Contents lists available at ScienceDirect

Journal of Econometrics

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

Least squares estimation of large dimensional threshold factor models[☆]

Daniele Massacci

Bank of England, Threadneedle Street, London, EC2R 8AH, United Kingdom

ARTICLE INFO

Article history:
Received 8 September 2015
Received in revised form
27 June 2016
Accepted 5 November 2016
Available online xxxx

JEL classification:
C12
C13
C33
C52
G10

Keywords:
Large threshold factor model
Least squares estimation
Model selection
Linearity testing
Connectedness

ABSTRACT

This paper studies large dimensional factor models with threshold-type regime shifts in the loadings. We estimate the threshold by concentrated least squares, and factors and loadings by principal components. The estimator for the threshold is superconsistent, with convergence rate that depends on the time and cross-sectional dimensions of the panel, and it does not affect the estimator for factors and loadings: this has the same convergence rate as in linear factor models. We propose model selection criteria and a linearity test. Empirical application of the model shows that connectedness in financial variables increases during periods of high economic policy uncertainty.

1. Introduction

Factor models are widely used tools to explain the common variations in large scale macroeconomic and financial data. An extensive literature analyzes factor models under the maintained assumption of constant loadings over the entire sample period: see Connor and Korajczyk (1986, 1988, 1993), Forni et al. (2000, 2004, 2015), Forni and Lippi (2001), Bai and Ng (2002), Stock and Watson (2002), and Bai (2003) for seminal contributions on linear factor models. Economic models are however unlikely to have constant parameters over time and factor models with time-dependent loadings are called for. Time-dependence in the loadings may be easily implemented through a change-point mechanism: this may be parameterized as either a structural break or a regime shift driven by the threshold principle, depending on the underlying data generating process.

Structural breaks in the loadings may arise as a consequence of events such as technological or policy changes. Several important contributions deal with large dimensional factor models subject to loadings instabilities. Breitung and Eickmeier (2011) show that ignoring breaks leads to overestimation of the number of factors and develop statistical tests for the null hypothesis of stability in the loadings. Bates et al. (2013) study the robustness properties of the principal components estimator of the factors under neglected loadings instability. Chen et al. (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) develop further statistical tools to detect breaks. Chen (2015) considers least squares estimation of the break date. Cheng et al. (2015) propose shrinkage estimation of large dimensional factor models with structural breaks.

☆ This work was carried out when the author was Franco Modigliani Research Fellow in Economics and Finance at the Einaudi Institute for Economics and Finance (EIEF); it was revised after the author joined the Bank of England. The views in this paper are the author's and do not necessarily reflect those of the Bank of England, or its policy committees. The author really is highly indebted to Marco Lippi for introducing him to factor models and for several enlightening conversations. Kind hospitality by the Department of Finance at Bocconi University is acknowledged. This paper benefits from comments from seminar participants at University of Rome Tor Vergata, Bocconi University and the Bank of Italy; from suggestions from conference participants at the Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance, and at the IAAE 2015 Annual Conference; and from conversations with Domenico Giannone, Alessandro Giovannelli, Hashem Pesaran and Stefano Soccorsi. The author thanks Jianqing Fan (the co-editor), the associate editor, and two anonymous referees for the insightful comments. Errors and omissions are the author's own responsibility. Financial support from the Associazione Borsisti Marco Fanno and from UniCredit and Universities Foundation is gratefully acknowledged.

E-mail address: dm355@cantab.net.

http://dx.doi.org/10.1016/j.jeconom.2016.11.001

Regime shift representations of the dependent variables are suitable when "history repeats", as with financial returns (Timmermann, 2008; Ang and Timmermann, 2012). Ng and Wright (2013) introduce a threshold mechanism in large dimensional factor models to simulate data and investigate the effects of nonlinearities on business cycle dynamics.[1] We take Ng and Wright (2013) intuition as a starting point and propose a large dimensional factor model with regime changes in the loadings governed by the threshold principle. We let the threshold value be unknown and focus on estimation, model selection and linearity testing. To the very best of our knowledge, we are the first to tackle this problem.

Let R^0 be the true number of factors. Under the maintained assumption that R^0 is known, we propose to estimate the threshold value by concentrated least squares, and factors and loadings by principal components (Hansen, 2000; Bai and Ng, 2002). We obtain a number of novel theoretical results. Let N and T denote the cross-sectional and time series dimensions, respectively. We first provide sufficient conditions to ensure that our model is identified from a linear factor model: formally, for $0.5 < \alpha^0 \leq 1$, we require that *at least* a fraction $O\left(N^{\alpha^0}\right)$ of the N cross-sectional units experiences a regime shift in the loadings, so that the shift resists to the aggregation induced by the principal components estimator. We then show that the estimator for the threshold parameter is consistent at a rate equal to $N^{\alpha^0}T$: this depends on the time series dimension T and the number of cross-sectional units N^{α^0} subject to the threshold effect. The convergence rate monotonically increases in α^0 and it is such that $\sqrt{N}T < N^{\alpha^0}T \leq NT$: this shows the direct relationship between identification of the model and convergence rate of the estimator for the threshold. As a consequence of this superconsistency property, we finally show that the principal components estimator for both regime-specific loadings and factors have convergence rate equal to $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$: despite the threshold effect, the convergence rate C_{NT} is equal to the one derived in Bai and Ng (2002) for linear factor models.

We next let the true number of factors R^0 be unknown so that it has to be estimated. Breitung and Eickmeier (2011) show that structural instability in the loadings leads to a factor representation with a higher dimensional factor space: due to an analogy argument, the same issue arises when a regime shift drives time variation in the loadings. Since the convergence rate C_{NT} of the estimator for loadings and factors is the same as in linear factor models, we make Bai and Ng (2002) information criteria robust to the threshold effect by accounting for the induced higher dimensional factor space representation.

As a last theoretical contribution, we propose a linearity test. Following Chen et al. (2014) and Han and Inoue (2015), we check whether the covariance matrix of the estimated factors is regime-dependent: we use the regression approach of Chen et al. (2014) and extend Hansen (1996) seminal contribution to derive the asymptotic distribution of the test statistic under the null hypothesis of linearity.

We finally show how our theoretical framework may be used to measure connectedness in financial markets (Acharya et al., 2010; Billio et al., 2012; Engle and Kelly, 2012; Diebold and Yilmaz, 2014; Adrian and Brunnermeier, 2016). We extend Billio et al. (2012) measure based on principal components analysis to allow for regime-specific connectedness. Using Baker et al. (2016) index of economic policy uncertainty as threshold variable, we show that connectedness in financial markets increases during periods of high uncertainty: this may be relevant for risk measurement and management.

The remainder of the paper is organized as follows. Section 2 describes the model. Section 3 deals with estimation. Section 4 looks at model selection. Section 5 develops a linearity test. Section 6 performs a Monte Carlo analysis. Section 7 provides an empirical application. Section 8 outlines directions for future research. Finally, Section 9 concludes. Appendix provides technical proofs.

Concerning notation, $\mathbb{I}(\cdot)$ denotes the indicator function; given a square matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$; the norm of a generic matrix $\mathbf{A}$ is $\|\mathbf{A}\| = \left[\mathrm{tr}\left(\mathbf{A}'\mathbf{A}\right)\right]^{1/2}$; for a given scalar A, $|A|$, $\mathbf{I}_A$ and $\mathbf{0}_A$ are the absolute value of A, the $A \times A$ identity matrix and the zero matrix, respectively; $\xrightarrow{p}$ denotes convergence in probability; $\xrightarrow{d}$ denotes convergence in distribution; $\Rightarrow$ denotes weak convergence with respect to the uniform metric.

2. The approximate threshold factor model

We consider the model

$$\mathbf{x}_t = \mathbb{I}(z_t \leq \theta)\,\boldsymbol{\Lambda}_1\mathbf{f}_t + \mathbb{I}(z_t > \theta)\,\boldsymbol{\Lambda}_2\mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \ldots, T, \tag{1}$$

where T denotes the time series dimension of the available sample; $\mathbf{x}_t = (x_{1t}, \ldots, x_{Nt})' \in \Re^N$ is the $N \times 1$ vector of observable dependent variables; $\mathbf{f}_t = (f_{1t}, \ldots, f_{Rt})' \in \Re^R$ is the $R \times 1$ vector of latent factors; $z_t \in \Re$ is an observable covariate and θ is the unknown threshold value; $\mathbf{e}_t = (e_{1t}, \ldots, e_{Nt})' \in \Re^N$ is the $N \times 1$ vector of idiosyncratic errors; $\boldsymbol{\Lambda}_j = \left(\boldsymbol{\lambda}_{j1}, \ldots, \boldsymbol{\lambda}_{jN}\right)'$ is the $N \times R$ matrix of factor loadings with ith row defined as $\boldsymbol{\lambda}_{ji} = \left(\lambda_{ji1}, \ldots, \lambda_{jiR}\right)'$, for $j = 1, 2$ and $i = 1, \ldots, N$.

The model in (1) belongs to the class of threshold models proposed in Tong and Lim (1980): see Tsay (1989, 1998), Chan (1993) and Hansen (1996, 1999, 2000) for methodological contributions; and Hansen (2011) for a survey of the literature. According to the threshold principle introduced in Pearson (1900), the regime prevailing at time t depends on the position of z_t with respect to the unknown threshold θ. Ng and Wright (2013) simulate data from a large dimensional threshold factor model to investigate the effects of nonlinearities on business cycle dynamics[2]: we explicitly focus on estimation, model selection and linearity testing. Our results extend to the case in which the threshold variable is more generally defined as a linear combination of covariates (Massacci, 2014): this would be relevant when the driver of the regimes is not *a priori* known.

The model in (1) extends large dimensional linear factor models to allow for a threshold effect on the loadings. Given Assumption C3 stated in Section 3.1, we follow Chamberlain and Rothschild (1983) and allow for some degree of correlation in the idiosyncratic components within each regime: (1) then is an *approximate threshold factor model*; it is more general than an *exact threshold factor model*, which would extend the arbitrage pricing theory of Ross (1976) and would not allow for any correlation in the idiosyncratic components in any regime.

3. Estimation

As in Stock and Watson (2002), we study estimation of (1) under the assumption that the true number of factors R^0 (i.e., the true dimension of $\mathbf{f}_t$) is known. We extend the theory in Bai and Ng (2002) based on principal components estimation to allow for concentrated least squares estimation, as motivated in Hansen (2000) for threshold regressions. The plan is as follows: Section 3.1 states the assumptions; Section 3.2 deals with identification; Section 3.3 describes the principal components estimator; Section 3.4 proves the consistency of the estimator; and Section 3.5 derives the convergence rates.

[1] See Ng and Wright (2013), p. 1147.

[2] See Ng and Wright (2013), p. 1147.

Please cite this article in press as: Massacci, D., Least squares estimation of large dimensional threshold factor models. Journal of Econometrics (2016), http://dx.doi.org/10.1016/j.jeconom.2016.11.001