Measurement of Factor Strength: Theory and Practice

Natalia Bailey[1] George Kapetanios[2] M. Hashem Pesaran[3]

[1]Monash University, Melbourne [2]King's College, London

[3]University of Southern California, and Trinity College, Cambridge

ASSA Virtual Conference

January 3-5, 2021

Introduction: factor models in finance and macroeconomics

- Factor models are used increasingly in empirical finance and macroeconomics to capture common shocks and to price risk factors.
- New theoretical advances have been made to deal with high dimensional factor models (Bai and Ng and others).
- In many applications common factors are latent and are proxied by Principal Components, or some dynamic variants thereof.
- Typically it is assumed that these factors are **strong**, in the sense that they impact all units in the panel.
- There is the additional difficulty of how to **interpret** the latent factors.

Introduction: Factor strengths in networks

- One important example is the role of dominant units in production or financial networks, and how to identify and measure their degree of dominance.
- When interconnections are known (e.g. via input-output tables) outdegrees of networks ($\mathbf{W}$) can be used to detect dominant units that act as strong factors.
  - See, for example, Acemoglu et al. (2012), Pesaran and Yang (2018), with applications to OECD countries by Dungey and Volkov (2018).
- However, in many applications such network information is not available.
  - For example, Kapetanios, Pesaran and Reese (2019) propose a thresholding method to detect pervasive units (if any) in large panel data sets.

Introduction: Measures of cross-sectional dependence

- The issue of factor strength is also related to the analysis of cross-sectional dependence (CSD) which has received attention over the past decade.
- Bailey et al (2016; BKP) propose an estimator of the degree or exponent of CSD which they denote by α.
- Bailey et al (2019) extend the analysis of CSD by focussing on the proportion of statistically significant pair-wise correlations, and show that the estimation method is also applicable to the residuals obtained from panel data regressions.
- The **present paper** addresses the more basic issue of how to measure and estimate factor strength, distinguishing between cases where the factors are observed or latent.

This paper: Measurement of factor strength

- Consider the single factor model $(n, T \to \infty)$

$$x_{it} = c_i + \gamma_i f_t + u_{it},\ i = 1, 2, \ldots, n;\ t = 1, 2, \ldots, T.$$

f_t: known factor, c_i: unit-specific effect, $u_{it} \sim IID(0, \sigma_i^2)$: idiosyncratic error, and γ_i: factor loading for unit i.

- This paper relates the estimation of factor strength to the degree to which f_t has pervasive effects on all the units in the panel.

- We focus on the rate of increase of $\sum_{i=1}^{n} \gamma_i^2$ with n, which we denote by α:
  - $\alpha = 1$: factor is strong; standard (latent) factor model - Bai and Ng (2002), Bai (2003).
  - $0 \leq \alpha < 0.5$: factor is weak - Onatski (2012).
  - $\alpha \in [0.5, 1)$: factor is semi-strong - see also Freyaldenhoven (2019), Uenatsu et al (2019).

Relevance of factor strength in practice

- In most empirical applications, the value of α is unknown:
  - Incorrectly setting $\alpha = 1$ can result in misleading inference.
  - Factors with $\alpha < 0.5$ can be absorbed into the error term and are of little consequence.
  - Values of α that are of interest and of consequence are in the range $\alpha \in (0.5, 1]$.
- Pesaran and Smith (2019) show that factor strengths play a crucial role in the identification and estimation of risk premia in arbitrage asset pricing models.
- Notions of strong and weak factors are used in other recent financial studies; e.g. Lettau and Pelger (2018) and Anatolyev and Mikusheva (2019).
- The strength of macroeconomic shocks is also of special interest, as its value has important bearing on forecasting and policy analysis.

Proposed factor strength estimator

- This paper proposes an estimator of factor strength and establishes its consistency and asymptotic (non-standard) distribution when $\alpha > 1/2$.
- The proposed estimator is based on the number of statistically significant factor loadings, taking account of the multiple tests being carried out.
- The cases where the factors are observed and unobserved are studied seperately.
- The proposed estimator of α has satisfactory small sample properties, especially when α is close to unity, even for moderate sample sizes. We establish an ultra-consistency property when $\alpha = 1$.
- The relevance of the proposed estimator is shown by means of two empirical applications, using well known datasets in finance and macroeconomics.

Multi-(observed) factor model

- Consider T observations on n cross section units: $\{x_{it}, i = 1, 2, \ldots, n, t = 1, 2, \ldots, T\}$ which we model using a multi-(observed) factor model:

$$x_{it} = c_i + \gamma_i' \mathbf{f}_t + u_{it}. \tag{1}$$

$$\text{for } i = 1, 2, \ldots, n \text{ and } t = 1, 2, \ldots, T$$

  - c_i: unit-specific effect
  - $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$: m-dimensional observed vector with $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{im})'$
  - $u_{it} \sim IID(0, \sigma_i^2)$: idiosyncratic (possibly) non-Gaussian errors

- For some unknown ordering of units over i, the loadings γ_i are generated as

$$|\gamma_{ij}| > 0 \text{ a.s. for } i = 1, 2, \ldots, [n^{\alpha_{j0}}],$$
$$|\gamma_{ij}| = 0 \text{ a.s. for } i = [n^{\alpha_{j0}}] + 1, [n^{\alpha_{j0}}] + 2, \ldots, n.$$

  where $1/2 < \alpha_{j0} \leq 1$, for $j = 1, 2, \ldots, m$ (m is fixed).

- The exponents α_{j0} measures the degree of pervasiveness or strength of the j^{th} factor.

Estimation strategy

- Our proposed estimator of α_j is based on the significance of the individual estimates of the factor loadings, γ_{ij}.
- For a given unit i, consider the least squares regression of $\{x_{it}\}_{t=1}^{T}$ on an intercept and $\mathbf{f}_t$. $\hat{c}_{iT}$ and $\hat{\gamma}_{iT}$ are the OLS estimates of this regression.
- Denote by $t_{ijT} = \hat{\gamma}_{ijT} / \text{s.e.}\left(\hat{\gamma}_{ijT}\right)$, the t-statistic corresponding to γ_{ij}:

$$t_{ijT} = \frac{\left(\mathbf{f}'_{j\circ}\mathbf{M}_{F_{-j}}\mathbf{f}_{j\circ}\right)^{-1/2}\left(\mathbf{f}'_{j\circ}\mathbf{M}_{F_{-j}}\mathbf{x}_i\right)}{\hat{\sigma}_{iT}}, \; j = 1, 2, \ldots, m; \; i = 1, 2, \ldots, n,$$

where:

- $\mathbf{f}_{j\circ} = (f_{j1}, f_{j2}, \ldots, f_{jT})'$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$,
- $\mathbf{M}_{F_{-j}} = \mathbf{I} - \mathbf{F}_{-j}\left(\mathbf{F}'_{-j}\mathbf{F}_{-j}\right)^{-1}\mathbf{F}'_{-j}$, $\mathbf{F}_{-j} = (\mathbf{f}_{1\circ}, \ldots, \mathbf{f}_{j-1\circ}, \mathbf{f}_{j+1\circ}, \ldots, \mathbf{f}_{m\circ})'$,
- $\hat{\sigma}^2_{iT} = T^{-1}\sum_{t=1}^{T}\hat{u}^2_{it}$, where $\hat{u}_{it} = x_{it} - \hat{c}_{iT} - \hat{\gamma}'_{iT}\mathbf{f}_t$.

Estimator of α_j: multi-(observed) factor model

- We consider the following estimator of α_j, for $j = 1, 2, \ldots, m$

$$\hat{\alpha}_j = \begin{cases} \tilde{\alpha}_j = 1 + \frac{\ln \hat{\pi}_{nT,j}}{\ln n}, \text{ if } \hat{\pi}_{nT,j} > 0 \\ 0, \text{ if } \hat{\pi}_{nT,j} = 0 \end{cases}, \quad (2)$$

- where $\hat{\pi}_{nT,j}$ denotes the proportion of regressions with statistically significant coefficients γ_{ij}:

$$\hat{\pi}_{nT,j} = n^{-1} \sum_{i=1}^{n} \hat{d}_{ij,nT} = n^{-1} \sum_{i=1}^{n} \mathbf{1}\left[|t_{ijT}| > c_p(n) \right], \quad (3)$$

with $\mathbf{1}(A) = 1$ if $A > 0$, and zero otherwise.

- The critical value function, $c_p(n)$, is given by

$$c_p(n) = \Phi^{-1}\left(1 - \frac{p}{2n^{\delta}}\right), \quad (4)$$

where p is the nominal size of the individual tests, $\delta > 0$ is the critical value exponent and $\Phi^{-1}(\cdot)$ is the inverse cumulative standard normal distribution.

Remarks

1. Clearly $\hat{\alpha}_j \in [0, 1]$ a.s.; also, $\hat{\alpha}_j$ and $\tilde{\alpha}_j$ are asymptotically equivalent since for $\alpha_j > 0$ then $\mathbb{P}(n\ \hat{\pi}_{nT,j} = 0) \to 0$ as $n \to \infty$.

2. It is tempting to argue in favour of using the proportion of non-zero loadings, π_j, instead of the exponent α_j. The two measures are related - $\pi_j = n^{\alpha_j - 1}$:
   - They coincide only when $\alpha_j = 1$.
   - But when $\alpha_j < 1$, π_j becomes smaller and smaller as $n \to \infty$, and eventually tends to 0, for all values of $\alpha_j < 1$.

3. The rate at which π_j tends to zero with n is determined by α_j, and hence α_j is a more discriminating measure of pervasiveness than π_j.

Assumptions

Assumption 1

- The error terms, u_{it}, and demeaned factors $\mathbf{f}_t - E(\mathbf{f}_t)$, are martingale difference processes with respect to $\mathcal{F}_{t-1}^{u_i} = \sigma(u_{i,t-1}, u_{i,t-2}, \ldots)$ and $\mathcal{F}_{t-1}^{f} = \sigma(\mathbf{f}_t, \mathbf{f}_{t-1}, \ldots)$.
- $E\left\{[\mathbf{f}_t - E(\mathbf{f}_t)][\mathbf{f}_t - E(\mathbf{f}_t)]'\right\} = \mathbf{\Sigma}$, some positive definite matrix.
- u_{it} are independent over i, and of $\mathbf{f}_t$, and have constant variances, $0 < \sigma_i^2 < C < \infty$.

Assumption 2

There exist sufficiently large positive constants C_0, C_1,and $s > 0$ such that

$$\sup_{i,t} \Pr(|x_{it}| > \nu) \leq C_0 \exp(-C_1 \nu^s), \text{ for all } \nu > 0, \tag{5}$$

$$\sup_{j,t} \Pr(|f_{jt}| > \nu) \leq C_0 \exp(-C_1 \nu^s), \text{ for all } \nu > 0. \tag{6}$$

Asymptotic distribution

Theorem 1

Consider model (1) with m observed factors and let Assumptions 1 and 2 hold. Then, when $\alpha_{j0} = 1, j = 1, 2, \ldots, m$

$$(\ln n)\left(\hat{\alpha}_j - 1\right) = O_p\left[n^{-1}\exp\left(-C_2 T\right)\right] + O\left[\exp\left(-C_1 T\right)\right],$$

for some $C_1, C_2 > 0$. Otherwise, for any $\alpha_{j0} < 1, j = 1, 2, \ldots, m,$

$$\psi_n(\alpha_{j0})^{-1/2}(\ln n)\left(\hat{\alpha}_j - \alpha_{j0}\right) \to_d N(0, C) \tag{7}$$

for some $C < 1$, where

$$\psi_n(\alpha_{j0}) = p\left(n - n^{\alpha_{j0}}\right) n^{-\delta - 2\alpha_{j0}}\left(1 - \frac{p}{n^{\delta}}\right). \tag{8}$$

Remarks

1. When $\alpha_{j0} = 1$, the distribution of $\hat{\alpha}_j$ is degenerate at $\alpha_{j0} = 1$, with $\hat{\alpha}_j \to_p 1$ exponentially fast. We refer to this property as ultraconsistency.

2. When $1/2 < \alpha_{j0} < 1$, the convergence rate of $\hat{\alpha}_j$ depends on the choice of δ:
    - for values of α_{j0} close to unity (from below) it is sufficient that $\delta > 0$.
    - for values of α_{j0} close to $1/2$, we need $\delta > 1/2$.
    - In the absence of a priori knowledge of α_{j0}, it is sufficient to set $\delta = 1/2$.
    - In practice, factors with $\alpha_{j0} \in [2/3, 1]$ are likely to be of greater interest, and for their estimation it is sufficient to set $\delta = 1/4$.
    - Our Monte Carlo results show that the estimates of factor strengths are reasonably robust to the choice of $\delta \in [1/4, 1/2]$.

Remarks

1. A test based on $\psi_n(\alpha_{j0})^{-1/2}(\ln n)(\hat{\alpha}_j - \alpha_{j0})$ will be conservative:
   - The rejection probability under the null hypothesis will be bounded from above by the significance level.
   - This is why in general we cannot get an asymptotic approximation for the variance of $\hat{\alpha}_j - \alpha_{j0}$ but only an upper bound.
2. Assumptions 1 and 2 can be relaxed:
   - One can assume some spatial mixing condition for u_{it} over i; this would still allow the central limit theorem underlying (7) to hold.
   - One can assume a suitable moment condition in Assumption 2; then one can derive the variance bound needed to construct a test statistic.
   - Relaxing the martingale difference assumption for $\mathbf{f}_t$ to a mixing assumption is possible - see results in Chudik et al (2018).

Case of unobserved factors

- When the factors are unobserved the factor strengths can be identified, but not the factors.
- This is because latent factors are identified only up to a non-singular $m \times m$ rotation matrix, $\mathbf{Q} = (q_{ij})$.
- Consider the multi-factor model (1) where $\mathbf{f}_t$ is unobserved. Suppose that:
  - $m = 2$ so that $\mathbf{f}_t = (f_{1t}, f_{2t})'$.
  - the factor strengths are given by $\alpha_1, \alpha_2 > 1/2$ and $\alpha_1 > \alpha_2$.

Strengths of Principal Components

- The principal component (PC) estimates of these factors are given by $\widehat{\mathbf{g}}_t = (\hat{g}_{1t}, \hat{g}_{2t})'$, and (as n and $T \to \infty$)

$$f_{1t} = q_{11}\hat{g}_{1t} + q_{12}\hat{g}_{2t} + o_p(1),$$
$$f_{2t} = q_{21}\hat{g}_{1t} + q_{22}\hat{g}_{2t} + o_p(1).$$

- Then the estimates of the loadings, $\widetilde{\gamma}_i = (\tilde{\gamma}_{i1}, \tilde{\gamma}_{i2})'$ associated with these PCs are given by - noting that $\widehat{\mathbf{G}} = (\widehat{\mathbf{g}}_1, \widehat{\mathbf{g}}_2, \ldots, \widehat{\mathbf{g}}_T)'$:

$$\widetilde{\gamma}_i = \left(\widehat{\mathbf{G}}'\mathbf{M}_\tau\widehat{\mathbf{G}}\right)^{-1}\widehat{\mathbf{G}}'\mathbf{M}_\tau\mathbf{x}_i.$$

- Since $\mathbf{Q}$ is non-singular, then $\widehat{\mathbf{G}} \to_p \mathbf{F}\mathbf{Q}^{-1}$ and $\widetilde{\gamma}_i \to_p \mathbf{Q}\gamma_i$.

- The strength of f_{1t} (or f_{2t}) computed using the estimates, $\tilde{\gamma}_{i1}$, $i = 1, 2, \ldots, n$ may not provide consistent estimates of the associated factor strengths.

Strengths of Principal Components

- To see this write $\widetilde{\gamma}_i \rightarrow_p \mathbf{Q}\gamma_i$ in an expanded format as

$$\tilde{\gamma}_{i1} = q_{11}\gamma_{i1} + q_{12}\gamma_{i2} + o_p(1),$$
$$\tilde{\gamma}_{i2} = q_{21}\gamma_{i1} + q_{22}\gamma_{i2} + o_p(1).$$

- Then, squaring both sides and summing over i gives us, in general:

$$\sum_{i=1}^{n} \tilde{\gamma}_{i1}^2 = \ominus(n^{\alpha_1}), \sum_{i=1}^{n} \tilde{\gamma}_{i2}^2 = \ominus(n^{\alpha_1}).$$

- The PCs do not allow us to distinguish/identify the factors by their strength.

- Using covariance eigenvalues does not help resolve this problem.

- This is because the eigenvectors associated with the largest eigenvalues are not uniquely determined.

Strength of strongest (unobserved) factor

- We focus on estimation of the true value of $\alpha = \max(\alpha_{1,}\alpha_2)$, α_0, which, in general, can be identified.
- The exponent α_0 can be estimated using the estimators proposed in Bailey et al (2016, 2019).
- The proposed estimator of this paper can also be used by computing the strength of:
  - the first PC
  - the simple cross section average, namely $\bar{x}_t = n^{-1}\sum_{i=1}^{n} x_{it}$
  - or the weighted cross section average $\bar{x}_{t,\gamma} = \sum_{i=1}^{n} \hat{w}_i x_{it}$, where $\hat{w}_i$ are the slope coefficients of $\bar{x}_t$ in the OLS regression of x_{it} on an intercept and $\bar{x}_t$.

Multi-(unobserved) factor model

- We recast the multi-factor model (1) with unobserved factors as follows:

$$x_{it} = c_i + \gamma_i f_t + v_{it}, \text{ for } i = 1, 2, \dots, n \text{ and } t = 1, 2, \dots, T \tag{9}$$

$$v_{it} = \sum_{j=2}^{m} \gamma_{ij} f_{jt} + u_{it}, \tag{10}$$

- The strongest factor f_t has strength $\alpha_0 = \alpha_{10}$, while the rest of the factors have strengths $\alpha_{20} \geq \alpha_{30} \geq \dots \geq \alpha_{m0} > 1/2$.

- For some unknown ordering of units over i, we assume for the loadings:

$$\begin{aligned} |\gamma_i| &> 0 \text{ a.s. for } i = 1, 2, \dots, [n^{\alpha_0}], \\ |\gamma_i| &= 0 \text{ a.s. for } i = [n^{\alpha_0}] + 1, [n^{\alpha_0}] + 2, \dots, n. \end{aligned} \tag{11}$$

$$\begin{aligned} |\gamma_{ij}| &> 0 \text{ a.s. for } i = 1, 2, \dots, [n^{\alpha_{j0}}], \ j = 2, \dots, m \\ |\gamma_{ij}| &= 0 \text{ a.s. for } i = [n^{\alpha_{j0}}] + 1, [n^{\alpha_{j0}}] + 2, \dots, n, \ j = 2, \dots, m. \end{aligned} \tag{12}$$

Estimation of α_0: multi-(unobserved) factor model

- For each i, consider the least squares regression of $\{x_{it}\}_{t=1}^{T}$ on an intercept and the cross section average of x_{it}, $\bar{x}_t$.

- $\alpha_0 = \max_j(\alpha_{j0})$ is estimated using the t-statistic corresponding to γ_i, with $\mathbf{f}$ replaced by $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T)'$:

$$\bar{t}_{iT} = \frac{\left(\bar{\mathbf{x}}'\mathbf{M}_\tau\bar{\mathbf{x}}\right)^{-1/2}\left(\bar{\mathbf{x}}'\mathbf{M}_\tau\mathbf{x}_i\right)}{\hat{\sigma}_{iT}},$$

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})',\ \hat{\sigma}^2_{iT} = T^{-1}\mathbf{x}_i'\mathbf{M}_{\bar{H}}\mathbf{x}_i,\ \mathbf{M}_{\bar{H}} = \mathbf{I}_T - \overline{\mathbf{H}}\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1}\overline{\mathbf{H}}',\ \overline{\mathbf{H}} = (\boldsymbol{\tau}_T, \bar{\mathbf{x}}).$$

- Then, we have

$$\hat{\alpha} = \begin{cases} 1 + \frac{\ln \bar{\pi}_{nT}}{\ln n},\ \text{if } \bar{\pi}_{nT} > 0 \\ 0,\ \text{if } \bar{\pi}_{nT} = 0 \end{cases},$$

where $\bar{\pi}_{nT} = n^{-1}\sum_{i=1}^{n} \mathbf{1}\left[|\bar{t}_{iT}| > c_p(n)\right]$, and $c_p(n)$ is given by (4).

Asymptotic distribution

Theorem 2

Consider model (9)-(10) with factor loadings given by (11)-(12), where $\mathbf{f}_t$ is unobserved. Let Assumptions 1 and 2 hold and denote by α_0 the true value of α. Then, as long as $\sqrt{T}n^{(\alpha_{20}-\alpha_0)} \to 0$, for any $\alpha_0 < 1$,

$$\psi_n(\alpha_0)^{-1/2} (\ln n) (\hat{\alpha} - \alpha_0) \to_d N(0, C)$$

for some $C < 1$, where α_{20} denotes the strength of the second strongest factor, and

$$\psi_n(\alpha_0) = p \left(n - n^{\alpha_0}\right) n^{-\delta-2\alpha_0} \left(1 - \frac{p}{n^{\delta}}\right).$$

Estimation of α_j: multi-(unobserved) factor model

- A possible way to provide some information on $\alpha_{j0}, j > 1$, may be based on a sequential application of weighted cross section averages.
- In particular, one can obtain residuals from the least squares regression of $\{x_{it}\}_{t=1}^{T}$ on an intercept and the cross section average of x_{it}, $\bar{x}_t$.
- Then, weighted cross section averages of these can be constructed, and the t-statistics of the relevant loadings can be used, in a similar way to that discussed above, to construct estimators for α_{20}.
- The procedure can continue sequentially, via the construction of further sets of residuals, for $\alpha_{j0}, j > 2$.

Asymptotic distribution

Theorem 3

Consider model (9)-(10) with factor loadings given by (11)-(12), where $\mathbf{f}_t$ is unobserved. Let Assumptions 1 and 2 hold and and denote by α_{j0} the true value of α_j. Then, as long as $\sqrt{T}n^{(\alpha_{j+1,0}-\alpha_{j0})} \to 0$, $j > 1$, for any $0.5 < \alpha_{j+1,0} < \alpha_{j0} < 1$,

$$\psi_n(\alpha_{j0})^{-1/2} (\ln n) (\hat{\alpha}_j - \alpha_{j0}) \to_d N(0, C)$$

for some $C < 1$, and

$$\psi_n(\alpha_{j0}) = p (n - n^{\alpha_{j0}}) n^{-\delta-2\alpha_{j0}} \left(1 - \frac{p}{n^\delta}\right).$$

Small sample properties of $\hat{\alpha}$

- We consider the following two-factor data generating process (DGP):

$$x_{it} = c_i + \gamma_{1i} f_{1t} + \gamma_{2i} f_{2t} + u_{it}, \tag{13}$$

for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, with factors, $\mathbf{f}_t = (f_{1t}, f_{2t})'$.

- Unit specific effects: $c_i \sim IIDN\,(0, 1)$.
- Factors:
  - are multi-variate normal, $\mathbf{f}_t \sim N\,(\mathbf{0}, \mathbf{\Sigma}_f)$, with variances $\sigma^2_{f_1} = \sigma^2_{f2} = 1$, and correlation $\rho_{12} = corr\,(f_{1t}, f_{2t}) = 0.0, 0.3$.
  - follow AR(1) processes with autocorrelation coefficients given by $\rho_{f_1} = \rho_{f_2} = 0.5$.
- Innovations, u_{it} can be:
  - Gaussian: $u_{it} \sim IIDN(0, \sigma_i^2)$, and $\sigma_i^2 \sim IID(1 + \chi^2_{2,i})/3$.
  - non-Gaussian: $u_{it} = \frac{\sigma_i}{2}\left(\chi^2_{2,it} - 2\right)$, and $\sigma_i^2 \sim IID(1 + \chi^2_{2,i})/3$.

Experiments

EXP 1A: Single observed factor in (13) ($\gamma_{i2} = 0$, for all i) and Gaussian errors.

EXP 1B: Single observed factor in (13) ($\gamma_{i2} = 0$, for all i) and non-Gaussian errors.

EXP 2A: Two observed correlated factors in (13) ($\rho_{12} = 0.3$) and Gaussian errors.

EXP 2B: Two observed correlated factors in (13) ($\rho_{12} = 0.3$) and non-Gaussian errors.

EXP 3A: Single unobserved factor in (13) ($\gamma_{i2} = 0$, for all i) and non-Gaussian errors. $\alpha_0 = \alpha_{10}$ is computed using the simple CSA, $\bar{x}_t$.

EXP 3B: Two unobserved factors in (13) ($\rho_{12} = 0.3$) and non-Gaussian errors. $\alpha_{10} = 0.95, 1.00$, and $\alpha_{20} = 0.51, 0.75, 0.95, 1.00$. $\alpha_0 = \max(\alpha_{10}, \alpha_{20})$ is estimated using the simple CSA, $\bar{x}_t$.

EXP 4: Two factors in (13) ($\rho_{12} = 0.3$) and Gaussian errors. $\alpha_{10} = 1$, and $\alpha_{20} = 0.75, 0.80, \ldots, 0.95, 1.00$. Estimates of α_{10} are computed based on misspecified model: $x_{it} = c_i + \beta_i f_{1t} + e_{it}$.

design stats obs unobs

Table: Bias, RMSE and Size ($\times$100) of estimating different strengths of first factor in the case of experiment 2B (two observed factors - non-Gaussian errors), when $\alpha_{20} = 0.85$

	Bias ($\times$100)				RMSE ($\times$100)				Size ($\times$100)			
$n\backslash$T	120	200	500	1000	120	200	500	1000	120	200	500	1000
	$\alpha_{10} = 0.75$, $\alpha_{20} = 0.85$											
100	1.20	1.13	1.08	1.04	1.58	1.50	1.46	1.41	4.35	2.80	3.25	2.30
200	1.43	1.39	1.30	1.31	1.59	1.54	1.45	1.47	10.00	7.65	7.00	7.55
500	1.30	1.23	1.17	1.14	1.38	1.30	1.24	1.21	13.55	10.25	8.20	7.20
1000	1.27	1.18	1.12	1.11	1.31	1.22	1.16	1.15	17.25	11.05	7.45	7.60
	$\alpha_{10} = 0.80$, $\alpha_{20} = 0.85$											
100	0.71	0.66	0.63	0.61	1.03	1.00	0.96	0.95	17.75	18.80	18.15	19.45
200	0.95	0.93	0.85	0.86	1.09	1.05	0.97	0.99	13.10	11.35	9.20	9.75
500	0.91	0.86	0.82	0.80	0.96	0.92	0.87	0.86	11.80	9.25	5.95	7.40
1000	0.85	0.81	0.76	0.76	0.88	0.83	0.79	0.78	18.10	11.00	8.85	7.80
	$\alpha_{10} = 0.85$, $\alpha_{20} = 0.85$											
100	0.68	0.67	0.64	0.62	0.87	0.86	0.83	0.81	9.70	9.40	7.70	7.35
200	0.61	0.59	0.54	0.54	0.72	0.69	0.65	0.65	5.95	3.90	4.10	3.05
500	0.51	0.50	0.47	0.46	0.56	0.54	0.51	0.51	10.80	7.70	7.35	7.75
1000	0.50	0.47	0.45	0.44	0.52	0.49	0.47	0.46	12.45	8.45	5.45	5.40
	$\alpha_{10} = 0.90$, $\alpha_{20} = 0.85$											
100	0.40	0.40	0.38	0.36	0.56	0.55	0.53	0.51	5.35	3.75	3.55	3.05
200	0.27	0.26	0.23	0.24	0.38	0.36	0.33	0.34	14.95	12.45	13.20	13.35
500	0.28	0.29	0.27	0.26	0.32	0.32	0.30	0.29	9.85	8.35	6.85	6.20
1000	0.28	0.27	0.26	0.25	0.30	0.28	0.27	0.27	12.60	8.25	6.50	6.05
	$\alpha_{10} = 0.95$, $\alpha_{20} = 0.85$											
100	0.06	0.07	0.07	0.06	0.25	0.24	0.24	0.22	6.70	3.45	3.40	2.75
200	0.10	0.11	0.10	0.10	0.18	0.18	0.17	0.17	9.15	4.05	3.85	4.45
500	0.10	0.11	0.11	0.11	0.13	0.14	0.13	0.13	13.35	8.75	8.85	7.40
1000	0.09	0.09	0.09	0.09	0.11	0.11	0.10	0.10	11.50	5.75	5.65	5.05
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.85$											
100	-0.02	0.00	0.00	0.00	0.07	0.02	0.00	0.00	-	-	-	-
200	-0.02	0.00	0.00	0.00	0.05	0.01	0.00	0.00	-	-	-	-
500	-0.02	0.00	0.00	0.00	0.03	0.01	0.00	0.00	-	-	-	-
1000	-0.02	0.00	0.00	0.00	0.03	0.00	0.00	0.00	-	-	-	-

Figure: Empirical power functions associated with testing different strengths of first factor in the case of experiment 2B (two observed factors - non-Gaussian errors), when $\alpha_{20} = 0.85$, $n = 100, 200, 500, 1000$ and $T = 200$

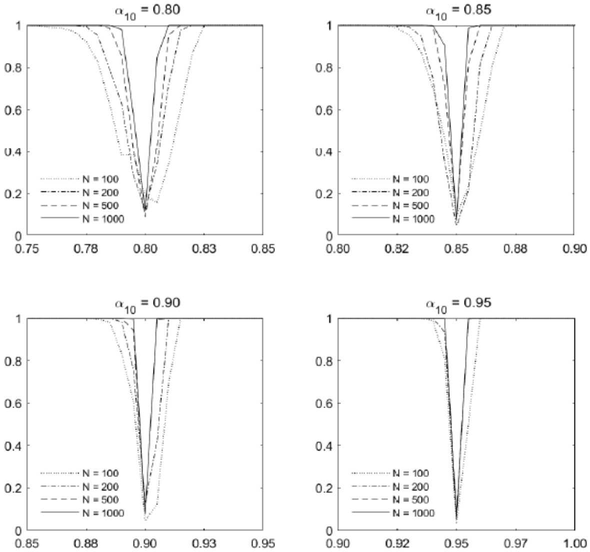


Table: Bias, RMSE and Size ($\times 100$) of estimating the strength of strongest factor in the case of experiment 3A (unobserved single factor - non-Gaussian errors) using CSA

	Bias ($\times 100$)				RMSE ($\times 100$)				Size ($\times 100$)			
n\T	120	200	500	1000	120	200	500	1000	120	200	500	1000
					$\alpha_{10} = 0.75$							
100	2.40	2.79	4.25	6.73	2.84	3.22	4.60	7.02	27.05	37.35	73.75	98.40
200	2.09	2.12	2.60	3.47	2.32	2.35	2.82	3.67	30.20	34.05	51.60	81.60
500	1.62	1.55	1.59	1.81	1.77	1.68	1.71	1.92	30.95	26.65	29.85	44.80
1000	1.46	1.39	1.38	1.41	1.54	1.46	1.44	1.47	31.95	26.00	27.70	28.00
					$\alpha_{10} = 0.80$							
100	1.26	1.47	2.14	3.42	1.61	1.81	2.43	3.66	28.05	32.20	55.35	87.45
200	1.24	1.24	1.42	1.76	1.39	1.40	1.57	1.90	24.75	27.35	35.40	54.80
500	1.03	0.98	0.97	1.04	1.11	1.05	1.03	1.10	21.80	17.75	15.40	21.95
1000	0.92	0.88	0.85	0.86	0.96	0.92	0.89	0.89	27.00	21.00	18.10	16.95
					$\alpha_{10} = 0.85$							
100	0.91	1.00	1.24	1.78	1.11	1.19	1.44	1.96	19.30	24.15	37.10	64.75
200	0.72	0.71	0.76	0.88	0.83	0.82	0.88	0.99	10.30	10.70	12.75	20.35
500	0.57	0.54	0.52	0.54	0.63	0.59	0.56	0.59	15.25	12.35	10.30	11.90
1000	0.52	0.50	0.48	0.47	0.55	0.52	0.50	0.50	15.50	10.15	8.75	7.45
					$\alpha_{10} = 0.90$							
100	0.49	0.51	0.60	0.79	0.63	0.67	0.76	0.95	6.75	8.40	12.50	22.90
200	0.32	0.31	0.32	0.35	0.42	0.40	0.42	0.44	13.75	11.50	13.50	13.40
500	0.31	0.30	0.28	0.29	0.35	0.33	0.31	0.32	12.50	9.95	8.85	8.50
1000	0.29	0.28	0.27	0.27	0.31	0.30	0.28	0.28	14.50	10.30	8.25	6.95
					$\alpha_{10} = 0.95$							
100	0.10	0.12	0.13	0.18	0.26	0.27	0.29	0.34	5.40	5.25	6.55	11.05
200	0.13	0.12	0.12	0.13	0.20	0.19	0.18	0.20	7.85	5.65	5.35	6.50
500	0.12	0.12	0.11	0.11	0.15	0.15	0.14	0.14	12.40	9.10	7.80	8.75
1000	0.10	0.10	0.09	0.09	0.11	0.11	0.11	0.10	8.75	5.35	4.90	4.90
					$\alpha_{10} = 1.00$							
100	-0.01	0.00	0.00	0.00	0.04	0.01	0.00	0.00	-	-	-	-
200	-0.01	0.00	0.00	0.00	0.03	0.01	0.00	0.00	-	-	-	-
500	-0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	-	-	-	-
1000	-0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	-	-	-	-

Figure: Empirical power functions associated with testing different strengths of strongest factor in the case of experiment 3A (unobserved single factor - non-Gaussian errors) using CSA, when $n = 100, 200, 500, 1000$ and $T = 200$

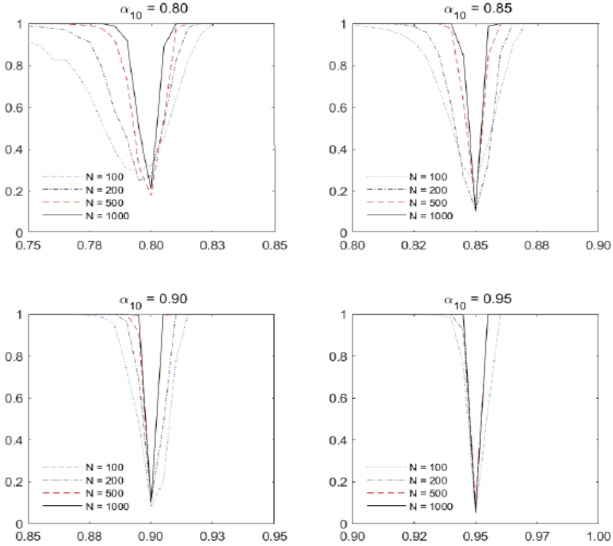


Table: Bias and RMSE ($\times 10,000$) of estimating the strength of strongest factor in the case of experiment 3B (two unobserved factors - non-Gaussian errors) using CSA, when $\alpha_{10} = 1.00$

	Bias ($\times 10,000$)				RMSE ($\times 10,000$)			
$n\backslash T$	120	200	500	1000	120	200	500	1000
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.51$							
100	-0.76	-0.05	0.00	0.00	4.20	1.09	0.00	0.00
200	-0.95	-0.05	0.00	0.00	3.14	0.67	0.00	0.00
500	-1.12	-0.07	0.00	0.00	2.39	0.46	0.00	0.00
1000	-1.25	-0.08	0.00	0.00	2.04	0.37	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.75$							
100	-0.92	-0.04	0.00	0.00	4.58	0.98	0.00	0.00
200	-0.94	-0.03	0.00	0.00	3.11	0.52	0.00	0.00
500	-1.12	-0.06	0.00	0.00	2.39	0.45	0.00	0.00
1000	-1.26	-0.09	0.00	0.00	2.08	0.38	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.95$							
100	-1.44	-0.15	0.00	0.00	5.78	1.83	0.00	0.00
200	-2.05	-0.19	0.00	0.00	5.31	1.37	0.00	0.00
500	-2.08	-0.19	0.00	0.00	3.99	0.82	0.00	0.00
1000	-2.27	-0.23	0.00	0.00	3.99	0.82	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 1.00$							
100	-0.02	0.00	0.00	0.00	0.69	0.00	0.00	0.00
200	-0.01	0.00	0.00	0.00	0.30	0.00	0.00	0.00
500	-0.02	0.00	0.00	0.00	0.30	0.00	0.00	0.00
1000	-0.02	0.00	0.00	0.00	0.19	0.00	0.00	0.00

Table: Bias and RMSE ($\times$100) of estimating the strength of strongest factor in the case of experiment 3B (two unobserved factors - non-Gaussian errors) using CSA, when $\alpha_{10} = 0.95$

	Bias ($\times$100)				RMSE ($\times$100)			
$n\backslash$T	120	200	500	1000	120	200	500	1000
	$\alpha_{10} = 0.95$, $\alpha_{20} = 0.51$							
100	0.18	0.21	0.40	0.57	0.36	0.38	0.55	0.70
200	0.16	0.16	0.22	0.30	0.23	0.24	0.29	0.38
500	0.13	0.13	0.15	0.17	0.16	0.16	0.18	0.20
1000	0.10	0.10	0.10	0.11	0.12	0.12	0.12	0.13
	$\alpha_{10} = 0.95$, $\alpha_{20} = 0.75$							
100	1.27	1.56	1.73	1.79	1.41	1.65	1.80	1.85
200	0.98	1.24	1.52	1.54	1.10	1.31	1.55	1.57
500	0.61	0.86	1.19	1.26	0.72	0.92	1.21	1.27
1000	0.42	0.59	0.95	1.07	0.51	0.67	0.97	1.08
	$\alpha_{10} = 0.95$, $\alpha_{20} = 0.95$							
100	3.98	4.03	4.04	4.05	4.00	4.05	4.06	4.07
200	3.87	3.95	3.95	3.96	3.88	3.96	3.96	3.97
500	3.74	3.82	3.84	3.83	3.74	3.82	3.84	3.83
1000	3.62	3.71	3.72	3.73	3.63	3.72	3.72	3.73
	$\alpha_{10} = 0.95$, $\alpha_{20} = 1.00$							
100	-0.02	0.00	0.00	0.00	0.07	0.02	0.00	0.00
200	-0.02	0.00	0.00	0.00	0.05	0.01	0.00	0.00
500	-0.02	0.00	0.00	0.00	0.04	0.01	0.00	0.00
1000	-0.02	0.00	0.00	0.00	0.04	0.01	0.00	0.00

Empirical applications

- We consider two empirical applications, using well-known datasets in finance and macroeconomics:
- (I) Identification of risk factors in asset pricing models (case of observed factors)
- (II) quantifying the strength of common macroeconomic shocks (case of unobserved factors)

(I) Identifying risk factors in asset pricing models

- CAPM (Sharpe (1964), Lintner (1956)) and its multi-factor APT extension (Ross (1976)) are widely used models in modern empirical finance.
- In turn, the Fama-MacBeth two-pass procedure (Fama and MacBeth (1973)) assesses the linear pricing relationships implied by these models.
- The first stage in the Fama-MacBeth procedure entails choosing the risk factors to be included in the asset pricing model.
- There is an upsurge in the number of factors deemed relevant to asset pricing in the past few years; Harvey and Liu (2019) document over 400 such factors.
- This has led to a rapidly growing area in the finance literature which evaluates the contribution of potential factors to these models.
- Recent contributions allow for the possibility of false discovery when the number of potential factors is large and multiple testing issues arise - Feng et al (2020).

Consistent estimation of risk premia

- We focus on determining the strength of these factors as a means of evaluating whether their risk can be priced correctly.
- Pesaran and Smith (2019) show that the APT theory requires risk factors to be sufficiently strong if their associated risk premium is to be estimated consistently.
- They explain that the risk premium of a factor with strength α can be estimated at the rate of $n^{-\alpha/2}$, where n is the number of individual securities studied.
- $\sqrt{n}$ consistent estimation of the risk premium of a given factor requires that factor to be strong, i.e. $\alpha = 1$.
- Factors with $\alpha < 0.5$ cannot be priced and are absorbed in pricing errors.
- In principle, it should be possible to identify the risk premium of semi-strong factors (factors with $\alpha \in (1/2, 1)$), but a very large n is required.
- In practice, where n is not sufficiently large, at best only factors with strength sufficiently close to unity can be priced.

Data

- We consider *monthly excess returns* of the securities included in the S&P 500 index over the period from September 1989 to December 2017:

$$\tilde{r}_{it} = r_{it} - r_{ft},$$

  where r_{it} is monthly return on security i, inclusive of dividend payments (if any) and r_{ft} is the risk-free rate (one-month US treasury bill rate).

- Returns were compiled on those securities (out of 500 in total) that had at least 120 months of history at the end of each month in the sample.
- On average, we ended up with $n = 442$ securities at the end of each month.
- We consider 146 factors in total: (i) the market factor (measured as the excess market return); (ii) an additional 145 factors considered by Feng et al (2020).
- In order to account for time variations in factor strength, we use rolling samples (340 in total) of 120 months (10 years).

Factor models for individual securities

- First, we consider the original CAPM specification

$$r_{it} - r_{ft} = a_{im} + \beta_{im} \left(r_{mt} - r_{ft} \right) + u_{it,m}, \text{ for } i = 1, 2, \ldots, n_{\tau}, \tag{14}$$

  where n_{τ} are the number of securities in rolling sample $\tau = 1, 2, \ldots, 340$.

- We obtain estimates of the strength of the market factor across the rolling windows, $\hat{\alpha}_{m,\tau}$, $\tau = 1, 2, \ldots, 340$.

- Next, we add the 145 factors to (14), one at a time and run

$$r_{it} - r_{ft} = a_{is} + \beta_{im|s} \left(r_{mt} - r_{ft} \right) + \beta_{is} f_{st} + u_{it,s}, \; i = 1, 2, \ldots, n_{\tau}$$

  for each $s = 1, 2, \ldots, 145$ and $\tau = 1, 2, \ldots, 340$.

- We assess the effect on the market factor strength estimates of adding more factors to (14) by computing rolling $\hat{\alpha}_{m,\tau|s}$, with the s^{th} factor included.

- We quantify the strength of these additional factors by obtaining rolling strength estimates for these factors, $\hat{\alpha}_{s,\tau}$.

Figure: Comparison of the market factor strength estimates obtained from CAPM ($\hat{\alpha}_{m,\tau}$) and the average estimates of its strength when computed using 145 two-factor asset pricing models ($\bar{\hat{\alpha}}_{m,\tau}$), over 10-year rolling windows ($p = 0.10$ and $\delta = 1/4$)

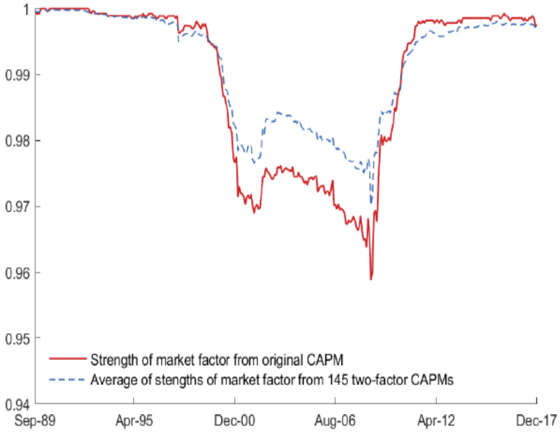


Figure: Percentage of factors (out of 145) whose estimated strength ($\hat{\alpha}_{s,\tau}$), $\tau = 1, 2, \ldots, 340$ exceeds the thresholds of 0.85, 0.90 and 0.95, in each 10-year rolling window

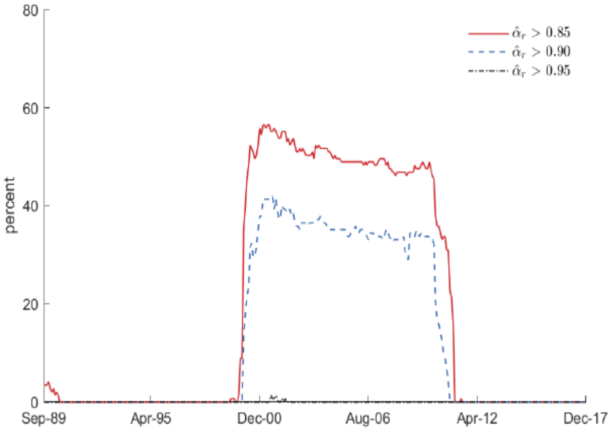
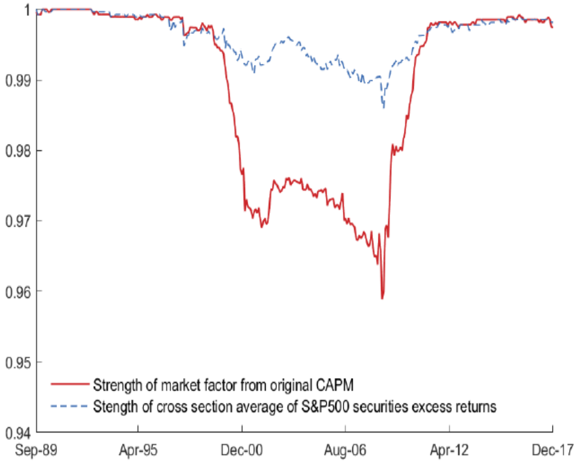


Figure: Comparison of the market factor strength estimates obtained from CAPM ($\hat{\alpha}_{m,\tau}$) and those from using CSA of S&P500 securities' excess returns ($\hat{\alpha}_{csa,\tau}$), over 10-year rolling windows



Commentary: market factor

- The rolling market factor estimates implied by CAPM are $\hat{\alpha}_{m,\tau} \approx 1$, suggesting that the market factor is strong and its risk premium is $\sqrt{n}$ consistent.
- There is some evidence of departure from unity over the period December 1999 to January 2011 which saw a number of sizeable financial events.
- $\hat{\alpha}_{m,\tau}$ recorded its minimum value of 0.958 in August 2008, around the time of the Lehman Brothers collapse.
- The estimates of market factor strength are generally unaffected if we consider the augmented CAPM regressions. It is clear that $\bar{\hat{\alpha}}_{m,\tau}$ closely track $\hat{\alpha}_{m,\tau}$ in general.

Commentary: additional factors

- The 10-year rolling estimation of the strength of the strongest unobserved factor captured by the strength of the CSA of the excess returns, $\hat{\alpha}_{csa,\tau}$, again tracks $\hat{\alpha}_{m,\tau}$ very closely - $corr(\hat{\alpha}_{csa,\tau}, \hat{\alpha}_{m,\tau}) = 0.93$.
- The 10-year rolling estimates of the strength of the remaining 145 factors, $\hat{\alpha}_{s,\tau}$, show considerable time variations, especially during December 1999 to January 2011.
- However, at no point during the full sample do any of these factors become strong.
- Apart from the market factor, there is no other factor whose strength exceeds 0.95 throughout our sample period.

(II) Strength of common macroeconomic shocks

- Similar considerations apply to macroeconomic shocks and their pervasive effects on different parts of the macroeconomy.
- The advent of 'high-dimensional' datasets has seen the development of predictive models using:
  - latent factors (e.g. Stock and Watson (2015)).
  - variable selection from the pool of regressors implied by the data (e.g. Hastie et al (2015), Beloni et al (2011)).
- The use of PC's in empirical macro presumes that the underlying factors are strong.
- It is therefore of interest to provide estimates of the strength of macro shocks.

Data

- We make use of an updated version of the macroeconomic dataset of Stock and Watson (2012, SW).
- We assume the macroeconomic shocks are unobserved and estimate the strength of the strongest of such shocks implied by the data.
- The dataset consists of balanced quarterly observations over the period 1988Q1-2019Q2 ($T = 126$) on $n = 187$ out of the 200 macroeconomic variables used in SW.

Table: Strength estimates of the strongest unobserved factor using the CSA of the SW dataset ($n = 187$ variables) and the corresponding exponent of cross section dependence (CSD)

	Q1 1988 - Q4 2007 ($T = 80$)			Q1 1988 - Q2 2019 ($T = 126$)		
	$\hat{\alpha}^*_{0.05}$	$\hat{\alpha}$	$\hat{\alpha}^*_{0.95}$	$\hat{\alpha}^*_{0.05}$	$\hat{\alpha}$	$\hat{\alpha}^*_{0.95}$
			$p = 0.10$			
Strength of CSA ($\delta = 1/4$)	0.962	0.964	0.966	0.928	0.930	0.933
Strength of CSA ($\delta = 1/2$)	0.957	0.958	0.959	0.918	0.920	0.922
Exponent of CSD	0.833	0.873	0.913	0.858	0.920	0.981
			$p = 0.05$			
Strength of CSA ($\delta = 1/4$)	0.962	0.964	0.966	0.927	0.929	0.931
Strength of CSA ($\delta = 1/2$)	0.957	0.958	0.959	0.912	0.914	0.915
Exponent of CSD	0.833	0.873	0.913	0.856	0.918	0.979

Notes: *90% confidence bands. In the computation of the strength of CSA, parameters p and δ are used when setting the critical value for $\hat{\alpha}$). The exponent of CSD corresponds to the most robust estimator of cross-sectional dependence proposed in Bailey et al (2016) and corrects for both serial correlation in the factors and weak cross-sectional dependence in the error terms.

Commentary: unobserved common shocks

- We estimated the strength (α) of the strongest shock using the cross section average of the variables in the dataset.
- We computed estimates of α for the pre-crisis period, 1988Q1 to 2007Q4, as well as for the full sample period ending on 2019Q2.
- They are clustered around 0.94 and strictly below 1, and are quite robust to the choice of the parameters p and δ, and to the time period considered.
- These estimates suggest that whilst there exist strong macroeconomic shocks, their effects are not nearly as pervasive as have been assumed in the factor literature.
- This finding is further corroborated by the estimates of the exponent of cross-sectional dependence (CSD) of BKP.

Conclusion

- This paper addresses the problem of estimating the strength of individual factors that underlie cross-sectional dependencies in large panels.
- This can be of interest, for example, for pricing of risk in empirical finance, or for quantifying the pervasiveness of macroeconomic shocks.
- It proposes a novel estimator of factor strength based on the number of statistically significant t-statistics in a regression of each unit in the panel on the factor under consideration.
- It provides inferential theory for the proposed estimator.
- Detailed and extensive Monte Carlo and empirical analyses showcase the potential of the proposed method.
- Further research is required to link our analysis to the problem of factor selection discussed by Feng et al (2020).

Appendix

Design

- Factor loadings, γ_{i1} and γ_{i2}:
  - We generate $v_{ij} \sim IIDU(\mu_{v_j} - 0.2, \mu_{v_j} + 0.2)$, for $i = 1, 2, \ldots, n$ and $j = 1, 2$ (such that $E(v_{ij}) = \mu_{v_j}$).
  - We set $\mu_{v_1} = \mu_{v_2} = 0.71$.
  - We randomly assign $[n^{\alpha_{10}}]$ and $[n^{\alpha_{20}}]$ of these r.v.s as elements of vectors $\gamma_j = (\gamma_{2j}, \gamma_{2j}, \ldots, \gamma_{nj})', j = 1, 2$.
  - We consider values of $(\alpha_{10}, \alpha_{20})$ in the range $[0.75, 1.00]$ starting with 0.75 and rising to 1 at 0.05 increments.
  - This amounts to 36 experiments for all combinations of a_{10} and a_{20}.

Back to mc

Reported statistics

- We report bias and RMSE of $\hat{\alpha}_j, j = 1, 2$ which is estimated by (**??**), with $p = 0.10$ and $\delta = 1/4$.

- We compute size and power of tests of $H_0 : \alpha_j = \alpha_{j0}$ against $H_1 : \alpha_j = \alpha_{ja}$ for $\alpha_{ja} = \alpha_{j0} + \kappa$, $\kappa = -0.05, -0.045, \ldots, 0.045, 0.05$ at 5% level, for all $\alpha_{j0} \in [0.75, 1.00)$.

- For both tests, we use the test statistic:

$$z_{\hat{\alpha}_j:\alpha_{j0}} = \frac{(\ln n)\left(\hat{\alpha}_j - \alpha_{j0}\right) - p\left(n - n^{\hat{\alpha}_j}\right) n^{-\delta-\hat{\alpha}_j}}{\left[p\left(n - n^{\hat{\alpha}_j}\right) n^{-\delta-2\hat{\alpha}_j}\left(1 - \frac{p}{n^{\delta}}\right)\right]^{1/2}}, \; j = 1, 2.$$

- Size and power results are not reported when α_{j0} and/or α_{ja} is equal to unity (estimator is ultraconsistent).

- All combinations of $n = \{100, 200, 500, 1000\}$ and $T = \{120, 200, 500, 1,000\}$ are considered for each experiment. The no. of replications is set to $R = 2,000$.

- The values of c_i and γ_{ij} are redrawn at each replication.

Back to mc

Commentary: observed factors

- Overall, bias, RMSE, size and power results for all experiments when errors are generated as Gaussian or as non-Gaussian are very similar.
- In the case of a single factor, bias and RMSE are universally low and gradually decrease as n, T and α_{10} rise.
- Especially when $\alpha_{10} = 1$, bias and RMSE are negligible even when $T = 120$.
- Given the low variance of our estimator, rejection probabilities under the null are sensitive to the bias of $\hat{\alpha}_1$, especially for smaller values of α_{10}.
- For $\alpha_{10} = 0.95$ correct empirical size is achieved even for moderate values of T.
- The corresponding power functions show that the proposed estimator is very precisely estimated for all values of α_{10} considered, and for all (n, T) combinations.
- Similar results hold for models with two observed factors, irrespective of whether the factors are orthogonal ($\rho_{12} = 0$) or moderately correlated ($\rho_{12} = 0.3$).
- In the case of model misspecification when $\alpha_{10} = 1$, omitting a second relevant and correlated factor does not unduly affect the performance of our estimator.

Back to mc

Commentary: unobserved factors

- When a single factor is generated, bias and RMSE of $\hat{\alpha}$ deteriorate somewhat compared to the observed factor case, especially for smaller values of α_0.
- The empirical size is particularly elevated for values of $\alpha_0 \leq 0.9$.
- However, for large sample sizes and values for α_0 close to unity, our estimator seems to be reasonably well behaved.
- In the case of two unobserved factors, when $\alpha_{10} = 1$ and $\alpha_{20} = 0.51$, then bias and RMSE results are universally very low, as in the one unobserved factor case.
- As α_{20} rises towards unity, a slight deterioration in results can be detected, for small values of T, e.g. $T = 120$, but the size distortions vanish as T increases.
- The ultraconsistency of our estimator when $\alpha_{10} = 1$ is evident by the values for both bias and RMSE measures (values scaled by 10,000).
- When $\alpha_{10}, \alpha_{20} < 1$, estimating α_0 becomes more challenging.
- Setting $\alpha_{10} = 0.95$ and $\alpha_{20} \in [0.51, 1]$, results worsen when α_{10} is close to α_{20}, but improve as $|\alpha_{10} - \alpha_{20}|$ grows larger, for all n and T (see conditions of Theorem 2).

Back to mc

Table: Bias and RMSE ($\times 10,000$) of estimating factor strength in the case of experiment 4 (observed misspecified single factor - Gaussian errors) when set to 1.00

	Bias ($\times 10,000$)				RMSE ($\times 10,000$)			
n\T	120	200	500	1000	120	200	500	1000
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.75$							
100	-0.68	-0.01	0.00	0.00	4.03	0.49	0.00	0.00
200	-0.62	-0.03	0.00	0.00	2.55	0.52	0.00	0.00
500	-0.76	-0.04	0.00	0.00	1.86	0.38	0.00	0.00
1000	-0.76	-0.05	0.00	0.00	1.42	0.27	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.80$							
100	-0.70	-0.01	0.00	0.00	4.03	0.49	0.00	0.00
200	-0.54	-0.02	0.00	0.00	2.38	0.47	0.00	0.00
500	-0.72	-0.04	0.00	0.00	1.82	0.35	0.00	0.00
1000	-0.71	-0.04	0.00	0.00	1.37	0.26	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.85$							
100	-0.61	-0.01	0.00	0.00	3.78	0.49	0.00	0.00
200	-0.45	-0.01	0.00	0.00	2.15	0.37	0.00	0.00
500	-0.62	-0.04	0.00	0.00	1.64	0.35	0.00	0.00
1000	-0.65	-0.04	0.00	0.00	1.27	0.24	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.90$							
100	-0.47	0.00	0.00	0.00	3.28	0.00	0.00	0.00
200	-0.39	-0.01	0.00	0.00	2.02	0.30	0.00	0.00
500	-0.48	-0.02	0.00	0.00	1.42	0.25	0.00	0.00
1000	-0.51	-0.03	0.00	0.00	1.11	0.22	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 0.95$							
100	-0.35	0.00	0.00	0.00	2.85	0.00	0.00	0.00
200	-0.31	-0.01	0.00	0.00	1.80	0.30	0.00	0.00
500	-0.34	-0.01	0.00	0.00	1.24	0.16	0.00	0.00
1000	-0.35	-0.02	0.00	0.00	0.88	0.18	0.00	0.00
	$\alpha_{10} = 1.00$, $\alpha_{20} = 1.00$							
100	-0.16	0.00	0.00	0.00	2.01	0.00	0.00	0.00
200	-0.13	0.00	0.00	0.00	1.10	0.00	0.00	0.00
500	-0.15	0.00	0.00	0.00	1.01	0.07	0.00	0.00
1000	-0.13	0.00	0.00	0.00	0.57	0.06	0.00	0.00

Table: Ranking of 65 factors in terms of the % of months their estimated strengths exceed the threshold of 0.90 during September 1989 to December 2017 and corresponding time averages of $\hat{\alpha}_{s,\tau}$, $s = 1, 2, \ldots, 65$, over different subsamples

	% of months when $\hat{\alpha}_{s,\tau} > 0.90$ over:	Time averages of $\hat{\alpha}_{s,\tau}$ over:			
Factor	Full sample	Full sample	September 1989 - August 1999	September 1999 - August 2009	September 2009 - December 2017
Market	100.0	0.990	0.999	0.974	0.997
Leverage	37.9	0.827	0.739	0.932	0.808
Sales to cash	37.9	0.817	0.716	0.936	0.793
Cash flow-to-price	37.9	0.832	0.765	0.933	0.792
Net debt-to-price	37.9	0.838	0.753	0.936	0.823
Earnings to price	37.9	0.811	0.743	0.935	0.745
Net payout yield	37.6	0.844	0.769	0.932	0.829
Years since first Compustat coverage	37.6	0.828	0.724	0.935	0.823
Cash flow to price ratio	37.6	0.818	0.737	0.934	0.775
Quick ratio	37.4	0.835	0.782	0.936	0.777
Altman's Z-score	37.4	0.828	0.740	0.931	0.808
Payout yield	37.1	0.851	0.785	0.932	0.831
Earnings volatility	37.1	0.852	0.779	0.936	0.840
Change in shares outstanding	37.1	0.805	0.671	0.932	0.815
Enterprise book-to-price	36.8	0.830	0.741	0.933	0.812
Cash holdings	36.8	0.826	0.740	0.935	0.797
Dividend to price	36.5	0.846	0.789	0.932	0.811
Depreciation / PP&E	36.5	0.851	0.813	0.930	0.801
Kaplan-Zingales Index	36.2	0.822	0.731	0.930	0.801
R&D-to-sales	36.2	0.815	0.731	0.923	0.786
Cash flow volatility	36.2	0.783	0.617	0.924	0.812
Accrual volatility	36.2	0.779	0.613	0.926	0.803
Current ratio	35.9	0.846	0.815	0.926	0.785
Idiosyncratic return volatility	35.6	0.851	0.799	0.923	0.828
Debt capacity/firm tangibility	35.6	0.829	0.735	0.920	0.832
Maximum daily return	35.3	0.838	0.764	0.927	0.821
Bid-ask spread	35.3	0.847	0.786	0.931	0.821
Cash productivity	35.3	0.819	0.751	0.911	0.789
Return volatility	34.7	0.844	0.786	0.922	0.820
...					