Economics Letters 116 (2012) 531–534

Contents lists available at SciVerse ScienceDirect

Economics Letters

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

Estimating the number of common factors in serially dependent approximate factor models[☆]

Ryan Greenaway-McGrevy[a,*], Chirok Han[b], Donggyu Sul[c]

[a] Bureau of Economic Analysis, 1441 L St. NW, BE-40, Washington, DC 20230, United States
[b] Korea University, Republic of Korea
[c] University of Texas at Dallas, United States

ARTICLE INFO

Article history:
Received 17 February 2012
Accepted 28 March 2012
Available online 25 April 2012

JEL classification:
C33

Keywords:
Factor model
Prewhitening
Least squares dummy variable (LSDV) filter
Factor number estimation
Cross-section dependence

ABSTRACT

A simple data-dependent filtering method is proposed before applying the Bai–Ng method to estimate the number of common factors in the conventional approximate factor model. The asymptotic justification is provided and the finite-sample performance is examined.

Published by Elsevier B.V.

1. Motivation and empirical example

In practice, factor number estimation can be difficult in panels that exhibit persistency. Intuitively, when either the cross-sectional units (N) or the time series observations (T) are not sufficiently large, weak dependence in the idiosyncratic component may be misinterpreted as dependence due to the factor structure, resulting in overfitting. Moreover, first-differencing (a typical treatment) can induce negative serial dependence, also leading to overestimation. This paper suggests that, in these situations, estimation can be sharpened by applying a data-dependent filter to the panel before applying the selection criteria, and we provide the theoretical justification of the filtering procedure. First, however, we provide an empirical example illustrating the benefits of data-dependent filtering using the Bai–Ng selection criteria.

For the approximate factor model $X_{it} = \lambda_i' F_t + e_{it}$, Bai and Ng (2002, BN hereafter) propose estimating the factor number r by minimizing

$$\mathrm{IC}(k) = \ln V_{NT}(k) + k g(N, T), \tag{1}$$

with respect to $k \in \{0, 1, \ldots, k_{\max}\}$ for some fixed $k_{\max}$, where $V_{NT}(k) = \min_{\{\lambda_i^k\},\{F_t^k\}} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_i^{k\prime} F_t^k)^2$ and $g(\cdot)$ is a penalty function. Setting $g(N, T) = \ln C_{NT}^2 [(N+T)/NT]$, where $C_{NT} = \min\left[\sqrt{T}, \sqrt{N}\right]$, yields the $\mathrm{IC}_{p2}(k)$ criterion, which is popular in practice, and has the largest penalty on the fitted factor number k, making it less sensitive to weak dependence in the idiosyncratic component than other $\mathrm{IC}(k)$ criteria.

We estimate the factor number to industry-level employment growth. Annual wage and salary employment by the North American Industry Classification System is obtained from table SA27 on the Bureau of Economic Analysis (BEA) website. We have 93 industries, and our sample spans 1990–2009. Employment growth is annual log-differences of total wage and salary employment. The panel is also standardized to remove the excessive heteroskedasticity.

$\mathrm{IC}_{p2}(k)$ in levels always selects the maximum number of factors. This result does not change with other $k_{\max}$. With first-differenced data, $\mathrm{IC}_{p2}(k)$ selects between 3 and 5 factors depending on the subsample (see Table 1). Note however that applying the $\mathrm{IC}_{p2}(k)$ criterion to the least squares dummy variable (LSDV) filtered panel yields a factor number estimate of 1. This result holds for both the full sample and the subsamples. Next, we discuss this filtering procedure and provide an asymptotic justification for the filter.

☆ The views expressed herein are those of the authors and not necessarily those of the Bureau of Economic Analysis or the Department of Commerce. The authors thank Yoosoon Chang and H. Roger Moon for helpful comments.

* Corresponding author. Tel.: +1 202 606 9844; fax: +1 202 606 5366.
E-mail address: ryan.greenaway-mcgrevy@bea.gov (R. Greenaway-McGrevy).

0165-1765/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.econlet.2012.03.031

Table 1
Estimated factor number; industry employment growth. Bai–Ng $\mathrm{IC}_{p2}(k)$ with $k_{\max} = 5, N = 92, T = 19$.

Sample	Level	First-difference (FD)	AR1
1990–2009	5	4	1
Subsample robustness			
1992–2009	5	5	1
1990–2007	5	3	1

2. Consistency of filtering procedure

The whitening filter has been used in many areas of econometrics. The basic idea is to attenuate the temporal dependence in the data to make the transformed data closer to white noise. We employ an autoregressive filtering, and as such we must first focus on two preliminary specification issues in order to ensure that the factor structure is preserved in the transformed data: (i) whether to perform an individual or a pooled filtering, and (ii) AR lag order.

To address the first issue, consider the transformed data

$$Z_{it} = X_{it} - \sum_{j=1}^{p} \phi_{ij} X_{it-j}, \quad X_{it} = \lambda_i' F_t + e_{it},$$

where the filter ϕ_{ij} is permitted to be different for each i. Writing Z_{it} as

$$Z_{it} = \lambda_i' \left(F_t - \sum_{j=1}^{p} \phi_{ij} F_{t-j} \right) + \left(e_{it} - \sum_{j=1}^{p} \phi_{ij} e_{it-j} \right),$$

we see that only when $\phi_{ij} = \phi_j$ (i.e., homogeneous for all i), are the common factors of $Z_{it} F_t - \sum_{j=1}^{p} \phi_j F_{t-j}$ and the dimension of factors is preserved under the transformation. Without the homogeneity restriction in the filtering coefficients, the filtered common component $\lambda_i'(F_t - \sum_{j=1}^{p} \phi_{ij} F_{t-j})$ cannot generally be expressed as a factor structure with the same dimension as F_t.

The second issue is the choice of the lag order p. Conveniently, an AR(1) fitting ($p = 1$)

$$Z_{it} = X_{it} - \phi X_{it-1} = \Delta X_{it} + (1 - \phi) X_{it-1} \tag{2}$$

is sufficient for consistent factor number estimation for many common panel processes, as we show below. Of course other orders p can also be used, but we do not see any particular advantage in using more lags unless e_{it} is more than once integrated. Hence we focus only on AR(1) filtering throughout the paper. Note that ϕ may not be a "true" AR(1) coefficient and that $X_{it} - \phi X_{it-1}$ may be dependent over t.

We therefore consider an LSDV estimator, $\hat{\phi}_{lsdv}$, obtained by regressing X_{it} on X_{it-1} and including individual intercepts. To show the validity of the AR(1) LSDV filtering, define $e_{it}^* = \sigma_{e,T}^{-2} e_{it}$ and $F_t^* = \Sigma_{FF,T}^{-1} F_t$. Note that e_{it} and F_t are divided by their variances rather than their standard deviations in the definition of e_{it}^* and F_t^*, so $(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E e_{it}^{*2} = \sigma_{e,T}^{-2}$ and $T^{-1} \sum_{t=1}^{T} F_t^* F_t^{*\prime} = \Sigma_{FF,T}^{-1}$. The reason for this normalization is both to ensure that the variables e_{it}^* and F_t^* behave regularly when the original processes e_{it} and F_t are stationary, and to ensure that e_{it}^* and F_t^* are negligible when e_{it} and F_t are integrated. Now Z_{it} of (2) can be rewritten as

$$Z_{it} = \lambda_i' [\Delta F_t + (1 - \phi) \Sigma_{FF,T} F_{t-1}^*] + [\Delta e_{it} + (1 - \phi) \sigma_{e,T}^2 e_{it-1}^*], \tag{3}$$

so the factors of the transformed series Z_{it} are $\Delta F_t + (1 - \phi) \Sigma_{FF,T} F_{t-1}^*$ and the idiosyncratic component is $\Delta e_{it} + (1 - \phi) \sigma_{e,T}^2 e_{it-1}^*$. If ϕ is chosen such that $(1 - \phi) \Sigma_{FF,T}$ and $(1 - \phi) \sigma_{e,T}^2$ are bounded, then these transformed factors and idiosyncratic components are likely to satisfy Bai and Ng's (2002) regularity (called BN-regularity hereafter). For a rigorous treatment along this line, we make the following assumptions.

Assumption 1. For any constant b_1 and b_2, $\{\lambda_i\}$, $\{\Delta F_t + b_1 F_{t-1}^*\}$ and $\{\Delta e_{it} + b_2 e_{it-1}^*\}$ are BN-regular.

Assumption 2. The common factors F_t and idiosyncratic errors e_{it} satisfy

$$\frac{1}{T} \sum_{t=1}^{T} E[F_{t-1} \Delta F_t'] = O(1) \quad \text{and}$$

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} E[e_{it-1} \Delta e_{it}] = O(1).$$

Theorem (*Consistency of LSDV Filtering*). *Given Assumptions 1 and 2, if $\sigma_{e,T}^2 = o(T)$ and $\Sigma_{FF,T} = o(T)$, then $\hat{k}(\hat{\phi}_{lsdv}) \to_p r$, where r is the true factor number and $\hat{k}\left(\hat{\phi}_{lsdv}\right)$ is the estimated factor from LSDV filtering.*

Assumptions 1 and 2 are satisfied by a variety of processes in e_{it} and F_t, including $I(1)$ and square summable $I(0)$, heterogenous, as well as local asymptotic processes. While first-differencing (FD) works well when the process is closer to unit root or even integrated, LSDV filtering typically performs better than first-differencing if the process does not exhibit strong serial correlation. In practice, the strength of the dependence is unknown, so it will be useful to provide a method which combines the two filtering methods and which is at least as good as the two filtering methods separately.

A simple way to enhance the small-sample performance is to choose the minimum factor number estimate from the first-differencing and the LSDV filtering, i.e.,

$$\hat{k}_{\min} = \min \left\{ \hat{k}(1), \hat{k}(\hat{\phi}_{lsdv}) \right\}. \tag{4}$$

This "minimum rule" is justified by the fact that serial correlation usually causes overestimation rather than underestimation of the factor number.

The method can also be applied to the restricted dynamic models based on $F_t = \sum_{j=1}^{p} \Pi_j F_{t-j} + G\eta_t$, where η_t is $q \times 1$ and G is $r \times q$ with full column rank (Amengual and Watson, 2007; Bai and Ng, 2007). For this model, filtering the data using the methods suggested above preserves the dimensions of both the static and dynamic factors, because

$$F_t - \phi F_{t-1} = \sum_{j=1}^{p} \Pi_j (F_{t-j} - \phi F_{t-j-1}) + G(\eta_t - \phi \eta_{t-1}),$$

where the transformed static factors $F_t - \phi F_{t-1}$ are still $r \times 1$ and the transformed primitive shocks $\eta_t - \phi \eta_{t-1}$ are still $q \times 1$.

3. Monte Carlo studies

We consider the following data-generating process (DGP):

$$X_{it} = \sum_{j=1}^{r} \lambda_{ji} F_{jt} + e_{it}, \quad F_{jt} = \theta F_{jt-1} + v_{jt} \text{ for } j = 1, \ldots, r;$$

$$e_{it} = \rho_i e_{it-1} + \varepsilon_{it}, \quad \varepsilon_{it} = \sum_{k=-J, k \neq 0}^{J} \beta u_{i-k,t} + u_{it} \text{ for } J = \lfloor N^{1/3} \rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the largest integer not exceeding the argument. Note that e_{it} can exhibit cross-sectional and time-series dependence through β and ρ_i, respectively. We draw $u_{it} \sim N(0, s_i^2)$, $v_{jt} \sim N(0, 1)$, and $\lambda_{ji} \sim N\left(0, r^{-1/2}\right)$. We set $r = 2$. We consider the $\mathrm{IC}_{p2}(k)$ criterion only because it uses the