Power Enhancement Principle in High Dimensional Cross-Sectional Tests

Yuan Liao

University of Maryland

with Jianqing Fan and Jiawei Yao

September 24, 2014

High dimensional sparse alternatives

- Test

$$H_0 : \theta = 0, \quad \dim(\theta) = N \gg T$$

- sparse alternatives: θ is a sparse vector.

Example I

$$y_{it} = \theta_i + \mathbf{b}_i'\mathbf{f}_t + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T,$$

Mean-variance efficiency:

$$H_0 : \theta_1 = ... = \theta_N = 0$$

Numerical evidence:

- Estimate Fama-French-three-factor regression, based on rolling windows from previous $T = 60$ months, Monthly returns of S&P 500 from Jan. 1980 to Dec. 2012.
- On average, $N = 618$, and 7.3 stocks are "significant"
- Market ineff. is mainly due to a few stocks with extra returns.

Example II

$$y_{it} = \alpha + \mathbf{x}_{it}'\beta + u_{it}, \quad i = 1, ..., n, \quad t = 1, ..., T,$$

Cross-sectional independence:

$$H_0 : \mathrm{cov}(u_{it}, u_{jt}) = 0, \quad i \neq j$$

θ : vector of cross-sectional correlations.

Sparse alternatives:

Often, the cross-sectional correlations are weak, yielding sparse covariance matrices

$$\Sigma_u = (\mathrm{cov}(u_{it}, u_{jt}))$$

Standardized Wald test

$$\frac{a_T\widehat{\theta}'\mathbf{V}\widehat{\theta} - N}{\sqrt{2N}} \to^d \mathcal{N}(0,1)$$

Two main challenges

- Estimating $\mathbf{V}$ is challenging when $N > T$.
- More fundamentally, has low power when $\|\theta\|$ is not large (sparse alternatives).

Low power of Wald test

$$a_T \widehat{\theta}' \mathbf{V} \widehat{\theta} \sim \chi^2_N(\|\theta\|_2^2)$$

Theorem

When $T = o(\sqrt{N})$, the Wald test has low power if

$$\Theta_a \subset \{\theta \in \Theta : \sum_{j=1}^{N} 1\{\theta_j \neq 0\} = o(\sqrt{N}/T)\}.$$

The asymptotic power $=$ size.

Power enhancement principle

$$J = J_0 + J_1, \quad J_0 \geq 0$$

- J_1 is a test with correct size, but small power under sparse alternative.
- $P(J_0 = 0 | H_0) \to 1$.
- J_0 is stochastically unbounded under sparse alternative.

■Hence power is enhanced without sacrificing size.

Construction of J_0

Screening set

$$\widehat{S} = \{j : |\widehat{\theta}_j| / \widehat{\mathrm{var}}^{1/2}(\widehat{\theta}_j) > \delta_{N,T}, j = 1, ..., N\}$$

$$J_0 = \sqrt{N} \sum_{j \in \widehat{S}} \widehat{\theta}_j^2 \widehat{\mathrm{var}}^{-1}(\widehat{\theta}_j),$$

- $\delta_{N,T}$ dominates the uniform estimation error:

$$P(\max_{j \leq N} |\widehat{\theta}_j - \theta_j| / \widehat{\mathrm{var}}^{1/2}(\widehat{\theta}_j) < \delta_{N,T}/2 | H_0 \cup H_a) \to 1. \quad (1)$$

- Under H_0, $\max_j |\widehat{\theta}_j| < \widehat{\mathrm{var}}^{1/2}(\widehat{\theta}_j) \delta_{N,T}/2$

$$P(J_0 = 0 | H_0) \to 1.$$

Power enhancement principle

Theorem Suppose $J_1 \to^d F$ under H_0 and has power against $H_a : \theta \in \Omega$, then

- $J_0 + J_1 \to^d F$ under H_0
- $J_0 + J_1$ has power against

$$H_a : \Omega \cup \{\theta \in \mathbb{R}^N : \max_{j \le N} |\theta_j| > C\delta_T\}$$

The power enhancement achieved uniformly over θ

If $J_1 =$ sd. Wald, $\Omega = \{\|\theta\|^2 \ge C\delta_T^2 N/T\}$.

Selection of significant components

$\widehat{S}$ mimics

$$S = \left\{ j : \frac{|\theta_j|}{\sigma_j} > 2\delta_T, j = 1, ..., N \right\},$$

$$P(\widehat{S} = S) \to 1.$$

Testing in factor models

$$y_{it} = \theta_i + \mathbf{b}_i'\mathbf{f}_t + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T,$$

$$H_0 : \theta_1 = ... = \theta_N = 0$$

- Construction of J_1: Pesaran and Yamagata 12:

$$\frac{\sqrt{T} a_f \widehat{\theta}' \mathbf{V} \widehat{\theta} - N}{\sqrt{2N}} | H_0 \rightarrow^d \mathcal{N}(0, 1)$$

$\mathbf{V} = \Sigma_u^{-1}$ under homoskedasticity.

Sparse estimation of Σ_u (Fan et al. 2011):

- Covariance matrix is conditionally sparse (given factors).

$$\mathbf{S}_u = (\frac{1}{T}\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt})$$

$$\widehat{\Sigma}_u = (S_{u,ij} I(S_{u,ij} > C\sqrt{\frac{\log N}{T}}))$$

- Finite sample positive definite.

$$J_1 = \frac{\sqrt{T} a_f \widehat{\theta}' \widehat{\Sigma}_u^{-1} \widehat{\theta} - N}{\sqrt{2Nc}} | H_0 \to^d \mathcal{N}(0,1)$$

A technical challenge

We need

$$\frac{T\widehat{\theta}'(\widehat{\Sigma}_u^{-1} - \Sigma_u^{-1})\widehat{\theta}}{\sqrt{N}} = o_p(1)$$

- Bound:

$$\frac{T}{\sqrt{2N}}\|\widehat{\theta}\|^2\|\Sigma_u^{-1} - \widehat{\Sigma}_u^{-1}\| = \frac{T}{\sqrt{N}} \times \frac{1}{T} \times N \times \frac{1}{\sqrt{T}} = \sqrt{\frac{N}{T}}$$

- Instead, need to consider a weighted convergence directly

$$\widehat{\theta}'(\Sigma_u^{-1} - \widehat{\Sigma}_u^{-1})\widehat{\theta}$$

Final test statistic:

$$J = \underbrace{\sqrt{N} \sum_{j \in \widehat{S}} \widehat{\theta}_j^2 \widehat{\mathrm{var}}^{-1}(\widehat{\theta}_j)}_{J_0} + \underbrace{\frac{cT\widehat{\theta}'\widehat{\Sigma}_u^{-1}\widehat{\theta} - N}{\sqrt{2N}}}_{J_1}$$

Fixed Effect Panel Data

$$y_{it} = \alpha + x_{it}'\beta + \mu_i + u_{it}, \quad H_0 : \Sigma_{u,ij} = 0, i \neq j$$

(Breusch Pagan, Pesaran et al. 08, Baltagi et al. 12)

- Strong correlations: **factor structure**, **cross-sec. AR.**
- Weak correlations: **sparse alternative**, **block diagonal**
- $\sum_{i \neq j} \widehat{\text{cov}}(u_{it}, u_{jt})^2$ has low power under weak correlations.
- Power enhancement:

$$J_0 = \sum_{ij} \hat{\rho}_{ij}^2 1(|\hat{\rho}_{ij}| > \delta_{N,T})$$

Numerical Studies

■Three-factor-model calibrated by the daily returns of top 100 constituents on S&P 500

■Σ_u is generated as a block-diagonal covariance.

■Two alternatives:

Sparse alternative:

$$\alpha_i = \begin{cases} 0.3, & i \le \frac{N}{T} \\ 0, & i > \frac{N}{T} \end{cases}$$

Weak signal:

$$\alpha_i = \begin{cases} \sqrt{\frac{\log N}{T}}, & i \le N^{0.4} \\ 0, & i > N^{0.4} \end{cases}.$$

Table: Size and power (%) of tests for simulated model

N	H_0			H_a^1			H_a^2		
	J_{wald}	PE	$\widehat{S}=\emptyset$	J_{wald}	PE	$\widehat{S}=\emptyset$	J_{wald}	PE	$\widehat{S}=\emptyset$
$T=300$									
500	5.2	5.4	99.8	48.0	97.6	2.6	69.0	76.4	64.6
800	5.4	6.2	99.2	60.0	99.0	1.2	69.2	76.2	62.2
1000	4.0	4.6	99.0	54.6	98.4	2.6	75.8	82.6	63.2
1200	5.0	5.4	99.6	64.2	99.2	0.8	74.2	81.0	63.6
$T=500$									
500	5.8	6.0	99.4	33.8	99.2	0.8	73.4	77.2	77.8
800	4.8	5.0	99.8	67.4	100.0	0.0	72.4	76.4	75.0
1000	5.0	5.2	99.8	65.0	100.0	0.2	76.8	80.4	74.0
1200	5.2	5.2	100.0	58.0	100.0	0.2	74.2	78.4	77.0

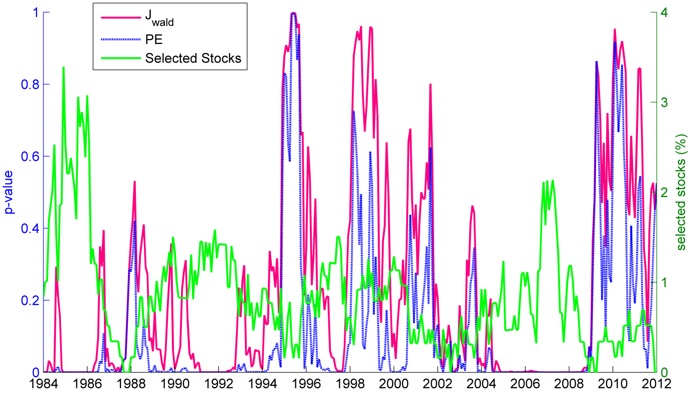
Empirical study

- Monthly returns of S&P 500 constituents from Jan. 1980 to Dec. 2012.
- Estimate Fama-French-three-factor regression
- Estimate and test based on rolling windows from previous $T = 60$ months
- Screening set:
  - On average, $N = 618$, and 7.3 stocks are selected by $\widehat{S}$
  - 81% of selected alphas are positive
  - Market inefficiency is mainly due to a few stocks with extra returns.

Table: Variable descriptive statistics for the FF-3 model

Variables	Mean	Std dev.	Median	Min	Max
N_τ	617.70	26.31	621	574	665
$\lvert\widehat{S}\rvert_0$	5.49	5.48	4	0	37
$\overline{\widehat{\alpha}}_i^{\tau}(\%)$	0.3729	0.1990	0.3338	-0.1735	0.9763
$\overline{\widehat{\alpha}}_{i\in\widehat{S}}^{\tau}(\%)$	3.3980	1.7210	3.7181	-6.2903	8.1299
p-value of J_{wald}	0.1861	0.2947	0.0150	0	0.9926
p-value of PEM	0.1256	0.2602	0.0003	0	0.9836

Figure: p-values and number of selected stocks



Summary

Screening test enhances the power

- Enhances power under sparse alternatives, maintaining correct size

High dimensional Wald test:

- Estimate large error covariance under *conditional sparsity*.
- Effect of estimation is negligible, but technically involved

Empirical study:

- Market inefficiency is mainly due to a few stocks with extra returns.