High-Dimensional Dynamic Factor Models: Theory and Applications to Forecasting and Macroeconomic Analysis.

Marco Lippi, Istituto Einaudi per l'Economia e la Finanza (EIEF), Roma

November 30, 2023

High Dimensional Data: Theory and applications

Laboratori Nazionali del Gran Sasso

The dataset

Much of the empirical research on High-Dimensional Dynamic factor Models has been conducted on a monthly Macroeconomic dataset containing about $n = 200$ time series for the US:
(1) output and income, (2) labor market, (3) housing, (4) consumption, orders and inventories, (5) money and credit, (6) bond and exchange rates, (7) prices, and (8) stock market. The series length T is about 500.

This is the meaning of “big data” in this literature. Indeed it is big data as compared to standard Applied Macroeconomic Analysis, in which VAR models are employed. A VAR, Vector Auto-Regression, is, for example,

$$X_t = AX_{t-1} + U_t,$$

where X_t is a, say, the 5-dimensional vector including GDP, Investment, Consumption, Unemployment, Interest Rate, so that A contains 25 parameters which must be estimated. And you see that with 10 variables you should estimate 100 parameters

This dependence of the number of parameters to estimate on the square of the number of variables is an example of "curse of dimensionality". You would never think of estimating a VAR with 100 variables.
Note that "curse of dimensionality" is relative to the number T. But 500, a little more or less, is the limit with monthly macroeconomic data (structural break).

Modeling high-dimensional datasets. Factors

High-dimensional factor models are based on the idea that the series of the dataset are determined by a small number of factors, that are common to all series, plus on individual cause of variation. Let me use an elementary example: for $i = 1, \dots, n$,

$$x_{it} = b_i F_t + \xi_{it},$$

where

(a) Everything is zero-mean covariance stationary.

(b) F_t and ξ_{it} are non-correlated.

(c) ξ_{it} and ξ_{jt}, $i \neq j$, are non-correlated, that is, the ξ's are individual specific.

(d) The variables x_{it} are observable, whereas F_t and ξ_{it} are latent.

$$x_{it} = b_i F_t + \xi_{it},$$

where

(a) Everything is covariance stationary and zero-mean.

(b) F_t and ξ_{it} are non-correlated.

(c) ξ_{it} and ξ_{jt}, $i \neq j$, are non-correlated.

(d) The variables x_{it} are observable, whereas F_t and ξ_{it} are latent.

Suppose that the ξ's are unpredictable and big with respect to $b_i F_t$. Then the x's will be little predictable. If we are able to estimate F_t and F_t is predictable, then our prediction of the x's improves. More on this later.

Estimation of F_t. Take the average:

$$X_{nt} = \frac{1}{n}\sum_{i=1}^{n} x_{it} = \left(\frac{1}{n}\sum_{i=1}^{n} b_i\right) F_t + \frac{1}{n}\sum_{i=1}^{n}\xi_{it}$$

The variances are

$$\mathrm{var}\frac{1}{n}\sum_{i=1}^{n} x_{it} \le \left(\frac{1}{n}\sum_{i=1}^{n} b_i\right)^2 \sigma_F^2 + \frac{1}{n^2} n \max_i \mathrm{var}\xi_{it} = \bar{b}_n^2\sigma_F^2 + \frac{1}{n}M_n$$

So we see that in the limit the ξ's disappear in the average under the assumption that M_n is bounded.

Now however, what if $\bar{b}_n \to 0$? We might use, instead of the weights $1/n$, the weights $b_i/\sqrt{b_1^2 + \cdots + b_n^2}$. Now the common component cannot tend to zero.

But this is not feasible because the coefficients b_i are not observable.

What we do is the following. Consider the $n \times n$ variance-covariance matrix of the x's, Γ_0^x. Then let W_x the column eigenvector corresponding to the first eigenvalue of Γ_n^x. We use W_x as weights. Let me recall you that $P_t^x = W_1^x x_{1t} + \cdots + W_n^x x_{nt}$ is known as the first Principal Component of the x vector.

We can show that

$$W_i^x P_t = W_i^x \left(W_1^x \cdots W_n^x \right) \begin{pmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{pmatrix} \to b_i F_t,$$

in mean square as $n \to \infty$, with rate $1/\sqrt{n}$.

But this is also non feasible because we do not observe Γ_x^n. We actually use the estimated covariance matrix $\hat{\Gamma}_n^x$. The corresponding estimator of $b_i F_t$, that is $\hat{W}_i^x \hat{P}_t^x$ converges in probability to $b_i F_t$ with rate $\max \left(1/\sqrt{n}, 1/\sqrt{T} \right)$.

Now suppose that, for example:

$$F_t = \alpha F_{t-1} + u_t,$$

where u_t is a white noise. Thus estimation of $b_i F_t$ and orthogonality of $b_i F_t$ and ξ_{it} allows predicting x_{it} by separately predicting F_t and ξ_{it}, which is an obvious advantage with respect to predicting x_{it} directly.

Of course it is necessary that we have a decent estimation of $b_i F_t$ and ξ_{it}.

The model can obviously be generalised for many factors.

$$x_{it} = b_{i1}F_{1t} + \cdots + b_{r1}F_{rt} + \xi_{it} = \chi_{it} + \xi_{it}.$$

The variables χ are called the common components and the ξ the idiosyncratic components. For example, the observable variables in our macroeconomic dataset are driven by a factor representing change in technology and another representing demand, so that $r = 2$. The model is estimated by means of the eigenvectors corresponding to the first r eigenvalues of $\hat{\Gamma}^x_n$ and

$$\hat{W}^x_{i1}\hat{P}^x_{1t} + \cdots + \hat{W}^x_r\hat{P}^x_{rt} \to \chi_{it},$$

as n and T at rate $\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)$.

Still this is not completely feasible because the integer r is not observable. So it must be estimated. More on this later.

Now there is something to say about the model:

$$x_{it} = b_{i1}F_{1t} + \cdots + b_{r1}F_{rt} + \xi_{it} = \chi_{it} + \xi_{it}.$$

Forget estimation, now we pretend to know the covariance matrix of the χ's and the ξ's.

Firstly, the assumption that the ξ's must be non correlated to one another can be relaxed. Some "weak" correlation can be allowed. For example, ξ_{1t} can be correlated to a finite number of other idiosyncratic components.

Secondly, some assumptions must be made on χ's to prevent that their covariance matrix become singular as $n \to \infty$.

Fom previous slide.

$$x_{it} = b_{i1} F_{1t} + \cdots + b_{r1} F_{rt} + \xi_{it} = \chi_{it} + \xi_{it}.$$

Firstly, the assumption that the ξ's must be non correlated to one another can be relaxed. Some "weak" correlation can be allowed. For example, ξ_{1t} can be correlated to a finite number of other idiosyncratic components.

Secondly, some assumptions must be made on χ's to prevent that their covariance matrix become singular as $n \to \infty$. So we assume:

(1) Let μ^{ξ}_{1n} be the first eigenvalue of Γ^{ξ}_{n}. There exists R such that $\mu^{\xi}_{1n} \leq R$ for all n.

(2) Let μ^{χ}_{rn} be the r-th eigenvalue of Γ^{χ}_{n}. $\lim_{n \to \infty} \mu^{\chi}_{rn} = \infty$.

From previous(1) Let μ^{ξ}_{1n} be the first eigenvalue of Γ^{ξ}_{n}. There exists R such that $\mu^{\xi}_{1n} \leq R$ for all n.

(2) Let μ^{χ}_{rn} be the r-th eigenvalue of Γ^{χ}_{n}. $\lim_{n\to\infty} \mu^{\chi}_{rn} = \infty$.

It is under (1) and (2) that we have

$$W^{x}_{i1} P^{x}_{1t} + \cdots + W^{x}_{r} P^{x}_{rt} \to \chi_{it} \quad \text{and} \quad \hat{W}^{x}_{i1} \hat{P}^{x}_{1t} + \cdots + \hat{W}^{x}_{r} \hat{P}^{x}_{rt} \to \chi_{it} \qquad (*)$$

in mean square and in probability, respectively.

Now you will object that these are assumptions on unobservable variables. But we have:

Theorem. Conditions (1) and (2) hold if and only if, as $n \to \infty$,
(3) $\mu^{x}_{rn} \to \infty$, there exists S such that $\mu^{x}_{r+1,n} \leq S$.
Thus under (3) we have $(*)$.

Previous (3) $\mu^x_{rn} \to \infty$, there exists S such that $\mu^x_{r+1,n} \leq S$.

Summing up, under (3)

(A) the integer r can be consistently estimated;

(B) $\hat{W}^x_{i1} \hat{P}^x_{1t} + \cdots + \hat{W}^x_{r} \hat{P}^x_{rt} \to \chi_{it}$.

Now suppose that

$$F_t = AF_{t-1} + R_t,$$

that is, F_t is generated by an r-dimensional VAR. Then the vector $\chi_t = (\chi_{1t} \ \cdots \ \chi_{rt})$ also is generated by a VAR

$$\chi_t = B\chi_{t-1} + V_t.$$

This VAR can be used to predict for example x_{1t} (a generalization of what we have seen in the elementary example with $r = 1$).

Now suppose that

$$F_t = AF_{t-1} + R_t, \qquad (**)$$

that is, F_t is generated by an r-dimensional VAR. Then the vector $\chi_t = (\chi_{1t} \ \cdots \ \chi_{rt})$ also is generated by a VAR

$$\chi_t = B\chi_{t-1} + V_t.$$

This VAR can be used to predict for example x_{1t} (a generalization of what we have seen in the elementary example with $r = 1$).
Of course this is an improvement with respect to using a VAR for $(x_{1t} \ \cdots \ x_{rt})$.
Lastly, note that estimations of the VAR $(**)$ implies that r is small, otherwise we fall again in the curse of dimensionality.

Again the model

$$x_{it} = b_{i1} F_{1t} + \cdots + b_{r1} F_{rt} + \xi_{it} = \chi_{it} + \xi_{it}.$$

I proposed the example in which $r = 2$, with the two factors interpreted as technology and demand. The model can accommodate also, for example,

$$x_{it} = b_i u_t + c_i u_{t-1} + \xi_{it}, \qquad (\dagger)$$

where u_t is a white noise. Indeed, by setting $F_{1t} = u_t$ and $F_{2t} = u_{t-1}$, we have $r = 2$ and

$$x_{it} = b_i F_{1t} + c_i F_{2t} + \xi_{it}. \qquad (\dagger\dagger)$$

In this case we say that the model has 2 static factors F_{1t} and F_{1t}, and 1 dynamic factor, u_t. Thus the dynamics in $(\dagger)$ has been transformed into the static representation $(\dagger\dagger)$.

However, there are fairly elementary cases in which this transformation is not possible. Suppose that χ_{it} is generated by the autoregressive equation

$$\chi_{it} = \alpha_i \chi_{i,t-1} + u_t,$$

where the coefficients α_i are drawn from the uniform distribution between $-.9$ and $.9$. We write

$$\chi_{it} = \frac{1}{1 - \alpha_i L} u_t = u_t + \alpha_i \chi_{t-1} + \alpha_i^2 u_{t-2} + \cdots ,$$

where L is the lag operator: $Ly_t = y_{t-1}$. Setting

$$x_{it} = \chi_{it} + \xi_{it} = \frac{1}{1 - \alpha_i L} u_t + \xi_{it},$$

the static representation $x_{it} = b_{i1} F_{1t} + \cdots + b_{r1} F_{rt} + \xi_{it}$ is impossible.

You see that in the example we have $r > q$, where q is the dimension of the dynamic. This is a general fact.

To accommodate the case

$$x_{it} = \chi_{it} + \xi_{it} = \frac{1}{1 - \alpha_i L} u_t + \xi_{it},$$

we introduce a more general model:

$$x_{it} = [b_{io} u_t + b_{i1} u_{t-1} + \cdots] + \xi_t = b_i(L) u_t + \xi_t$$

The analysis of this model requires consideration of the spectral density matrix of $(x_{1t} \ \cdots \ x_{nt})$. Under the assumptions that
u_t is a q-dimensional white noise,
$\xi_{it} \perp u_{t-s}$ for all i, t and s, we have

Previous

$$x_{it} = [b_{io} u_t + b_{i1} u_{t-1} + \cdots] + \xi_t = b_i(L) u_t + \xi_t$$

The analysis of this model requires consideration of the spectral density matrix of $(x_{1t} \ \cdots \ x_{nt})$. Under the assumptions that
u_t is a q-dimensional orthonormal white noise,
$\xi_{it} \perp u_{t-s}$ for all i, t and s, we have, setting $B_n(L) = (b_1(L) \ \cdots \ b_n(L))'$,

$$\Sigma_n^x(\theta) = B_n(e^{-i\theta}) B_n'(e^{i\theta}) + \Sigma_n^\xi(\theta).$$

Then:
We take the first q eigenvalues $\lambda_{nj}^x(\theta)$ and corresponding normalized $n \times 1$ eigenvectors $Z_{nj}^x(\theta)$.
Transform the eigenvectors back in the time domain where they produce $n \times 1$ filters $Z_{nj}^x(L)$ and produce the first q dynamic principal components:

$$P_{jt}^x = Z_{nj}^x(L)(x_{1t} \ \cdots \ x_{nt})'.$$

Lastly we obtain the estimator of χ_{it}:

$$Z^x_{n1,i}(L^{-1})P^x_{1t} + \cdots + Z^x_{nr,i}(L^{-1})P^x_{rt}.$$

EXAMPLE. The model is

$$\begin{cases} x_{it} = u_t + \xi_{it} \text{ if } i \text{ is odd} \\ x_{it} = u_{t-1} + \xi_{it} \text{ if } i \text{ is even} \end{cases}$$

where the ξ's are unit variance white noises. This is a stylised example with leading and lagging variables. For n even,

$$\Sigma^x_n(\theta) = \begin{pmatrix} 1 \\ e^{-i\theta} \\ \vdots \\ 1 \\ e^{-i\theta} \end{pmatrix} \begin{pmatrix} 1 & e^{i\theta} & \cdots & 1 & e^{i\theta} \end{pmatrix} + I_n$$

The first eigenvalue and corresponding eigenvector are

$$\lambda_n^x(\theta) = n + 1, \quad Z_{n1}^x(\theta) = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & e^{i\theta} & \cdots & 1 & e^{i\theta} \end{pmatrix}'.$$

The filter corresponding to $Z_{n1}^x(\theta)$ is

$$\frac{1}{\sqrt{n}} \begin{pmatrix} 1 & L^{-1} & \cdots & 1 & L^{-1} \end{pmatrix}'$$

and the first principal component

$$\begin{aligned} &\frac{1}{\sqrt{n}} \left(x_{1t} + L^{-1} x_{2t} \cdots + x_{n-1,t} + L^{-1} x_{nt} \right) \\ &\qquad = \sqrt{n}\, u_t + \frac{1}{\sqrt{n}} \left(\xi_{1t} + \xi_{2,t+1} \cdots + \xi_{n-1,t} + \xi_{n,t+1} \right). \end{aligned}$$

and the first principal component

$$\frac{1}{\sqrt{n}}\left(x_{1t} + L^{-1}x_{2t}\cdots + x_{n-1,t} + L^{-1}x_{nt}\right)$$
$$= \sqrt{n}\, u_t + \frac{1}{\sqrt{n}}\left(\xi_{1t} + \xi_{2,t+1}\cdots + \xi_{n-1,t} + \xi_{n,t+1}\right).$$

The estimates of χ_{1t} and χ_{2t}, for example, are

$$u_t + \frac{1}{n}\left(\xi_{1t} + \xi_{2,t+1}\cdots + \xi_{n-1,t} + \xi_{n,t+1}\right)$$

and

$$u_t + \frac{1}{n}\left(\xi_{1,t-1} + \xi_{2,t}\cdots + \xi_{n-1,t-1} + \xi_{n,t}\right)$$

respectively. You see the “realignment” induced by the dynamic principal component.

Modeling high-dimensional datasets. Example.

The model

$$\begin{cases} x_{it} = u_t + \xi_{it} \text{ if } i \text{ is odd} \\ x_{it} = u_{t-1} + \xi_{it} \text{ if } i \text{ is even} \end{cases}$$

can be rewritten as

$$x_{it} = b_{i1} F_{1t} + b_{i2} F_{2t} + \xi_{it},$$

where $F_{1t} = u_t$, $F_{2t} = u_{t-1}$ and the b's are defined in an obvious way.
In the static framework

$$S_n^x = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 0 & 1 \end{pmatrix} + I_n$$

Modeling high-dimensional datasets. Example.

Eigenvalues and eigenvectors are

$$\mu^x_{n1} = \mu^x_{n2} = n/2 + 1$$
$$W^x_{n1} = \frac{1}{\sqrt{n/2}} \begin{pmatrix} 1 & 0 & \cdots & 1 & 0 \end{pmatrix}$$
$$W^x_{n2} = \frac{1}{\sqrt{n/2}} \begin{pmatrix} 0 & 1 & \cdots & 0 & 1 \end{pmatrix}$$

The principal components are:

$$P^x_1 = \sqrt{(n/2)} u_t + \frac{1}{\sqrt{(n/2)}} (\xi_{1t} + \xi_{3t} + \cdots \xi_{n-1t})$$
$$P^x_2 = \sqrt{(n/2)} u_t + \frac{1}{\sqrt{(n/2)}} (\xi_{2t} + \xi_{4t} + \cdots \xi_{nt})$$

The principal components are:

$$P_1^x = \sqrt{(n/2)}u_t + \frac{1}{\sqrt{(n/2)}}\xi_{1t} + \xi_{3t} + \cdots + \xi_{n-1,t}$$

$$P_2^x = \sqrt{(n/2)}u_t + \frac{1}{\sqrt{(n/2)}}(\xi_{2t} + \xi_{4t} + \cdots \xi_{nt})$$

and the estimate of χ_{1t} is

$$u_t + \frac{1}{n/2}\left(\xi_{1t} + \xi_{3t} + \cdots + \xi_{n-1} + \xi_{2t} + \xi_{4t} + \cdots \xi_n\right)$$

So you see that the dynamic approach is two times more efficient in the elimination of the idiosyncratic component.

Modeling high-dimensional datasets. Example.

However, let us go back to:
The first eigenvalue and corresponding eigenvector are

$$\lambda_n^x(\theta) = n + 1, \quad Z_{n1}^x(\theta) = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & e^{i\theta} & \cdots & 1 & e^{i\theta} \end{pmatrix}'.$$

The filter corresponding to $Z_{n1}^x(\theta)$ is

$$\frac{1}{\sqrt{n}} \begin{pmatrix} 1 & L^{-1} & \cdots & 1 & L^{-1} \end{pmatrix}' \qquad (\ddagger)$$

and the first principal component

$$\frac{1}{\sqrt{n}} \left(x_{1t} + L^{-1} x_{2t} \cdots + x_{n-1,t} + L^{-1} x_{nt} \right)$$
$$= \sqrt{n}\, u_t + \frac{1}{\sqrt{n}} \left(\xi_{1t} + \xi_{2,t+1} \cdots + \xi_{n-1,t} + \xi_{n,t+1} \right).$$

You see that the filter $(\ddagger)$ is two sided, which a serious drawback.

First Papers

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: identification and estimation. The Review of Economics and Statistics 82, 540–554.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20, 147–162.

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

Literature

Relatively recent

Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. Journal of Econometrics 185, 359–371.

Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2017). Dynamic factor models with infinite dimensional factor space: Asymptotic analysis. Journal of Econometrics 199, 74–92.