Estimating the Number of Common Factors in Serially Dependent Approximate Factor Models*

Ryan Greenaway-McGrevy†
Bureau of Economic Analysis

Chirok Han
Korea University

Donggyu Sul
University of Texas at Dallas

February 17, 2012

Abstract

A simple data-dependent filtering method is proposed before applying the Bai-Ng method to estimate the factor number. The asymptotic justification is provided and the finite sample performance is examined.

Keywords: Factor Model, Prewhitening, LSDV filter, Factor Number Estimation, Cross Section Dependence

JEL Classification: C33

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 over-fitting. 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 a 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.

*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: r.greenaway-mcgrevy@bea.gov. Address: 1441 L St. NW, BE-40, Washington, DC 20230. Phone: +1 (202) 606-9844. Fax: +1 (202) 606 5366

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

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

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 IC(k) criteria.

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

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
sub-sample robustness			
1992-2009	5	5	1
1990-2007	5	3	1

$\mathrm{IC}_{p2}(k)$ in levels always selects the maximum number of factors. This result does not change with other $k_{\max}$. With FD data, $\mathrm{IC}_{p2}(k)$ selects between 3 to 5 factors depending on the sub-sample. Note however that applying the $\mathrm{IC}_{p2}(k)$ criterion to the LSDV filtered panel yields a factor number estimate of one. 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.

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), the common factors of Z_{it} are $F_t - \sum_{j=1}^{p} \phi_j F_{t-j}$ 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 $X_{it} - \phi X_{it-1}$ 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 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*

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

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 works well when the process is closer to unit root or even integrated, the 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:

$$\begin{aligned} X_{it} &= \textstyle\sum_{j=1}^{r} \lambda_{ji} F_{jt} + e_{it}, \quad F_{jt} = \theta F_{jt-1} + v_{jt} \text{ for } j = 1, \ldots, r; \\ e_{it} &= \rho_i e_{it-1} + \varepsilon_{it}, \quad \varepsilon_{it} = \sum_{k=-J, k\neq 0}^{J} \beta u_{i-k,t} + u_{it} \text{ for } J = \lfloor N^{1/3} \rfloor. \end{aligned}$$

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 largest penalty, so the probability of overestimation is the smallest amongst BN's IC criteria. We consider the following two cases.

Case 1: Moderate heterogenous idiosyncratic serial dependence We generate the moderate heterogeneity by setting $\rho_i \sim U[-0.1, 0.9]$, $s_i \sim U[0.5, 1.5], \theta = 0.5, \beta = 0.1$, and $\sigma_v^2 = 1$. Table 2 shows the results with 2,000 replications. Evidently the AR1 filter outperforms the FD filter by a wide margin either N or T is small. The minimum rule improves on the AR1 filter slightly, while $\mathrm{IC}_{p2}(k)$ in levels performs poorly.

Case 2: Extremely heterogenous autoregressive parameters In this set of simulations we draw ρ_i from *iid* $U[-0.1, 0.1]$ for $i = 1, \ldots, \frac{1}{2}N$ and *iid* $U[0.7, 0.9]$ for $i = \frac{1}{2}N + 1, \ldots, N$. Other settings are as in Case 1. As in Case 1 above, in this framework there is a marked disparity between the degree of serial dependence in the AR1 parameter. But within each subgroup the degree of heterogeneity is low. Table 3 exhibits the results. Notably the AR1 filter performs much better than the FD or LEV methods. As in case 1, this result is attributable to the FD filter inducing large negative correlation in many of the cross sections. The effect is more noticeable in this DGP because more cross sections have an AR(1) coefficient close to zero.

Table 2: Finite Sample Performances with Moderate Heterogeneity
$\rho_i \sim U[-0.1, 0.9]$, $s_i \sim U[0.5, 1.5], \theta = 0.5, \beta = 0.1, \sigma_v^2 = 1$

		LEV			FD			AR1			MIN		
N	T	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$
25	25	4.8	26.3	68.9	16.8	57.9	25.3	23.7	68.9	7.4	27.4	67.6	5.0
50	25	0.0	1.9	98.1	2.3	82.8	14.9	2.7	88.1	9.2	3.2	93.9	2.9
100	25	0.0	0.7	99.3	0.8	94.6	4.6	1.3	95.6	3.1	1.5	98.2	0.3
25	50	0.3	9.9	89.8	2.4	69.3	28.3	5.2	88.2	6.6	5.5	89.7	4.8
25	100	0.1	4.9	95.0	0.7	68.6	30.7	0.6	95.2	4.2	0.9	95.6	3.5

Note: $<, =, >$: under, correct, and over estimation, respectively.

Table 3: Finite Sample Peformance with Extreme Heterogeneity
$\rho_i \sim U[-0.1, 0.1]$ for $i \leq N/2$ and $\rho_i \sim U[0.7, 0.9]$ for $i > N/2$

		LEV			FD			AR1			MIN		
N	T	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$
25	25	4.8	23.4	71.8	17.1	40.6	42.3	25.0	60.6	14.4	28.4	60.7	10.9
50	25	0.1	4.0	95.9	2.3	51.3	46.4	4.1	80.2	15.7	4.8	84.3	10.9
100	25	0.0	0.1	99.9	0.7	79.1	20.2	1.5	86.6	11.9	1.8	95.4	2.8
25	50	0.0	5.5	94.5	2.0	43.4	54.6	5.6	80.8	13.6	5.9	82.7	11.4
25	100	0.0	1.3	98.7	0.5	36.9	62.6	1.3	85.0	13.7	1.4	86.7	11.9

We also considered other parameters settings and DGPs, as well as other estimation methods such as the Hallin and Liska (2007) cross validation approach. In the interests of brevity we do not report the results here, but are avaliable from the authors upon request.

4 Conclusion

We demonstrate that even a moderate degree of serial correlation in the idiosyncratic errors (relative to the given sample size) can cause the Bai-Ng criteria to overestimate the true factor number. To overcome this problem, we suggest filtering the panel. We theoretically analyze how the filtering method can work for general processes with serial correlation and verify the applicability of the method by simulations.

5 Appendix

The following three lemmas are helpful to prove Theorem.

Lemma 1 Under Assumption 1, if (i) $(1-\phi)\,\Sigma_{FF,T}$ converges to a finite limit, (ii) $(1-\phi)\,\sigma^2_{e,T} = O(1)$ and (iii) $(\hat{\phi}-\phi)\sigma^2_{X,T} \to_p 0$, then $\hat{k}(\hat{\phi}) \to_p r$.

Lemma 2 If $T^{-1}\sigma^2_{X,T} = O(1)$, then under Assumption 2, $(1-\phi_{lsdv})\,\sigma^2_{X,T} = O(1)$.

Lemma 3 Suppose that (i) $\mathrm{var}(X^2_{it-1}) \leq M\sigma^4_{X,T}$, and (ii) $\left|\sum_{k=1}^{\infty}\mathrm{cov}(X^2_{it}, X^2_{it+k})\right| \leq M\sigma^4_{X,T}$ for all i and t for some $M < \infty$. If $T^{-1}\sigma^2_{X,T} = o(1)$, then $(\hat{\phi}_{lsdv} - \phi_{lsdv})\sigma^2_{X,T} = o_p(1)$.

Proof of Lemma 1 Let $\hat{a} = (\hat{\phi} - \phi)\sigma^2_{X,T}$, $\hat{h}(k) = h_{NT}(k; \hat{Z}_{it})$ and $h(k) = h_{NT}(k; Z_{it})$. The goal is to show that (i) $\hat{h}(k)$ does not shrink to zero for $k < r$, and (ii) $\hat{h}(k) = O(C_{NT}^{-2})$ for $k > r$. (See Bai and Ng, 2002, proof of Theorem 2.) Note that $\hat{Z}_{it} = Z_{it} + \hat{a}X^*_{it-1}$, where $X^*_{it} := X_{it}/\sigma^2_{X,T}$.

(i) When $k < r$ **:** It suffices to show that $\hat{h}(k) - h(k) \to_p 0$. But $\hat{h}(k) - h(k) = \hat{\xi}_r - \hat{\xi}_k$, where

$$\hat{\xi}_j = \max_{F \in \Re^{T\times j}} \frac{1}{NT}\sum_{i=1}^{N} \hat{Z}_i' P_F \hat{Z}_i - \max_{F \in \Re^{T\times j}} \frac{1}{NT}\sum_{i=1}^{N} Z_i' P_F Z_i.$$

Note that $|\max f - \max g| \leq \max |f - g|$, so $|\hat{\xi}_j| \to_p 0$ since $\hat{a} \to_p 0$ and all the averages are stochastically bounded.

(ii) For the case with $k > r$ **:** All conditions satisfy BN-reqular. But Lemma 4 in BN should be proved with new assumptions. Particularly we have to show that

$$\max_{F \in \Re^{T\times k}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(u_i - \hat{a}e^*_{i,-1})' M_F (u_i - \hat{a}e^*_{i,-1}) = O_p(C_{NT}^{-2}),$$

which corresponds to (1) of Bai and Ng (2006). Since both u_{it} and and e^*_{it-1} satisfy the assumptions of Bai and Ng (2002, 2006) and $\hat{a} \to_p 0$, the left term becomes $O_p(C_{NT}^{-2})$.

Proof of Lemma 2 Define $\hat{\phi}_{lsdv} = \hat{\Gamma}_0^{-1}\hat{\Gamma}_1$ and ϕ_{lsdv} =$\mathrm{E}\left(\hat{\Gamma}_0^{-1}\right)\mathrm{E}\left(\hat{\Gamma}_1\right) = \Gamma_{0,T}^{-1}\Gamma_{1,T}$. Since $(1 - \phi_{lsdv})\sigma^2_X = \Gamma_{0,T}^{-1}(\Gamma_{0,T} - \Gamma_{1,T})\sigma^2_{X,T}$ and $\Gamma_{0,T}^{-1}$ is finite, it suffices to show that $(\Gamma_{0,T} - \Gamma_{1,T})\sigma^2_{X,T} = O(1)$ or $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\tilde{X}_{it-1}\Delta\tilde{X}_{it} = O(1)$ where the "~" notation stands for the within-group transformation. It is easy to show since

$$\frac{1}{T}\sum_{t=1}^{T} E\tilde{X}_{it-1}\Delta\tilde{X}_{it} = \frac{1}{T}\sum_{t=1}^{T} EX_{it-1}\Delta X_{it} - \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} EX_{it-1}\Delta X_{is},$$

and the first and second terms are bounded by Assumption 2.

Proof of Lemma 3 Under Assumption 2 and conditions in Lemma 3, the proof of Lemma 3 is boiled down to the boundary condition of

$$\frac{1}{T^2}\sum_{t=1}^{T} \mathrm{var}\left(X_{it-1}X^*_{it-1}\right) + \frac{2}{T^2}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \mathrm{cov}\left(X_{it-1}X^*_{it-1}, X_{is-1}X^*_{is-1}\right).$$

Since $\mathrm{var}(X_{it}X^*_{it})$ =$\mathrm{var}(X^2_{it})/\sigma^4_X = O(\sigma_X^{-2}) = O(1)$, so the first term is $O(T^{-1})$, and the second term is also $O(T^{-1})$. By using this boundary condition, Lemma 3 can be proved. The detailed proof is omitted.

Proof of Theorem The first differenced process ΔX_{it} clearly gives a consistent estimate. For $X_{it} - \hat{\phi}_{lsdv} X_{it-1}$, we note that the assumptions that $T^{-1}\sigma^2_{e,T} = o(1)$ and $T^{-1}\Sigma_{F,T} = o(1)$ imply that $T^{-1}\sigma^2_{X,T} = o(1)$. Then it is straightforward to see that conditions for Lemma 2 and 3 are satisfied under the regularity Assumptions 1-2. The result follows from Lemmas 2 and 3.

References

[1] Amengual, D. and M. W. Watson, 2007, Consistent estimation of the number of dynamic factors in a large N and T panel, Journal of Business and Economic Statistics 25, 91–96.

[2] Bai, J. and S. Ng, 2002, Determining the number of factors in approximate factor models, Econometrica 70, 191–221.

[3] Bai, J. and S. Ng, 2006, Determining the number of factors in approximate factor models, Errata, available at http://www.columbia.edu/~sn2294/papers/correctionEcta2.pdf.

[4] Bai, J. and S. Ng, 2007, Determining the number of primitive shocks in factor models, Journal of Business and Economic Statistics 26, 52–60.

[5] Hallin, M. and R. Liska, 2007, Determining the number of factors in the generalized dynamic factor model, *Journal of the American Statistical Association* 102, 603–617.