hat{Z}_{1:p}\hat{\beta}_{1:p} - Z\beta)'(\hat{Z}_{1:p}\hat{\beta}_{1:p} - Z\beta)$. The problems of prediction, model selection, and model averaging with this loss function were extensively studied by Phillips (1979), Kunitomo and Yamamoto (1985), Shao (1997), and Hansen (2007), to name just a few studies.

Since $\Lambda F'$ is unobserved, L_p cannot be evaluated directly. In Section 2, we derive asymptotic approximations for L_p that are easy to analyze and estimate. Section 2.1 considers the standard strong factors asymptotic regime (Bai and Ng, 2008).

The strong factors asymptotics has been criticized by Boivin and Ng (2006), Heaton and Solo (2006), DeMol et al. (2008), Onatski (2010, 2012b), Kapetanios and Marcellino (2010), and Chudik et al. (2011) for not providing accurate finite sample approximations in applications where the factors are moderately or weakly influential. Therefore, in Section 2.2 we derive asymptotic approximations for L_p using Onatski's (2012b) weak factors assumptions.

Using the derived asymptotic approximations, Section 3 develops four different estimators of L_p. All these estimators use a preliminary estimator $\hat{r}$ of the true number of factors r. Under the strong factors asymptotics, if $\hat{r} \xrightarrow{p} r$, all the corresponding loss estimators are consistent for L_p after a shift by a constant that does not depend on p.

Under the weak factors asymptotics, in general, no preliminary estimator $\hat{r}$ can consistently estimate r. As explained in Onatski (2012b, p. 250), one can, instead, estimate the number q of theoretically detectable, or "effective", factors. If $\hat{r} \xrightarrow{p} q$, then two of the corresponding proposed loss estimators provide the asymptotic upper and lower bounds on the shifted loss. We show that the minimizers of these estimators bracket the actual loss minimizer with probability approaching one as n and T go to infinity. The other two loss estimators are consistent for the shifted loss when there is either no cross-sectional or no temporal correlation in the idiosyncratic terms. In these special cases, the number of factors that minimizes the corresponding estimator of the loss is consistent for the number of factors that minimizes the actual loss. The latter is not necessarily equal to the true number of factors r or to the "effective" number of factors q.

All the proposed loss estimators are simple functions of the eigenvalues of the sample covariance matrix. Monte Carlo exercises in Section 4 show that their quality is excellent when simulated factors are relatively strong. When the factors become weaker, the quality gradually deteriorates, but remains reasonably good in intermediate cases.

In Section 5, we provide an empirical example of model comparison based on our loss estimators. We compare a two- and a three-factor model of excess stock returns, and find that estimating the third factor leads to a loss deterioration for the monthly data covering the period from 2001 to 2012. That is, a PC estimate of the three-factor model provides a worse description of the undiversifiable risk portion of the excess returns than a PC estimate of the two-factor model. Interestingly, this loss-based ordering is reversed when we use the data from 1989 to 2000, which suggests a decrease in the signal-to-noise ratio in the more recent excess returns data.

Section 6 discusses possible extensions, establishes a connection with the literature on sparse models (see, for example, Belloni et al. (2012)), and concludes. All proofs are given in Appendix A.

2. Asymptotic approximation for the loss

2.1. Strong factors asymptotics

In what follows, $\mu_i(M)$ denotes the ith largest eigenvalue of a Hermitian matrix M. Further, $A_{\cdot j}$ and $A_{j\cdot}$ denote the jth column and jth row of a matrix A, respectively. We make the following assumptions.

A1 There exists a diagonal matrix D_n with elements $d_{1n} \geq d_{2n} \geq \cdots \geq d_{rn} > 0$ along the diagonal, such that $F'F/T = I_r$ and $\Lambda'\Lambda/n = D_n$.

This assumption is a convenient normalization. The only non-trivial constraint it implies is the requirement that rank $F = r$ and rank $\Lambda = r$.

A2 As $n \to \infty$, $\Lambda'\Lambda/n \to D$, where D is a diagonal matrix with decreasing elements $d_1 > d_2 > \cdots > d_r > 0$ along the diagonal.

Assumption A2 is sometimes called the factor pervasiveness assumption. It requires that the cumulative explanatory power of factors, measured by the diagonal elements of $\Lambda'\Lambda$, increases proportionally to n. The assumption is standard, but may be too strong in some applications. In Section 2.2, we consider an alternative assumption that allows $\Lambda'\Lambda$ to remain bounded as $n \to \infty$.

Let $n, T \to_c \infty$ denote the situation where both n and T diverge to infinity so that $n/T \to c \in (0, \infty)$. This asymptotic regime is particularly useful for the analysis of data with comparable cross-sectional and temporal dimensions, such as many financial and macroeconomic datasets. It also does not preclude situations where n/T is small or large as long as n/T does not go to zero or to infinity.

A3 As $n, T \to_c \infty$, (i) there exists $\varepsilon > 0$ such that $\Pr\left(\mathrm{tr}\,[ee']/(nT) > \varepsilon\right) \to 1$; (ii) for any $j, k \leq r$, $\Lambda_{\cdot j}' e F_{\cdot k}/\sqrt{nT} = O_P(1)$; (iii) $\mu_1\left(ee'/T\right) = O_P(1)$.

Part (i) of A3 rules out uninteresting cases where the idiosyncratic terms e_{it} are zero or very close to zero for most of i and t. Part (ii) of A3 is in the spirit of assumptions E (d,e) in Bai and Ng (2008). Validity of the central limit theorem for sequences $\left\{\Lambda_{ij}e_{it}F_{tk};\, i, t \in \mathbb{N}\right\}$ with $j, k \leq r$ is sufficient but not necessary for A3(ii). Part (iii) of A3 further bounds the amount of dependence in the idiosyncratic terms.

Assumption A3(iii) is technically very convenient and has been previously used by Moon and Weidner (2010a). They provide several examples of primitive conditions implying A3(iii). Proposition 6, which we formulate and prove in Appendix A, shows that A3(iii) holds for very wide classes of stationary processes $\{e_{\cdot t}, t \in \mathbb{Z}\}$.

Proposition 1. *Let $P_{i:j}$ be a $T \times T$ matrix of projection on the space spanned by $F_{\cdot i}, \ldots, F_{\cdot j}$, and let $Q_{i:j}$ be an $n \times n$ matrix of projection*