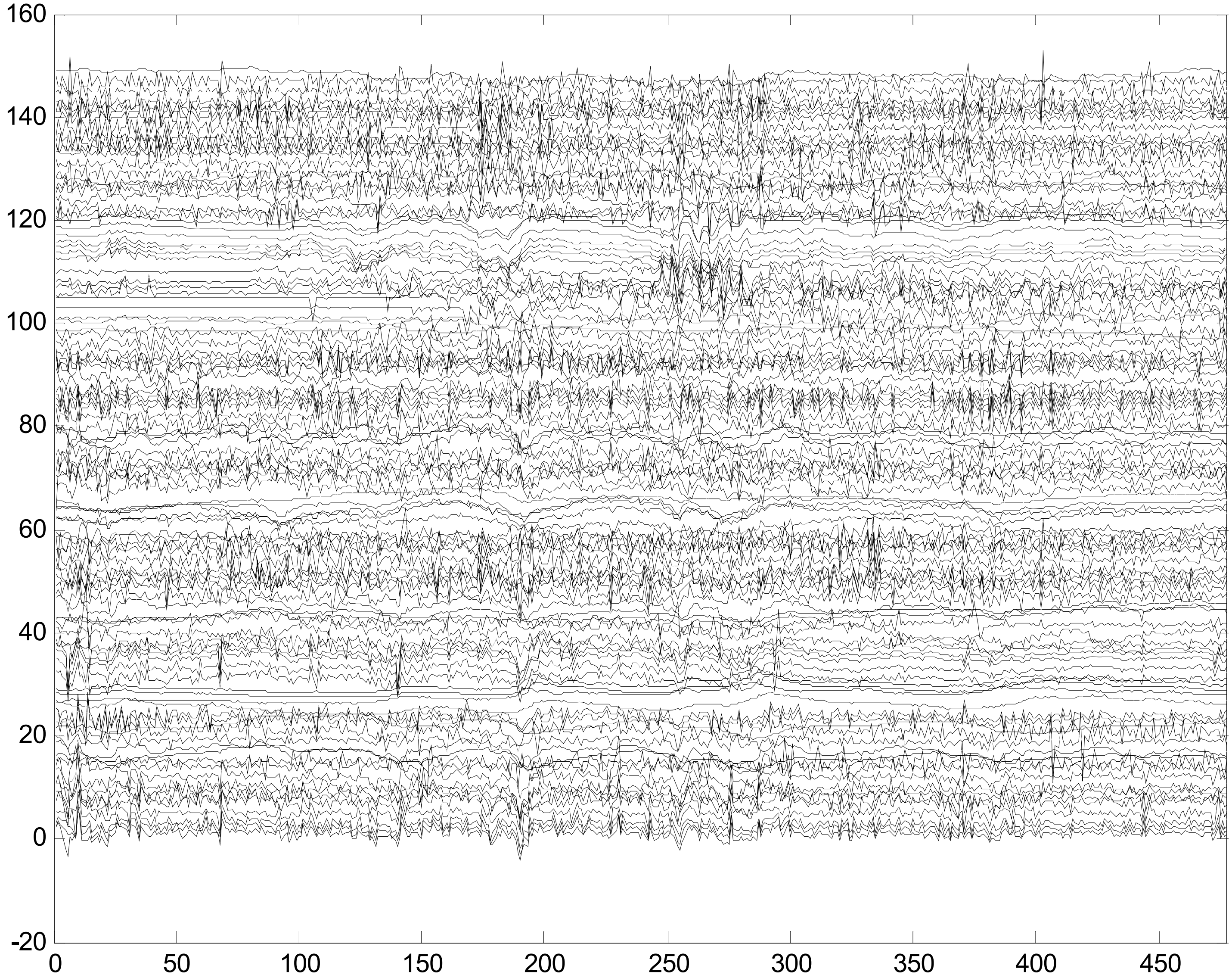
The principal components estimator of large factor models

Alexei Onatski
Columbia University

US macroeconomic data, 2-nd half of 20th century



Large factor models in economics

$$\underset{n\times T}{X} = \underset{n\times r}{L}\ \underset{r\times T}{F'} + \underset{n\times T}{e}$$

- Applications: arbitrage pricing theory, factor augmented VARs, diffusion index forecast models

- Factor domination idea:

$$\max eval\left(\frac{ee'}{T}\right) < u, \quad \min eval\left(L'L\frac{F'F}{T}\right) \to \infty$$

- PC estimator: $\hat{F} = \sqrt{T}\, evec\left(X'X/T\right),\ \hat{L} = X\hat{F}/T$

My assumptions

Assumption 1: Factors are either deterministic or stationary, $E(F'F/T)=I,\quad L'L$ is diagonal

Assumption 2 : Idiosyncratic terms are i.i.d. $N(0,\sigma^2)$

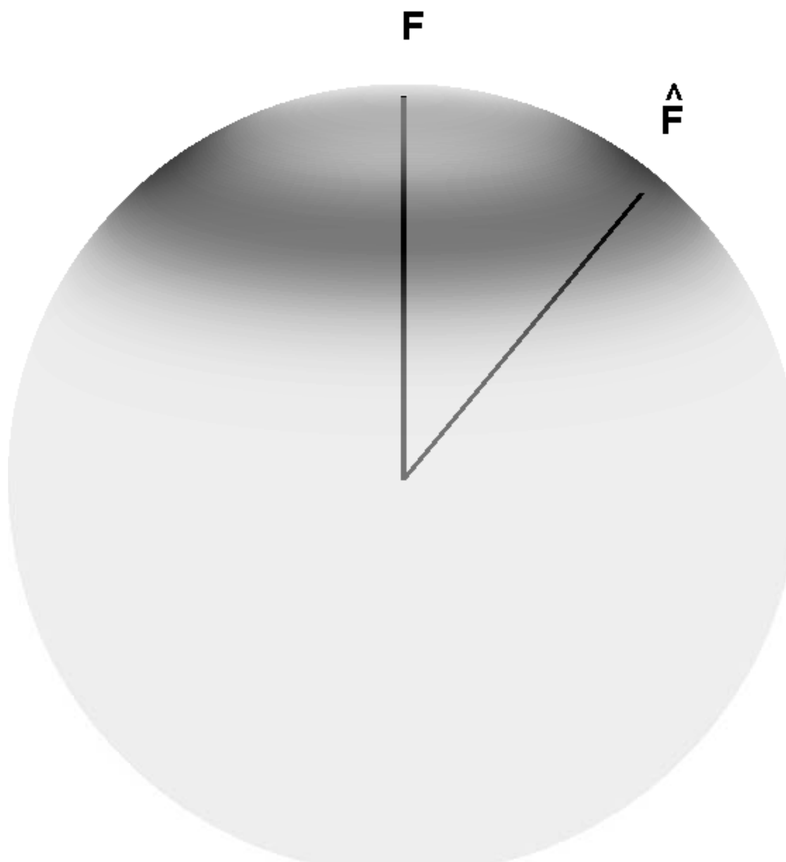
Assumption 3 : $L'L = diag\left(d_1,...,d_k\right)+o(n^{-1/2}),$

$$n/T = c + o(n^{-1/2})$$

$$\sqrt{T}\left(F'F/T - I\right)\rightarrow \Phi$$

Theorem 1: $\hat{F} = F \cdot Q + F^{\perp}, \quad Q = Q^{(1)} + \frac{1}{\sqrt{T}} Q^{(2)},$

$Q^{(1)} < I, \quad Q^{(2)} \to N(0, \Omega)$



Workings of the PC estimator

Let $F = \left(\sqrt{T}, 0, 0, \ldots, 0\right)'$ and $L = (\sqrt{d}, 0, 0, \ldots, 0)'$

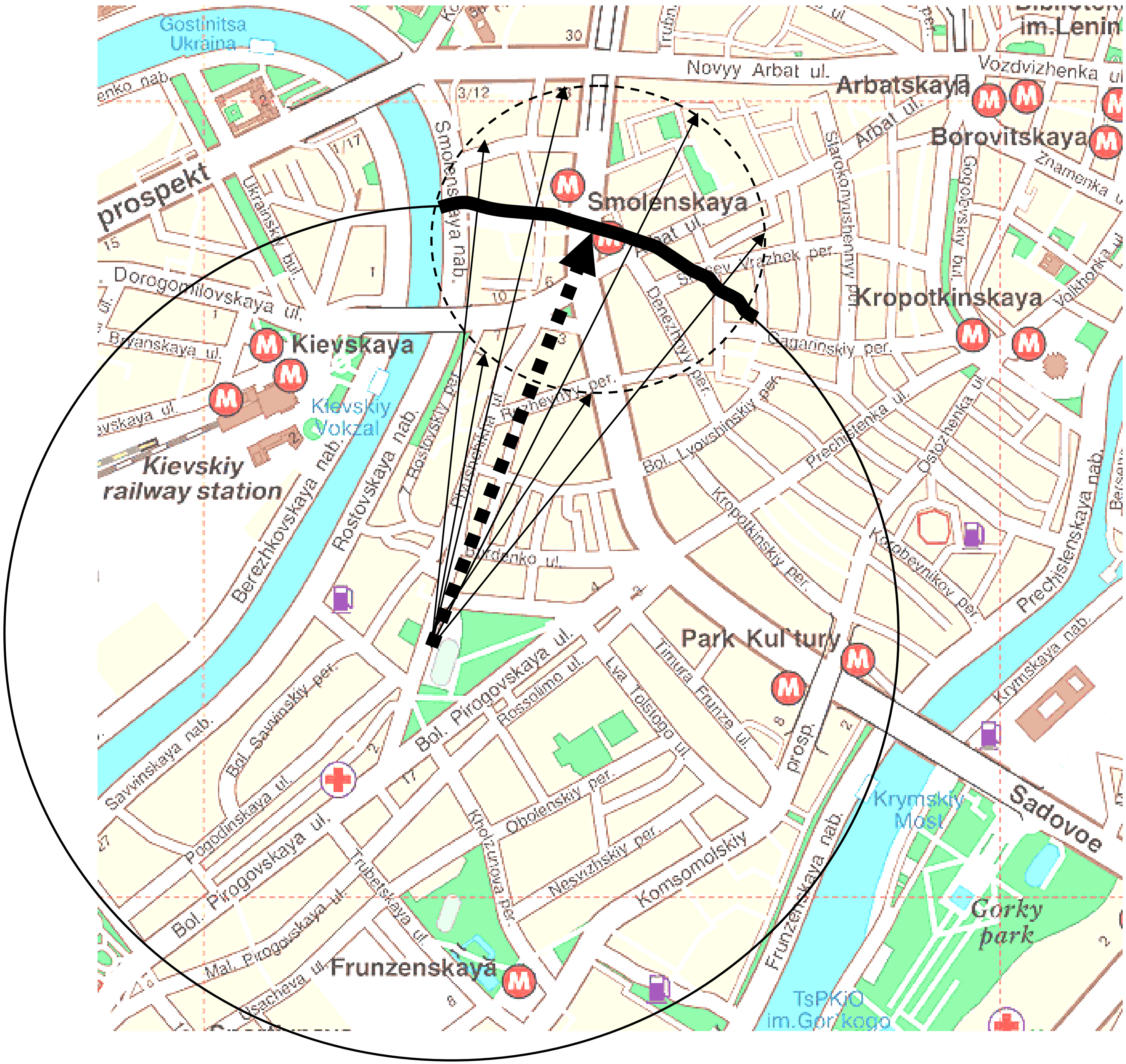
Then $\frac{X'X}{T} = d\,\frac{\left(F+\varepsilon_1\right)\left(F+\varepsilon_1\right)'}{T} + \frac{\varepsilon_{-1}'\varepsilon_{-1}}{T}$

The PC direction is u such that

$$d \cdot u' \frac{\left(F+\varepsilon_1\right)\left(F+\varepsilon_1\right)'}{T} u + u' \frac{\varepsilon_{-1}'\varepsilon_{-1}}{T} u \to \max$$

The first term is maximized when $u \uparrow\uparrow F + \varepsilon_1$

The second, when $u \uparrow\uparrow \max evec\left(\frac{\varepsilon_{-1}'\varepsilon_{-1}}{T}\right)$



Gostinitsa
Ukraina
Novyy Arbat ul.
Arbatskaya
Vozdvizhenka ul.
im.Lenin
Borovitskaya
Smolenskaya
Arbat ul.
prospekt
Dorogomilovskaya ul.
Kievskaya
Kievskiy
Vokzal
Kievskiy
railway station
Kropotkinskaya
Gagarinskiy per.
Smolenskaya nab.
Starokonyushennyy per.
Gogolevskiy bul.
Prechistenka ul.
Ostozhenka ul.
Kropotkinskiy per.
Rostovskaya nab.
Berezhkovskaya nab.
Burdenko ul.
Park Kul`tury
Bol. Pirogovskaya ul.
Savvinskaya nab.
Timura Frunze ul.
Lva Tolstogo ul.
Krymskiy
Most
Sadovoe
Prechistenskaya nab.
Krymskaya nab.
Komsomolskiy
Frunzenskaya nab.
Frunzenskaya
Gorky
park
TsPKiO
im.Gor`kogo
Obolenskiy per.
Nesvizhskiy per.
Mal. Pirogovskaya ul.

Poincaré's observation

- A random direction in a high-dimensional space is nearly orthogonal to any fixed direction with high probability

- More precisely, let S^n be the Euclidean sphere with rotation-invariant measure μ. Fix x_0 on S^{n+1}. Then:

$$\mu\left\{x \in S^{n+1} : \left|x' x_0\right| > \varepsilon\right\} \leq \sqrt{\frac{\pi}{2}} \exp\left(-\varepsilon^2 n / 2\right)$$

The good, the bad, and the ugly

$$\sqrt{T}\left(\hat{L}_{ij} - \rho L_{ij}\right) \to N(0,\Gamma), \quad d_j > \sigma^2\sqrt{c}$$

- $\Gamma = \frac{\sigma^2}{d_j}$
- $\Gamma = \sum_{\substack{s=1 \\ s\neq j}}^{k} L_{is}^2 \frac{\left(d_j + \sigma^2\right)\left(d_s + \sigma^2\right)}{\left(d_j - d_s\right)^2} + \left(1 - \sum_{s=1}^{k} L_{is}^2\right)\frac{\sigma^2\left(d_j + \sigma^2\right)}{d_j^2}$
- $\Gamma = \sum_{\substack{s=1 \\ s\neq j}}^{k} L_{is}^2 \frac{d_j\left(d_j + \sigma^2\right)\left(d_s + \sigma^2\right)}{\left(d_j + c\sigma^2\right)\left(d_j - d_s\right)^2} + \left(1 - \sum_{s=1}^{k} L_{is}^2\right)\frac{\sigma^2\left(d_j + \sigma^2\right)}{d_j\left(d_j + c\sigma^2\right)}$

$$+ L_{ij}^2 \frac{c\sigma^4 d_j\left(d_j + \sigma^2\right)^2}{\left(d_j + c\sigma^2\right)\left(d_j^2 - c\sigma^4\right)^2}\left(1 + c\left(\frac{d_j + \sigma^2}{d_j + c\sigma^2}\right)^2\right)$$

Conclusion

- We find explicit formulae for the amount of inconsistency and the asymptotic distribution of the PC estimator of large spherical factor models
- More work is needed to generalize results to correlated idiosyncratic terms and to dynamic factor models