Factor models with many assets: strong factors, weak factors, and the two-pass procedure

Stanislav Anatolyev\textsuperscript{1} Anna Mikusheva\textsuperscript{2}

\textsuperscript{1}CERGE-EI and NES
\textsuperscript{2}MIT

December 2017
Linear factor-pricing models

- Factor-pricing model:

\[ Er_{it} = \lambda' \beta_i, \text{ where } \beta_i = \text{var}(F_t)^{-1}\text{cov}(F_t, r_{it}) \]

- Classical estimation approach is the two-pass procedure (Fama and MacBeth, 1973) with standard error correction (Shanken, 1992)
  1. Estimate \( \beta_i \) for each portfolio from time-series regression;
  2. Estimate \( \lambda \) from cross-sectional regression of average returns on estimated betas.

- Quality control:
  - Is price of risk non-zero? Test: \( H_0 : \lambda \neq 0 \);
  - Do these risks price market? Specification test \( H_0 : Er_{it} = \lambda' \beta_i \);
  - How much does risk exposure explain a variation in average returns? Second-pass \( R^2 \).
Linear factor-pricing models

- First and most known: CAPM (Sharpe 1964, Linner 1965)
- The second most well-known is Fama-French (1993): includes market portfolio, size factor ‘SMB’ (small-minus-big) and book-to-market factor ‘HML’ (high-minus-low).
- Some models have factors based on market behavior: example-momentum factor ‘MOM’ (Jegadeesh and Titman, 1993);
- Some have macroeconomic factors: example- consumption-to-wealth ratio ‘cay’ (Lettau and Ludvigson, 2001)
Problem 1: weak identification?

- If some of the observed factors are only weakly correlated with returns, then the second-pass parameters may be weakly identified.
  - Kan and Zhang (1999): useless factors lead to spurious inference
  - Kleibergen and Zhan (2015): weak factors may arise from poor measurement of true factors
Empirical fact found in Kleibergen and Zhan (2015): many well-known linear factor-pricing models have very strong remaining factor structure present in the residuals.

Example: for all Lettau and Ludvigson (2001) specifications first three principle components of residuals explain 82% - 96% of remaining cross-sectional variation.

One found exception to this rule: Fama and French.
Observation in our paper: Large $T$ and large $N$?

- Traditionally (and in all mentioned papers) the asymptotic results are derived under assumption:

  $$N \text{ is fixed, } T \to \infty$$

- However, the most often used datasets are:
  - Fama-French: $N = 25, T = 141$;

- $N$ and $T$ are comparable in size
- More adequate asymptotic approximations may result from both $N \to \infty$ and $T \to \infty$
Our setup includes simultaneously

- **Weak observed factors:** Some observed factors are only weakly correlated: we model corresponding risk exposure coefficients $\beta_i$ as being of order $O(1/\sqrt{T})$. Thus, first-stage estimation error is of the same order of magnitude as the coefficients themselves.

- **Missing factors:** There is a strong factor structure present in error terms.

- **Large-$N$-large-$T$ asymptotics:** Many assets-long time span:

  $$N, T \to \infty$$
Findings of our paper

- We prove that the classical two-pass procedure fails in our setting: inconsistent estimates of the premia on weak factors, invalid inferences and significant finite-sample bias for estimate of risk premia on strong observed factor.
- We propose new procedures that provide consistent estimators for risk premia and guarantee asymptotically gaussian inferences.
Findings of our paper

We develop an estimation procedure for risk premia in an environment with many assets, weak included factors and strong excluded factors with the following features:

- it yields consistent estimates when the traditional two-pass procedure fails;
- it yields consistent estimates without knowledge of which factors are strong and which are weak;
- it does not lose efficiency if the traditional two-pass procedure works;
- it is a procedure of the ‘press button’ type: easy-to-implement, uses standard estimation techniques.
Outline

1. Introduction
2. Setup and main assumptions
3. Two-pass procedure fails: Why?
4. Our proposed solution
5. Some famous papers revisited
Setup

- We observe excess returns on assets or portfolios \( \{r_{it}, i = 1, \ldots, N, t = 1, \ldots, T\} \) and \( k_F \times 1 \) risk factors \( \{F_t, t = 1, \ldots, T\} \) that follow the correctly-specified linear factor-pricing model:

\[
Er_{it} = \lambda' \beta_i, \quad \text{where} \quad \beta_i = \text{var}(F_t)^{-1} \text{cov}(F_t, r_{it})
\]

- This is equivalent to assuming that

\[
r_{it} = \lambda' \beta_i + (F_t - EF_t)' \beta_i + \varepsilon_{it},
\]

where the random error terms \( \varepsilon_{it} \) have mean zero and are uncorrelated with \( F_t \). We treat \( \lambda \) and \( \beta_i \) as non-random, while \( r_{it}, F_t, \varepsilon_{it} \) are random.
Setup: weak observed factors

- We will divide factors $F_t = (F'_{t,1}, F'_{t,2})'$ and exposures $\beta_i = (\beta'_{i,1}, \beta'_{i,2})'$ into “strong” and “weak”:
  - $\beta_{i,2} = \frac{b_i}{\sqrt{T}}$, where we make the same assumptions about size of $\beta_{i,1}$ and size of $b_i$ (they are $O(1)$).
  - Estimation error for each $\beta_i$ is of order $O_p(1/\sqrt{T})$, similar to size of $\beta_{i,2}$.
  - In setting with $N$-fixed and $T \to \infty$, this corresponds to weak identification.

- We do not assume that econometrician knows which factors are weak or the number of weak factors (our results hold for more general assumptions, that some linear combination of factors is weak).
Setup: missing factors

Model:

\[ r_{it} = \lambda' \beta_i + (F_t - EF_t)' \beta_i + \varepsilon_{it}, \]

We assume that error terms are not auto-correlated (efficient market hypothesis) but have non-trivial cross-sectional dependence - they have unobserved factor structure:

\[ \varepsilon_{it} = \nu_t' \mu_i + e_{it}, \]

where

- \( \nu_t \) are unobserved random variables; have mean zero and unit variance (normalization); uncorrelated with \( e_{it} \);
- \( \mu_i \) - unknown constant loadings of size \( O(1) \);
- \( e_{it} \) are weakly cross-sectionally correlated. 
Outline

1. Introduction
2. Setup and main assumptions
3. Two-pass procedure fails: Why?
4. Our proposed solution
5. Some famous papers revisited
Two-pass procedure fails: Why?

Asymptotics of the two-pass procedure

- If all observed factors are strong: \( \Rightarrow \sqrt{T}(\hat{\lambda}_{TP} - \lambda) \Rightarrow N(0, V) \).
- If some observed factors are weak, but no missing factors in errors: \( \Rightarrow \) "errors-in-variables" bias:
  - \( \hat{\lambda}_{TP,1} \) is consistent and Gaussian, but biased (inferences are not valid),
  - \( \hat{\lambda}_{TP,2} \) is inconsistent
- If some observed factors are weak, and some missing factors in errors:
  \( \Rightarrow \) "errors-in-variables" + "omitted variable":
  - \( \hat{\lambda}_{TP,1} \) is consistent, but biased and non-standard distribution,
  - \( \hat{\lambda}_{TP,2} \) is inconsistent
Why two-pass fails? No missing factors case

- Assume some observed factors are weak, but no factor structure in errors
  \[ r_{it} = \lambda' \beta_i + (F_t - EF_t)' \beta_i + e_{it}, \]
- \( e_{it} \) are weakly dependent
- First-pass estimates:
  \[ \hat{\beta}_i = \left( \sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t' \right)^{-1} \sum_{t=1}^{T} \tilde{F}_t r_{it} = (\beta_i + u_i)(1 + o_p(1)), \]

where \( u_i = \frac{1}{T} \sum_{t=1}^{T} F_t e_{it} \) are 'asymptotically uncorrelated' for different \( i \) and unrelated to \( \beta_i \).
Why two-pass fails? No missing factors case

- Ideal regression: if one regresses $\bar{r}_i = \frac{1}{T} \sum_{t=1}^{T} r_{it}$ on $\beta_i$, then will have consistent estimate of $\lambda$

- But we have instead only estimates and $u_i = O(1/\sqrt{T})$

$$
\begin{pmatrix}
\hat{\beta}_{i,1} \\
\hat{\beta}_{i,2}
\end{pmatrix}
= 
\begin{pmatrix}
\beta_{i,1} \\
\beta_{i,2}
\end{pmatrix}
+ 
\begin{pmatrix}
u_{i,1} \\
u_{i,2}
\end{pmatrix}
= 
\begin{pmatrix}
\beta_{i,1}(1 + o(1)) \\
\beta_{i,2} + u_{i,2}
\end{pmatrix}
$$

- Mistake in $\beta_{i,2}$ is of the same order of magnitude as coefficient itself. It behaves like classical measurement error!

- Regression of $\bar{r}_i$ on $\hat{\beta}_i$ has an attenuation bias!
Two-pass procedure fails: Why?

No missing factors case: Solution

Idea:

- Split sample in two $T_1 \sqcup T_2 = \{1, \ldots, T\}$
- Estimate $\beta_i$ twice:

$$\hat{\beta}_i^{(j)} = \left(\sum_{t \in T_j} \tilde{F}_t \tilde{F}_t'\right)^{-1} \sum_{t \in T_j} \tilde{F}_t r_{it} = (\beta_i + u_i^{(j)})(1 + o_p(1)), \quad j = 1, 2$$

- Estimation mistakes $u_i^{(1)}$ and $u_i^{(2)}$ are (asymptotically) uncorrelated
- Use $\hat{\beta}_i^{(1)}$ as a regressor and $\hat{\beta}_i^{(2)}$ as instrument (or vice versa, or both and average final estimates)
- Idea of sample-splitting (and its extreme version: leave-one-out or jackknife) has been used in many-weak-IV model (Hansen, Hausman and Newey, 2008)
Factors in errors. Why two-pass fails?

- Model with factor structure in errors:

\[ r_{it} = \lambda' \beta_i + (F_t - EF_t)' \beta_i + v_t' \mu_i + e_{it}, \]

\( v_t \) is unobserved and \( \mu_i \) are unknown, \( e_{it} \) are weakly cross-correlated.

- First step

\[ \hat{\beta}_i = \left( \sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t' \right)^{-1} \sum_{t=1}^{T} \tilde{F}_t r_{it} = \left( \beta_i + \frac{\eta_T \mu_i}{\sqrt{T}} + u_i \right) (1 + o_p(1)), \]

where

\[ \eta_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{-1}^{F_t} \tilde{F}_t v_t' \]

is coming from unobserved factor structure.
Factors in errors. Why two-pass fails?

\[ \hat{\beta}_i = \left( \beta_i + \eta T \mu_i \sqrt{T} + u_i \right) \left( 1 + o_p(1) \right), \]

- Now the estimation error \( \eta T \mu_i \sqrt{T} + u_i \) is NOT classical measurement error:
  - both terms \( \eta T \mu_i \sqrt{T} \) and \( u_i \) are stochastically of order \( O_p \left( \frac{1}{\sqrt{T}} \right) \)
  - estimation errors are cross-correlated (for different \( i \)) due to term \( \frac{\eta T \mu_i}{\sqrt{T}} \)
  - estimation error may be ‘correlated’ with regressor if ‘sample correlation’ between \( \beta_i \) and \( \mu_i \) is non-zero
Factors in errors. Why two-pass fails?

- Model with factor structure in errors:

\[ r_{it} = \lambda' \beta_i + (F_t - EF_t)' \beta_i + \nu_t \mu_i + e_{it}, \]

- Ideal regression:

\[ y_i = \sqrt{T} \tilde{r}_i = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} r_{it} = \tilde{\lambda}' \left( \sqrt{T} \beta_i \right) + \eta'_v \mu_i + \epsilon_i, \]

- If there is \( \mu_i \) but you know \( \beta_i \) only- we have omitted variable, it will cause omitted variable bias if 'sample correlation' between \( \beta_i \) and \( \mu_i \) is non-zero.
Factors in errors. Why two-pass fails?

Summary:
- if there is no factor structure in errors - we have classical error-in-variables problem and associated attenuation bias
- If we have factor structure in errors we additionally have:
  - non-classical error-in-variable (mistakes in regressor $\hat{\beta}_{i,2}$ are cross-correlated and ‘correlated’ with $\beta_i$)
  - even if we know $\beta_i$ there is omitted variable bias in the ‘ideal’ regression if ’sample correlation’ between $\beta_i$ and $\mu_i$ is non-zero.
Outline

1. Introduction
2. Setup and main assumptions
3. Two-pass procedure fails: Why?
4. Our proposed solution
5. Some famous papers revisited
Our proposed solution: Idea

- We reconsider sample-splitting.
- We have an estimate of $\beta_i$ for each sub-sample

$$\hat{\beta}^{(j)}_i = \left( \sum_{t \in T_j} \tilde{F}_t \tilde{F}_t' \right)^{-1} \sum_{t \in T_j} \tilde{F}_t r_{it} = \left( \beta_i + \frac{\eta_j \mu_i}{\sqrt{T}} + u^{(j)}_i \right) \left(1 + o_p(1)\right),$$

where

$$\eta_j = \frac{1}{\sqrt{|T_j|}} \sum_{t \in T_j} \sum_{F}^{-1} \tilde{F}_t v'_t \Rightarrow N(0, \Omega_{Fv}).$$

- $\eta_j$ are independent for different $j$ and independent from errors $u^{(j)}_i$. 
Our proposed solution: Idea

\[ \hat{\beta}_i^{(j)} = \left( \beta_i + \frac{\eta_i \mu_i}{\sqrt{T}} + u_i^{(j)} \right) \left(1 + o_p(1)\right), \]

- We can construct proxy for \( \mu_i \) (!!!)

\[ \hat{\beta}_i^{(1)} - \hat{\beta}_i^{(2)} = \left( \frac{\eta_1}{\sqrt{|T_1|}} - \frac{\eta_2}{\sqrt{|T_2|}} \right) \mu_i + (u_i^{(1)} - u_i^{(2)}) \]

- If \( |T_j| = T/4 \), then ‘random’ coefficient \( \left( \frac{\eta_1}{\sqrt{|T_1|}} - \frac{\eta_2}{\sqrt{|T_2|}} \right) = O\left(\frac{1}{\sqrt{T}}\right) \)

  and error \( (u_i^{(1)} - u_i^{(2)}) = O\left(\frac{1}{\sqrt{T}}\right) \)

- Proxy \( \hat{\beta}_i^{(1)} - \hat{\beta}_i^{(2)} \) mis-measures \( \mu_i \), but measurement error is classical: not cross-correlated and not correlated with regressors.
Our proposed solution: Idea

- Split sample into 4 equal sub-samples.
- Estimate $\hat{\beta}^{(j)}_i$ for $j = 1, \ldots, 4$.
- Run IV regression of $\tau_i$ on regressors $\hat{\beta}^{(1)}_i$ and proxy based on $\hat{\beta}^{(1)}_i - \hat{\beta}^{(2)}_i$ with instruments $\hat{\beta}^{(3)}_i$ and $\hat{\beta}^{(3)}_i - \hat{\beta}^{(4)}_i$.
- For efficiency considerations you may repeat this 4 times circulating indices 1-4.
- Average estimates you obtain for $\lambda$.
- We also provide formula for how to calculate covariance matrix for our estimate.
Our proposed solution

- The exact asymptotic distribution of $\hat{\lambda}_{4S}$ is not Gaussian but rather mixed Gaussian. The estimated variance matrix is asymptotically random though non-degenerate with probability 1.
- This is due to the fact that the coefficient on proxy for $\mu_i$ is random. It leads to information contained in second stage IV being random, though NOT weak with probability 1.
- Our 4-split estimator:
  - it yields consistent estimates when the traditional two-pass procedure fails;
  - it yields consistent estimates without knowledge of which factors are strong and which are weak;
  - it does not lose efficiency if the traditional two-pass procedure works;
  - it is a procedure of the ‘push-button’ type: easy-to-implement, uses standard estimation techniques.
Outline

1. Introduction
2. Setup and main assumptions
3. Two-pass procedure fails: Why?
4. Our proposed solution
5. Some famous papers revisited
### Empirical application (Fama–French portfolios)

<table>
<thead>
<tr>
<th>no.</th>
<th>specification</th>
<th>5 main principal components in residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Market, SMB, HML</td>
<td>0.29 0.14 0.11 0.07 0.04</td>
</tr>
<tr>
<td>2</td>
<td>Market, HML</td>
<td>0.62 0.10 0.05 0.03 0.03</td>
</tr>
<tr>
<td>3</td>
<td>Market, HML, cay</td>
<td>0.62 0.10 0.05 0.03 0.03</td>
</tr>
</tbody>
</table>
### Empirical application (Fama–French portfolios)

<table>
<thead>
<tr>
<th>no.</th>
<th>specification</th>
<th>5 main principal components in residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Market, SMB, HML</td>
<td>0.29, 0.14, 0.11, 0.07, 0.04</td>
</tr>
<tr>
<td>2</td>
<td>Market, HML</td>
<td>0.62, 0.10, 0.05, 0.03, 0.03</td>
</tr>
<tr>
<td>3</td>
<td>Market, HML, cay</td>
<td>0.62, 0.10, 0.05, 0.03, 0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>no.</th>
<th>risk factor</th>
<th>Market, SMB, HML, cay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>conventional two-pass</td>
<td>2.70, 0.69, 1.96, 0.58</td>
</tr>
<tr>
<td></td>
<td>average four-split</td>
<td>2.80, 0.46, 1.29, 0.84</td>
</tr>
<tr>
<td>3</td>
<td>conventional two-pass</td>
<td>2.55, 1.92, 0.027</td>
</tr>
<tr>
<td></td>
<td>average four-split</td>
<td>2.06, 2.44, −0.009</td>
</tr>
</tbody>
</table>
### Empirical application (industry portfolios)

<table>
<thead>
<tr>
<th>Specification</th>
<th>5 main principal components in residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market, SMB, HML, MOM</td>
<td>0.14 0.12 0.08 0.06 0.04</td>
</tr>
</tbody>
</table>
Empirical application (industry portfolios)

<table>
<thead>
<tr>
<th>specification</th>
<th>5 main principal components in residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market, SMB, HML, MOM</td>
<td>0.14 0.12 0.08 0.06 0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>risk factor</th>
<th>Market</th>
<th>SMB</th>
<th>HML</th>
<th>MOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>conventional two-pass</td>
<td>1.05</td>
<td>−0.27</td>
<td>−0.00</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.19</td>
<td>0.15</td>
<td>0.35</td>
</tr>
<tr>
<td>average four-split</td>
<td>1.15</td>
<td>−1.10</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>0.24</td>
<td>0.18</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Conclusion

What we have done here:

- Showed that conventional two-pass procedure gives unreliable estimates of risk premia in empirically-relevant situations
- Proposed alternative “press buttons” procedure robust to weak factors and strong missing factors, based on split-sample IV
- Alternative procedure yields consistent and asymptotically normal estimates under many-asset, weak-factor asymptotics