Computational Statistics and Data Analysis 76 (2014) 95–115

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Parameter cascading for panel models with unknown number of unobserved factors: An application to the credit spread puzzle

Oualid Bada *, Alois Kneip

Institute for Financial Economics and Statistics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn, Germany

ARTICLE INFO

Article history:
Received 3 September 2012
Received in revised form 8 November 2013
Accepted 9 November 2013
Available online 26 November 2013

Keywords:
Large panel data
Factor error structure
Factor model
Common stochastic trends
Model selection criteria
Credit spread puzzle

ABSTRACT

The iterative least squares method for estimating panel models with unobservable factor structure is extended to cover the case where the number of factors is unknown a priori. The proposed estimation algorithm optimizes a penalized least squares objective function to estimate the factor dimension jointly with the remaining model parameters in an iterative hierarchical order. Monte Carlo experiments show that, in many configurations of the data, such a refinement provides more efficient estimates in terms of MSE than those that could be achieved if the feasible iterative least squares estimator is calculated with an externally selected factor dimension. The method is applied to the problem of the credit spread puzzle to estimate the space of the missing risk factors jointly with the effects of the observed credit and illiquidity risks.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, the use of panel data has attracted increasing attention in many empirical studies. This is motivated by the fact that such data sets allow statisticians to deal with the problem of the *unobserved heterogeneity*. Recent studies have discussed large panel data models in which the unobserved heterogeneity can be modeled by a factor structure; see, e.g., Bai (2009), Bai et al. (2009), Kneip et al. (2012), and Pesaran (2006). While most of the ongoing studies have focused on fitting the model for a given number of factors, the present work considers the problem of estimating the factor dimension jointly with the unknown model parameters. Our estimation algorithm can be applied for models of the form:

$$Y_{it} = X_{it}\beta + \underbrace{F_t \Lambda_i'}_{(1\times d)\times(d\times 1)} + \epsilon_{it} \quad \text{for } i \in \{1, \ldots, N\} \text{ and ; } t \in \{1, \ldots, T\}, \tag{1}$$

where X_{it} is a $(1 \times p)$ vector of observable regressors, β is a $(p \times 1)$ vector of unknown parameters, Λ_i is a $(1 \times d)$ vector of individual scores (or factor loadings), F_t is a $(1 \times d)$ vector of unobservable common time-varying factors, ϵ_{it} is the error term, and d is an unknown integer, which has to be estimated jointly with β, Λ_i and F_t.

The difference between (1) and the classical panel data models consists in the unobserved factor structure $F_t\Lambda_i'$. It is noted that (1) not only provides a generalization of panel data models with additive effects, where $d = 2$, $F_t = (f_{t,1}, 1)$, and $\Lambda_i = (1, \lambda_{2,i})$, but also includes the dynamic factor models in static form as in Stock and Watson (2005). To illustrate this case, consider a static factor model with autocorrelated idiosyncratic errors of order P such that

$$Y_{it} = F_t^* \Lambda_i' + e_{it} \quad \text{and}$$
$$e_{it} = \beta_1 e_{i,t-1} + \cdots + \beta_P e_{i,t-P} + \epsilon_{it}.$$

* Corresponding author. Tel.: +49 0 228 73 9271; fax: +49 0 228 73 5050.
E-mail addresses: bada@uni-bonn.de (O. Bada), akneip@uni-bonn.de (A. Kneip).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.11.007

It is easily verified that integrating the expansion of e_{it} in the first equation and using $e_{i,t-p} = Y_{i,t-p} - F^*_{t-p}\Lambda'_i$ for each $p = 1, \ldots, P$ results in a panel model of form (1), where the regressors are the lags of Y_{it}, i.e., $X_{it} = (Y_{i,t-1}, \ldots, Y_{i,t-P})$, and $F_t = F^*_t - \beta_1 F^*_{t-1} - \cdots - \beta_P F^*_{t-P}$.

Moreover, the static presentation of the unobserved factor structure in (1) can arise from q-dimensional dynamic factors, say $\mathbf{F}_t$, (also called primitive factors or primitive shocks). In this case, $F_t = [\mathbf{F}_t, \ldots, \mathbf{F}_{t-m}]$ and $d = q(m+1) \geq q$.

Stock and Watson (2005) propose to estimate dynamic factor models in static form by the iterated least squares method (also called iterative principal component). Bai (2009) studies (1) in the context of panel data models and provides asymptotic theory for the iterative least squares estimators when both N and T are large. However, Stock and Watson (2005) and Bai (2009) assume the factors to be stationary. Bai et al. (2009) extend the theoretical development of Bai (2009) to the case where the cross-sections share unobserved common stochastic trends of unit root processes. They prove that the asymptotic bias arising from the time series in such a case can be consistently estimated and corrected. Ahn et al. (2013) consider the classical case where T is small and N is large and estimate the model by using the generalized method of moments (GMM). They show that, under fixed T, the GMM estimator is more efficient than the estimator of Bai (2009).

A second criticism of the conventional iterative least squares method is that the number of unobserved factors has to be known a priori. In this regard, Bai and Ng (2002) and Bai (2004) propose new panel information criteria to assess the number of significant factors in large panels. The performance of these criteria depends, however, on the choice of an appropriate maximal number of factors to be considered in the selection procedure. Hallin and Liška (2007) propose similar criteria in the context of generalized dynamic factor models and provide a calibration strategy to adjust the height of the penalization; however, the calibration requires extensive computations. Alternatively, Kapetanios (2010) proposes a threshold approach based on the empirical distribution properties of the largest eigenvalue. The method requires i.i.d. errors. Onatski (2010) extends the approach of Kapetanios (2010) by allowing the errors to be either serially correlated or cross-sectionally dependent. Onatski (2009) proposes a test statistic based on the ratios of adjacent eigenvalues in the case of Gaussian errors. Alternative methods for assessing the number of factors in the context of principal component analysis and classical factor analysis can be found in Josse and Husson (2012), Dray (2008), and Chen et al. (2010). But note that all these approaches assume the factors to be extracted directly from observed variables and not estimated with other model parameters.

Kneip et al. (2012) consider the case of observed regressors and unobserved common factors and propose a semi-parametric estimation method and a sequential testing procedure to assess the dimension of the unobserved factor structure. The asymptotic properties of their approach rely on second order differences of the factors and i.i.d. idiosyncratic errors. Pesaran (2006) attempts to control for the hidden factor structure by introducing additional regressors into the model, which are the cross-section averages of the dependent variables and the cross-section averages of the observed explanatory variables. The advantage of this estimation procedure is its invariance to the unknown factor dimension. However, the method does not aim to consistently estimate the factor structure but only deals with the problem of its presence when estimating the remaining model parameters.

In this paper, we extend the iterative approach of Bai (2009) and Bai et al. (2009) in such a way that we allow for the number of factors to be unknown a priori. We integrate a penalty term into the objective function to be globally optimized and update iteratively the estimators of all required parameters in hierarchical order as described in Cao and Ramsay (2010). Our parameter-cascading strategy also includes the update of the penalty term in order to adjust the height of the penalization and avoid under- and over-parameterization. Monte Carlo experiments show that, in many configurations of the data, such a refinement provides more efficient estimates in terms of MSE than those that could be achieved if the feasible iterative least squares estimator is calculated with an externally selected factor dimension.

There are a lot of examples where the determination of the number of factors in the presence of additional observed regressors is of particular interest. As an example, we consider, in our application section, the problem of the so-called *credit spread puzzle*-the gap between the observed corporate bond yields and duration-equivalent government bond yields that classical financial models of credit risk fail to explain (see, e.g., Huang and Huang, 2012, and Elton et al., 2001). For a long time, the credit spread has been considered a simple compensation for credit default risk. Most empirical studies show, however, that default risk cannot be the unique explanatory variable. Kagraoka (2010) decomposes the credit spread into credit risk, illiquidity risk, and an unobservable risk component, which he defines as systematic risk premium; however, he assumes the unobserved part to be generated by only one factor. Castagnetti and Rossi (2013) adopt a heterogeneous panel model with multiple factors. Their results suggest that credit spreads are driven by observable as well as unobservable common risk factors.

In our application, we extend the empirical development of Kagraoka (2010) by allowing for the systematic risk premium to be composed of multiple hidden factors. Moreover, we allow for some slope coefficients to be temporally heterogeneous. This differs from the setting of Castagnetti and Rossi (2013), who use a panel model with cross-sectionally heterogeneous slope parameters. Our empirical study relies on daily observations of 111 US corporate bonds over a period of 397 business days. Our results suggest the presence of two unobserved primitive risk factors affecting US corporate bonds during the period between September 2006 and March 2008, while one single factor is sufficient to describe the data for the time periods prior to the beginning of the subprime crisis in 2007.

The remainder of this paper is organized as follows: Section 2 proposes an algorithmic refinement of the conventional iterative least squares estimation method. In Section 3, we extend the model with additional nominal variables, discuss the

Download English Version:

https://daneshyari.com/en/article/6869956

Download Persian Version:

https://daneshyari.com/article/6869956

Daneshyari.com