Determining the number of factors in approximate factor models, Errata
by Jushan Bai and Serena Ng[1]

The right hand side of equation (11) of Bai and Ng (2002) has the opposite signs intended, resulting an incorrect inequality on page 219 (line 2). To complete the proof of Lemma 4, we need to show that

$$(NT)^{-1} \sum_{i=1}^{N} \underline{e}_i' P_{\hat{F}}^k \underline{e}_i = O_p(C_{NT}^{-2}) \tag{1}$$

where $C_{NT}^2 = \min(N,T)$. We show in this correction that (1) is true by a different argument. The above is bounded by the sum of the first k largest eigenvalues of the matrix $A_{NT} = \frac{1}{NT}ee'$, where $e = (e_{it})$, $N \times T$. Therefore, it is sufficient to show the largest eigenvalue of A_{NT} is of order $O_p(C_{NT}^{-2})$.

Let $\rho(A)$ denote the largest eigenvalue of a matrix A. If e_{it} are iid with finite fourth moment, then $\rho(A_{NT}) = O_p(C_{NT}^{-2})$, see Yin, Bai, and Krishnaiah (1988). The iid assumption can be replaced by independence with uniformly bounded 7th moment, see Jonsson (1985), whose proofs only use independence though iid is assumed. In our context, we need to show that (1) also holds under weak cross-section and time series dependence as well as heteroskedasticities.

Let $\xi = (\xi_{it})$ be an arbitrary $N \times T$ matrix consisting of independent elements with uniformly bounded 7th moment and $E(\xi_{it}) = 0$. Let Σ $(N \times N)$ and R $(T \times T)$ be arbitrary non-random positive definite matrices. Let $e = \Sigma^{1/2}\xi R^{1/2}$. This allows correlation and heteroskedasticity in both dimensions of e. We can also permit $var(\xi_{it}) = \omega_{it}^2$. Weak cross-section and serial correlations dictate that Σ and R have bounded eigenvalues, as required by approximate factor structure. We have

$$ee' = \Sigma^{1/2}\xi R \xi' \Sigma^{1/2} \tag{2}$$

Note that $\rho(ee') \le \rho(\Sigma)\rho(R)\rho(\xi'\xi)$, and $\xi'\xi$ and $\xi\xi'$ have identical nonzero eigenvalues. By assumptions on ξ, $\rho(\frac{1}{NT}\xi'\xi) = O_p(C_{NT}^{-2})$, it follows that (1) holds if the correlations in both dimensions are weak. Note that (2) is somewhat more restrictive than our original assumptions since it requires a matrix ξ whose elements are independent. But the key features of correlation and heteroskedasticity in both dimensions of e are preserved. Furthermore, both Σ and R can be random, not necessarily independent of ξ for our purpose, provided that their eigenvalues are bounded, see Onatski (2005), and Bai and Silverstein (1999). Alternatively, Amengual and Watson (2005) show that if $\frac{1}{NT}E\{tr[(ee')^j]\} \le M[\max(N,T)]^{j-1}$ for all $j \ge 1$ with some $M < \infty$, then (1) holds.

References

1. Amengual, D. and M.W. Watson (2005), "Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel," Unpublished manuscript, Princeton University.

2. Bai, J. and Ng, S (2002). "Determining the number of factors in approximate factor models", Econometrica, 70, pp 191-221.

3. Bai, Z.D. and J. W. Silverstein (1999): "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability Vol. 27, No. 3, 1536-1555.

4. Jonsson, D. (1985): "On the largest eigenvalue of a sample covariance matrix," Multivariate Analysis VI (edited by P.R. Krishnaiah), North Holland, 327-333.

5. Onatski, A. (2005): "Determining the Number of Factors from Empirical Distribution of Eigenvalues," unpublished manuscript, Columbia University.

6. Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988): "On the limit of the largest eigenvalue of large dimensional sample covariance matrix," Probability Theory and Related Fields 78, 509-521.

[1] We thank Mark Watson for alerting us to the error and for useful discussions.