Journal of Econometrics 191 (2016) 1–18

Contents lists available at ScienceDirect

Journal of Econometrics

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

Efficient estimation of approximate factor models via penalized maximum likelihood

Jushan Bai [a,*,1], Yuan Liao [b]

[a] Department of Economics, Columbia University, 420 West 118 Street, New York, NY 10027, USA

[b] Department of Mathematics, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form
19 August 2015
Accepted 10 October 2015
Available online 20 October 2015

Keywords:
High dimensionality
Unknown factors
Principal components
Conditional sparse
Thresholding
Cross-sectional correlation
Penalized maximum likelihood
Adaptive LASSO
SCAD
Heteroskedasticity

ABSTRACT

We study an approximate factor model in the presence of both cross sectional dependence and heteroskedasticity. For efficient estimations it is essential to estimate a large error covariance matrix. We estimate the common factors and factor loadings based on maximizing a Gaussian quasi-likelihood, through penalizing a large covariance sparse matrix. The weighted ℓ_1 penalization is employed. While the principal components (PC) based methods estimate the covariance matrices and individual factors and loadings separately, they require consistent estimation of residual terms. In contrast, the penalized maximum likelihood method (PML) estimates the factor loading parameters and the error covariance matrix jointly. In the numerical studies, we compare PML with the regular PC method, the generalized PC method (Choi 2012) combined with the thresholded covariance matrix estimator (Fan et al. 2013), as well as several related methods, on their estimation and forecast performances. Our numerical studies show that the proposed method performs well in the presence of cross-sectional dependence and heteroskedasticity.

1. Introduction

In many applications of economics, finance, and other scientific fields, researchers often face a large panel dataset in which there are multiple observations for each individual; here individuals can be families, firms, countries, etc. One useful method for summarizing information in a large dataset is the factor model:

$$y_{it} = \alpha_i + \lambda_{0i}' f_t + u_{it}, \quad i \le N, t \le T, \tag{1.1}$$

where α_i is an individual effect, λ_{0i} is an $r \times 1$ vector of factor loadings and f_t is an $r \times 1$ vector of common factors; u_{it} denotes the idiosyncratic component of the model. Note that y_{it} is the only observable random variable in this model. If we write $y_t = (y_{1t}, \ldots, y_{Nt})'$, $\Lambda_0 = (\lambda_{01}, \ldots, \lambda_{0N})'$, $\alpha = (\alpha_1, \ldots, \alpha_N)'$ and $u_t = (u_{1t}, \ldots, u_{Nt})'$, then model (1.1) can be equivalently written as

$$y_t = \alpha + \Lambda_0 f_t + u_t.$$

An efficient estimation of the factor loadings and factors should take into account both cross-sectional dependence and heteroskedasticity. This paper uses the penalized maximum (quasi) likelihood estimation under large N, T. The maximum likelihood estimator depends on estimating a high-dimensional covariance matrix $\Sigma_{u0} = \mathrm{cov}(u_t)$, which is a difficult problem when it is non-diagonal and $N/T \to \infty$. Recently, Bai and Li (2012a) studied the maximum likelihood estimation when Σ_{u0} is a diagonal matrix. As was shown by Chamberlain and Rothschild (1983), it is desirable to allow dependence among the error terms $\{u_{it}\}_{i \le N, t \le T}$ not only serially but also cross-sectionally. This gives rise to the *approximate factor model*. With approximate factor models, Doz et al. (2012) considered the consistency of MLE for f_t, restricting a diagonal error covariance matrix. Bai and Li (2012b) estimated an approximate factor model for both factors and factor loadings with MLE, also restricting a diagonal error covariance matrix, and derived the limiting distributions of the estimators. These are shrinkage estimators that shrink the off diagonal elements of Σ_{u0} to zero.

In addition to the diagonal elements, this paper also estimates the off-diagonal elements of Σ_{u0}, which has $O(N^2)$ number of parameters. The key assumption we make is that the model is *conditionally sparse*, in the sense that Σ_{u0} is a sparse matrix with

* Corresponding author.
E-mail addresses: jb3064@columbia.edu (J. Bai), yuanliao@umd.edu (Y. Liao).
[1] Also at: School of Finance, Nankai University, China.

http://dx.doi.org/10.1016/j.jeconom.2015.10.003
0304-4076/

bounded eigenvalues. This assumption requires many off-diagonal elements of Σ_{u0} to be zero or nearly so, but still allows the identities of the sparse positions to be unknown. The conditional sparsity, though slightly stronger than the assumptions in Chamberlain and Rothschild (1983), is meaningful in practice. For example, when the idiosyncratic components represent firms' individual shocks, they are either uncorrelated or weakly correlated among the firms across different industries, because the industry specific components are not necessarily pervasive for the whole economy (Connor and Korajczyk, 1993). Under the sparsity assumption, Fan et al. (2013) proposed a thresholding method to consistently estimate Σ_{u0} when $N > T$. Their method is based on the traditional principal components method, and does not improve the estimation of factors and loadings. This paper proposes a maximum likelihood (ML)-based method that simultaneously estimate the error covariance matrix and loadings, taking into account both cross-sectional correlations and heteroskedasticity.

Let $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, and $S_y = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y})'$ be the sample covariance matrix based on the observed data. The quasi-likelihood function is

$$L(\Lambda, \Sigma_u, S_f) = \frac{1}{N}\log|\Lambda S_f \Lambda' + \Sigma_u| + \frac{1}{N}\mathrm{tr}(S_y(\Lambda S_f \Lambda' + \Sigma_u)^{-1}), \quad (1.2)$$

where $S_f = \frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})'$, with $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$. In addition, a weighted ℓ_1-penalty is attached to penalize the estimation of off-diagonal entries. So we are solving the following optimization problem:

$$\min_{\Lambda, \Sigma_u, S_f}\left[L(\Lambda, \Sigma_u, S_f) + \sum_{i \neq j}\mu_{N,T} w_{ij}|\Sigma_{u,ij}|\right]$$

where the weight w_{ij} is the entry-dependent weight; $\mu_{N,T}$ is a tuning parameter. We provide data-dependent choices for $\{w_{ij}\}_{i,j \le N}$ and $\mu_{N,T}$, as well as the corresponding theories.

There has been a large literature on estimating model (1.1). Stock and Watson (1998; 2002) and Bai (2003) considered the principal components analysis (PC), which essentially treats u_{it} to have the same variance across i, and is inefficient. Choi (2012) proposed a generalized PC; also see Breitung and Tenhofen (2011). Additional literature on factor models includes, for example, Tsai and Tsay (2010), Bai and Ng (2002), Wang (2009), Dias et al. (2013), Han (2012), among others. Most of these studies are based on the PC method, which is inefficient under cross-sectional heteroskedasticity with unknown dependence structures. Moreover, this paper studies high-dimensional static factor models although the factors and errors can be serially correlated. For generalized dynamic factor models, the readers are referred to Forni et al. (2000; 2005), Forni and Lippi (2001), Hallin and Liška (2007), among others. Our estimation method is maximum likelihood (ML) based, in which no spectral analysis is involved. The ML-based estimation allows for over-identification restrictions to be imposed on the loadings (in a similar way as Bai and Wang (2015)) and allows for forecasting in the spirit of Giannone et al. (2008).

The theoretical results of our paper are only about the consistency of the estimators, although some convergence rate of the covariance estimator is presented in Lemma B.2 in the Appendix, which is not minimax optimal. We admit that due to the technical difficulty, it is challenging to derive the optimal (or near optimal) rate of convergence, and further research on the optimal rate is needed in the future. This paper aims to propose a novel ML-based method for estimating approximate factor models, and illustrates its appealing features to use in practice. We shall elaborate the advantages of ML-based methods in Section 2.2. In addition, we assume the number of factors r to be known. Both N and T diverge to infinity and r is fixed. In practice, r can be estimated from the data, and there has been a large literature addressing its consistent estimation, for example, Bai and Ng (2002), Kapetanios (2010), Onatski (2010), Alessi et al. (2010), Hallin and Liška (2007), and Lam and Yao (2012), among others.

The recent work by Fan et al. (2013) focuses on the covariance estimation using the regular PC. In contrast, we focus on efficiently estimating the factors, loadings, and the covariance matrices simultaneously using penalized MLE. Hence we focus on different estimation problems. The maximum likelihood method has been one of the fundamental tools for statistical estimation and inference.

Our approach is also closely related to the large covariance estimation literature, which has been rapidly growing in recent years. Our penalization procedure is similar to the method in Lam and Fan (2009), Bien and Tibshirani (2011), etc. However, as we described above, our approach is still quite different from theirs in the sense that the penalized ML method considered in this paper estimates the loadings and error covariance matrix jointly. A major difficulty is that the likelihood function being considered contains a few fast-diverging eigenvalues thanks to $\Lambda_0\Lambda_0'$. One of our main objectives is to show that maximizing the Gaussian likelihood function involving fast-diverging eigenvalues can still achieve consistency. Other works on large covariance estimation include, for example, Cai and Zhou (2012), Bickel and Levina (2008), Fan et al. (2008), Jung and Marron (2009), Witten et al. (2009), Deng and Tsui (2013), Yuan (2010), Ledoit and Wolf (2012), El Karoui (2008), Pati et al. (2012), Rohde and Tsybakov (2011), Zhou et al. (2011) and Ravikumar et al. (2011), etc.

The paper is organized as follows. Section 2 defines the simultaneous estimation using penalized MLE, and discusses the advantages of ML-based methods. Section 3 presents theoretical analysis. Section 4 discusses computational issues and implementations. Section 5 numerically compares the proposed methods with competing ones in the literature on both estimation and time series forecasts, using simulated and real data. Finally, Section 6 concludes with further discussions. All proofs are given in the Appendix.

Notation

Let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalues of a matrix A respectively. Also let $\|A\|_1$, $\|A\|$ and $\|A\|_F$ denote the ℓ_1, spectral and Frobenius norms of A, respectively. They are defined as $\|A\|_1 = \max_i \sum_j |A_{ij}|$, $\|A\| = \sqrt{\lambda_{\max}(A'A)}$, $\|A\|_F = \sqrt{\mathrm{tr}(A'A)}$. Note that when A is a vector, both $\|A\|$ and $\|A\|_F$ are equal to the Euclidean norm. For two sequences a_T and b_T, we write $a_T \ll b_T$, and equivalently $b_T \gg a_T$, if $a_T = o(b_T)$ as $T \to \infty$. Also, $a_T \asymp b_T$ if $a_T = o(b_T)$ and $b_T = o(a_T)$.

2. Simultaneous estimation based on maximum likelihood

The approximate factor model (1.1) implies the following covariance decomposition:

$$\Sigma_{y0} = \Lambda_0 \,\mathrm{cov}(f_t)\, \Lambda_0' + \Sigma_{u0}, \quad (2.1)$$

assuming f_t to be uncorrelated with u_t, where Σ_{y0} and Σ_{u0} denote the $N \times N$ covariance matrices of y_t and u_t; $\mathrm{cov}(f_t)$ denotes the $r \times r$ covariance of f_t, all assumed to be time-invariant. The approximate factor model typically requires the idiosyncratic covariance Σ_{u0} have bounded eigenvalues and $\Lambda_0'\Lambda_0$ have eigenvalues diverging at rate $O(N)$. One of the key concepts of approximate factor models is that it allows Σ_{u0} to be non-diagonal.