Ex-post Risk Premia Estimation using Large Cross Sections of Assets: Regularization and Clustering

Soohun Kim — KAIST
Georgios Skoulakis — University of Piraeus

23rd CRETE Conference
Tinos
July 09, 2025

Agenda

- We are interested in testing implications of asset pricing models using large cross sections of stocks (as opposed to portfolios)
- Using all the available data at once is challenging due to significant amounts of missing data (missing-at-random assumptions seem unrealistic in this context)
- We focus on short time windows (small T) and use all stocks available (large N)
- Following Shanken (RFS 1992), we center our analysis around ex-post risk premia (as opposed to ex-ante risk premia)
- Cross-sectional regressions à la Fama-MacBeth are plagued by a severe error-in-variables (EIV) problem due to beta estimation error

Related Literature

- Shanken (RFS 1992)
- Jagannathan, Skoulakis, and Wang (HC 2010)
- Kim and Skoulakis (JE 2018) (Two-stage Regression Calibration approach with Clustering)
- Jegadeesh, Noh, Pukthuanthong, Roll and Wang (JFE 2019) (IV approach)
- Raponi, Robotti, and Zaffaroni (RFS 2020) (Inference for the Shanken estimator)
- Kim and Skoulakis (WP 2024) (Two-stage IV-GMM approach with Clustering)
- Kleibergen and Zhan (WP 2024) (alternative to Shanken estimator)

EIV Problem: Illustration

- Illustration using a single factor model
- For simplicity, ignore the role of the intercept
- Return generating process: $R_i = \beta_i \lambda^* + u_i$, $i = 1, \ldots, N$
- LLN assumptions on shocks and betas: as $N \to \infty$,

$$\frac{1}{N} \sum_{i=1}^{N} u_i \xrightarrow{p} 0, \quad \frac{1}{N} \sum_{i=1}^{N} \beta_i u_i \xrightarrow{p} 0$$

- If the true betas were known, we would regress R_i on β_i to get

$$\lambda^\dagger = \frac{\sum_{i=1}^{N} \beta_i R_i}{\sum_{i=1}^{N} \beta_i^2} \xrightarrow{p} \lambda^*$$

EIV Problem: Illustration

- Error-in-variables problem due to beta estimation error
- True beta β is estimated by OLS $\hat{\beta}$: $\hat{\beta}_i = \underbrace{\beta_i}_{\text{true}} + \underbrace{\xi_i}_{\text{estimation error}}$
- Assumptions on the cross section of betas:

$$\bar{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta_i \to \mu_\beta, \quad \frac{1}{N}\sum_{i=1}^{N} \left(\beta_i - \bar{\beta}\right)^2 \to \sigma_\beta^2$$

- Further natural LLN assumptions on the primitives (betas and shocks) imply that, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} \xi_i \xrightarrow{p} 0, \quad \frac{1}{N}\sum_{i=1}^{N} \xi_i^2 \xrightarrow{p} \sigma_\xi^2,$$
$$\frac{1}{N}\sum_{i=1}^{N} \beta_i\xi_i \xrightarrow{p} 0, \quad \frac{1}{N}\sum_{i=1}^{N} u_i\xi_i \xrightarrow{p} 0$$

EIV Problem: Illustration

- The error-in-variables problem gives rise to attenuation bias
- Regressing r_i on $\hat{\beta}_i$ yields

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} \hat{\beta}_i R_i}{\sum_{i=1}^{N} \hat{\beta}_i^2} \xrightarrow{p} \frac{\mu_\beta^2 + \sigma_\beta^2}{\mu_\beta^2 + \sigma_\beta^2 + \sigma_\xi^2} \lambda^* < \lambda^*$$

  under the typical assumption $\lambda^* > 0$

- The worse the beta estimation error, the higher σ_ξ^2 and the larger the bias in the $\hat{\lambda}$ estimates
- An adjustment is needed so that the risk premia estimates are pulled away from zero

Digression: Ridge Regression

- Linear regression: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$
- When $\mathbf{X}'\mathbf{X}$ is near-singular, the OLS estimator of $\boldsymbol{\beta}$ is unstable
- One remedy is the ridge regression approach which yields

$$\hat{\boldsymbol{\beta}}_{\mathtt{RDG}} = (\mathbf{X}'\mathbf{X} + \rho\mathbf{I})^{-1}\,\mathbf{X}'\mathbf{y}$$

  where $\rho > 0$ is a suitable constant

- The ridge estimator can be obtained by solving the optimization problem

$$\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\,(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \rho\,(\boldsymbol{\beta}'\boldsymbol{\beta})$$

Regularization

- Ridge regression is a particular example of regularization
- Other popular examples include lasso and elastic net
- The idea is to adjust the objective function towards a specific goal
- We will use the same broad idea in order to deal with the error-in-variables problem
- Instead of penalizing the estimates, as in the case of ridge regression, we need to boost the estimates to correct for the attenuation bias

Data Generating Process

- N assets and T time periods with $N >> T$. Formally, T is fixed and $N \to \infty$.
- $\mathbf{R}_t$ $(N \times 1)$ excess returns with $\boldsymbol{\mu}_R = \mathrm{E}[\mathbf{R}_t]$, $\mathbf{f}_t$ $(K \times 1)$ factors with $\boldsymbol{\mu}_f = \mathrm{E}[\mathbf{f}_t]$.
- Second-order moments $\boldsymbol{\Sigma}_f = \mathrm{E}[(\mathbf{f}_t - \boldsymbol{\mu}_f)(\mathbf{f}_t - \boldsymbol{\mu}_f)']$ $(K \times K)$, $\boldsymbol{\Sigma}_{Rf} = \mathrm{E}[(\mathbf{R}_t - \boldsymbol{\mu}_R)(\mathbf{f}_t - \boldsymbol{\mu}_f)']$ $(N \times K)$, beta matrix $\mathbf{B} = \boldsymbol{\Sigma}_{Rf}\boldsymbol{\Sigma}_f^{-1}$ $(N \times K)$.
- $\boldsymbol{\delta}_f$ $K \times 1$ ex-ante risk premia vector with $\boldsymbol{\mu}_R = \mathbf{B}\boldsymbol{\delta}_f$.

Data Generating Process

- Letting $\mathbf{u}_t = (\mathbf{R}_t - \boldsymbol{\mu}_R) - \mathbf{B}(\mathbf{f}_t - \boldsymbol{\mu}_f)$ and imposing $\boldsymbol{\mu}_R = \mathbf{B}\boldsymbol{\delta}_f$, we obtain the time-series representation:

$$\mathbf{R}_t = \mathbf{B}(\mathbf{f}_t - \boldsymbol{\mu}_f + \boldsymbol{\delta}_f) + \mathbf{u}_t, \text{ with } \mathrm{E}[\mathbf{u}_t] = \mathbf{0}_N, \ \mathrm{E}[\mathbf{u}_t \mathbf{f}_t'] = \mathbf{0}_{N \times K}.$$

- Cross-sectional averages

$$\bar{\mathbf{R}} = \mathbf{1}_N \lambda_0^* + \mathbf{B}\boldsymbol{\lambda}_f^* + \bar{\mathbf{u}} = \mathbf{X}\boldsymbol{\lambda}^* + \bar{\mathbf{u}}$$

where $\lambda_0^* = 0$,

$$\boldsymbol{\lambda}_f^* = \bar{\mathbf{f}} + (\boldsymbol{\delta}_f - \boldsymbol{\mu}_f) \text{ (ex-post risk premia)},$$

$\boldsymbol{\lambda}^* = [\lambda_0^* \ (\boldsymbol{\lambda}_f^*)']'$ and $\mathbf{X} = [\mathbf{1}_N \ \mathbf{B}]$.

Cross-sectional Assumptions

(i) For all $t = 1, \dots, T$, as $N \to \infty$,

$$\frac{1}{N}\mathbf{X}'\mathbf{u}_t = \left[\ \tfrac{1}{N}\textstyle\sum_{n=1}^{N} u_{n,t} \quad \tfrac{1}{N}\textstyle\sum_{n=1}^{N} u_{n,t}\boldsymbol{\beta}_n'\ \right]' \xrightarrow{p} \mathbf{0}_{K+1}.$$

(ii) For all $t, t' = 1, \dots, T$, as $N \to \infty$, $\frac{1}{N}\sum_{n=1}^{N} u_{n,t}u_{n,t'} \xrightarrow{p} v\mathbf{1}_{[t=t']}$, where v is a positive constant.

(iii) As $N \to \infty$, $\bar{\boldsymbol{\beta}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\beta}_n \to \boldsymbol{\mu}_\beta$.

(iv) As $N \to \infty$, $\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{\beta}_n - \boldsymbol{\mu}_\beta)(\boldsymbol{\beta}_n - \boldsymbol{\mu}_\beta)' \to \boldsymbol{\Sigma}_\beta$, where $\boldsymbol{\Sigma}_\beta$ is a symmetric and positive definite matrix.

Objective Function

- For a given candidate risk premia parameter vector $\boldsymbol{\lambda} = [\lambda_0 \ (\boldsymbol{\lambda}_f)']'$, define the vector of (estimated) pricing errors

$$\widehat{\alpha}(\boldsymbol{\lambda}) = \bar{\mathbf{R}} - (\mathbf{1}_N \lambda_0 + \hat{\mathbf{B}} \boldsymbol{\lambda}_f)$$

  where $\hat{\mathbf{B}}$ is the OLS estimator of $\mathbf{B}$

- In standard OLS fashion, we seek to minimize the average pricing error $\frac{1}{N}\widehat{\alpha}(\boldsymbol{\lambda})'\widehat{\alpha}(\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$

- Given $\boldsymbol{\lambda}_f$, the objective function is minimized by $\lambda_0 = \frac{1}{N}\mathbf{1}_N'(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f)$ and so the vector of pricing errors is

$$\widehat{\alpha}(\boldsymbol{\lambda}_f) = \mathbf{J}_N(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f)$$

  where $\mathbf{J}_N = \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N'$ (demeaning matrix)

Manifestation of the EIV Problem

- The objective function then becomes

$$Q_N(\boldsymbol{\lambda}_f) = \frac{1}{N}(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f)'\mathbf{J}_N(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f)$$

- Under standard assumptions on the primitives, as $N \to \infty$, we have $Q_N(\boldsymbol{\lambda}_f) \xrightarrow{p} Q(\boldsymbol{\lambda}_f)$, where

$$Q(\boldsymbol{\lambda}_f) = (\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*)'\boldsymbol{\Sigma}_\beta(\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*) + \frac{v}{T}(1 + \boldsymbol{\lambda}_f'\mathbf{P}_f\boldsymbol{\lambda}_f),$$

$\mathbf{P}_f = \left(\frac{1}{T}\mathbf{F}\mathbf{J}_T\mathbf{F}'\right)^{-1}$ (precision matrix) and v satisfies $\frac{1}{N}\mathbf{U}'\mathbf{U} \xrightarrow{p} v\mathbf{I}_T$.

Remedy by Regularization

- If the betas were known, then the above limit would be $(\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*)'\boldsymbol{\Sigma}_\beta(\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*)$. The term $\frac{v}{T}\left(1 + \boldsymbol{\lambda}_f'\mathbf{P}_f\boldsymbol{\lambda}_f\right)$ reflects beta estimation error.
- The limiting objective function $Q(\boldsymbol{\lambda}_f)$ is **not** minimized when $\boldsymbol{\lambda}_f$ equals the true ex-post risk premia vector $\boldsymbol{\lambda}_f^*$.
- We need to adjust the objective function so that $\boldsymbol{\lambda}_f^*$ is the unique minimizer of the limiting objective function.
- Proposal `S` (subtraction):

$$Q_{S,N}(\boldsymbol{\lambda}_f) = Q_N(\boldsymbol{\lambda}_f) - \boldsymbol{\lambda}_f'\mathbf{W}_S\boldsymbol{\lambda}_f$$

- Proposal `D` (division):

$$Q_{D,N}(\boldsymbol{\lambda}_f) = \frac{Q_N(\boldsymbol{\lambda}_f)}{1 + \boldsymbol{\lambda}_f'\mathbf{W}_D\boldsymbol{\lambda}_f}$$

Remedy by Regularization

- We select $\mathbf{W}_S$ and $\mathbf{W}_D$ so that the limiting objective functions are minimized by the true ex-post risk premia vector.
- The suitable choices are $\mathbf{W}_S = \frac{v}{T}\mathbf{P}_f$ and $\mathbf{W}_D = \mathbf{P}_f$.
- Under these choices, the objective functions become

$$Q_{S,N}(\boldsymbol{\lambda}_f) = \frac{1}{N}(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f)'\mathbf{J}_N(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f) - \frac{v}{T}(\boldsymbol{\lambda}_f'\mathbf{P}_f\boldsymbol{\lambda}_f)$$

and

$$Q_{D,N}(\boldsymbol{\lambda}_f) = \frac{\frac{1}{N}(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f)'\mathbf{J}_N(\bar{\mathbf{R}} - \hat{\mathbf{B}}\boldsymbol{\lambda}_f)}{1 + \boldsymbol{\lambda}_f'\mathbf{P}_f\boldsymbol{\lambda}_f}$$

Remedy by Regularization

- Actually **any** convex combination of $Q_{S,N}(\boldsymbol{\lambda}_f)$ and $Q_{D,N}(\boldsymbol{\lambda}_f)$ could be used as the objective function resulting in an entire class of estimators.

- The limiting objective functions are

$$Q_S(\boldsymbol{\lambda}_f) = (\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*)'\boldsymbol{\Sigma}_\beta(\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*) + \frac{v}{T}$$

and

$$Q_D(\boldsymbol{\lambda}_f) = \frac{(\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*)'\boldsymbol{\Sigma}_\beta(\boldsymbol{\lambda}_f - \boldsymbol{\lambda}_f^*)}{1 + \boldsymbol{\lambda}_f'\mathbf{P}_f\boldsymbol{\lambda}_f} + \frac{v}{T}$$

- The limiting objective functions $Q_S(\boldsymbol{\lambda}_f)$ and $Q_D(\boldsymbol{\lambda}_f)$ are indeed uniquely minimized at $\boldsymbol{\lambda}_f = \boldsymbol{\lambda}_f^*$.

Proposal S Estimator

- The risk premia estimator $\hat{\boldsymbol{\lambda}}_{\mathrm{S},f}$ is given by

$$\hat{\boldsymbol{\lambda}}_{S,f} = \left(\frac{1}{N}\hat{\mathbf{B}}'\mathbf{J}_N\hat{\mathbf{B}} - \frac{\hat{v}}{T}\mathbf{P}_f\right)^{-1}\left(\frac{1}{N}\hat{\mathbf{B}}'\mathbf{J}_N\bar{\mathbf{R}}\right)$$

where

$$\hat{v} = \frac{\mathrm{tr}(\widehat{\mathbf{U}}'\widehat{\mathbf{U}}/N)}{T-(K+1)} \xrightarrow{p} v$$

and $\widehat{\mathbf{U}}$ is the matrix of OLS residuals.

- This is the estimator proposed by Shanken (RFS 1992) and studied by Raponi, Robotti, and Zaffaroni (RFS 2020) (under cross-sectional near-uncorrelatedness).

Proposal D Estimator

- The risk premia estimator $\hat{\boldsymbol{\lambda}}_{\mathrm{D},f}$ is given by

$$\hat{\boldsymbol{\lambda}}_{\mathrm{D},f} = \left(\frac{1}{N}\hat{\mathbf{B}}'\mathbf{J}_N\hat{\mathbf{B}} - \mu_*\mathbf{P}_f\right)^{-1}\left(\frac{1}{N}\hat{\mathbf{B}}'\mathbf{J}_N\bar{\mathbf{R}}\right)$$

  where μ_* is the minimum eigenvalue of the matrix

$$\mathbf{K} = (\tilde{\mathbf{C}}_f^{-1})'\left(\frac{1}{N}[\ \bar{\mathbf{R}} \quad \hat{\mathbf{B}}\ ]'\mathbf{J}_N[\ \bar{\mathbf{R}} \quad \hat{\mathbf{B}}\ ]\right)\tilde{\mathbf{C}}_f^{-1},$$

$$\tilde{\mathbf{C}}_f = \begin{bmatrix} 1 & \mathbf{0}_K' \\ \mathbf{0}_K & \mathbf{C}_f \end{bmatrix},$$

  and $\mathbf{C}_f$ is the Cholesky factor of $\mathbf{P}_f$ ($\mathbf{P}_f = \mathbf{C}_f'\mathbf{C}_f$).

- This is the estimator proposed and analyzed by Kleibergen and Zhan (WP 2024) (under cross-sectional near-uncorrelatedness).

Consistency

- Under suitable LLN assumptions on the shocks and betas, we have

$$\frac{\hat{V}}{T} \xrightarrow{p} \frac{V}{T} \text{ and } \mu_* \xrightarrow{p} \frac{V}{T}, \text{ as } N \to \infty.$$

- As a result, as $N \to \infty$, we have

$$\hat{\boldsymbol{\lambda}}_{\mathrm{S},f} \xrightarrow{p} \boldsymbol{\lambda}_f^*$$

and

$$\hat{\boldsymbol{\lambda}}_{\mathrm{D},f} \xrightarrow{p} \boldsymbol{\lambda}_f^*.$$

Asymptotic Normality

- To establish the asymptotic normality of the ex-post risk premia estimators and develop feasible test statistics for the hypotheses of interest, we need to impose some structure on the relevant zero-mean shocks.

- Let **e** is the $\mathcal{T}$-dimensional random vector defined by

$$\mathbf{e} = \left[\begin{array}{c} \frac{1}{N}\mathbf{U}'\mathbf{1}_N \\ \text{vec}\left(\frac{1}{N}\mathbf{U}'\mathbf{U} - v\mathbf{I}_T\right) \\ \text{vec}\left(\frac{1}{N}(\mathbf{B} - \mathbf{1}_N\boldsymbol{\mu}_\beta')'\mathbf{U}\right) \end{array} \right] = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \left[\begin{array}{c} \mathbf{e}_{i,1} \\ \mathbf{e}_{i,2} \\ \mathbf{e}_{i,3} \end{array} \right],$$

where $\mathcal{T} = (1 + T + K)T$.

- Assumption (cross-sectional CLT): As $N \to \infty$,

$$\sqrt{N}\mathbf{e} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \mathbf{e}_i \xrightarrow{d} \text{N}(\mathbf{0}_{\mathcal{T}}, \mathbf{V}_e).$$

Asymptotic Normality of $\hat{\boldsymbol{\lambda}}_{\mathsf{S}}$

- The asymptotic distribution of the estimator $\hat{\boldsymbol{\lambda}}_{\mathsf{S}}$ is given by

$$\sqrt{N}(\hat{\boldsymbol{\lambda}}_{\mathsf{S}} - \boldsymbol{\lambda}^*) \xrightarrow{d} \mathrm{N}(\mathbf{0}_{K+1}, \boldsymbol{\Pi}_{\mathsf{S}}' \mathbf{V}_e \boldsymbol{\Pi}_{\mathsf{S}}),$$

where

$$\boldsymbol{\Pi}_{\mathsf{S}} = \begin{bmatrix} \mathbf{g} & \mathbf{O}_{T\times K} \\ -\mathbf{M}_{\mathsf{S}}\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta & \mathbf{M}_{\mathsf{S}}\boldsymbol{\Sigma}_\beta^{-1} \\ -(\mathbf{g}\otimes\mathbf{I}_K)\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta & (\mathbf{g}\otimes\mathbf{I}_K)\boldsymbol{\Sigma}_\beta^{-1} \end{bmatrix},$$

$$\mathbf{H} = \mathbf{J}_T - \mathbf{J}_T\mathbf{F}'(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-1}\mathbf{F}\mathbf{J}_T,$$

$$\mathbf{G} = \mathbf{J}_T\mathbf{F}'(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-1},$$

$$\mathbf{g} = \frac{1}{T}\mathbf{1}_T - \mathbf{G}\boldsymbol{\lambda}_f^*,$$

$$\mathbf{M}_{\mathsf{S}} = \mathbf{g}\otimes\mathbf{G} + \frac{1}{T}\frac{1}{\mathrm{tr}(\mathbf{H})}\mathrm{vec}(\mathbf{H})(\boldsymbol{\lambda}_f^*)'\mathbf{P}_f.$$

- The matrix $\boldsymbol{\Pi}_{\mathsf{S}}$ can be estimated consistently.

Asymptotic Normality of $\hat{\boldsymbol{\lambda}}_{\mathrm{D}}$

- The asymptotic distribution of the estimator $\hat{\boldsymbol{\lambda}}_{\mathrm{D}}$ is given by

$$\sqrt{N}(\hat{\boldsymbol{\lambda}}_{\mathrm{D}} - \boldsymbol{\lambda}^*) \xrightarrow{d} \mathrm{N}(\mathbf{0}_{K+1}, \boldsymbol{\Pi}_{\mathrm{D}}' \mathbf{V}_e \boldsymbol{\Pi}_{\mathrm{D}}),$$

  where

$$\boldsymbol{\Pi}_{\mathrm{D}} = \begin{bmatrix} \mathbf{g} & \mathbf{O}_{T\times K} \\ -\mathbf{M}_{\mathrm{D}}\boldsymbol{\Sigma}_{\beta}^{-1}\boldsymbol{\mu}_{\beta} & \mathbf{M}_{\mathrm{D}}\boldsymbol{\Sigma}_{\beta}^{-1} \\ -(\mathbf{g}\otimes\mathbf{I}_K)\boldsymbol{\Sigma}_{\beta}^{-1}\boldsymbol{\mu}_{\beta} & (\mathbf{g}\otimes\mathbf{I}_K)\boldsymbol{\Sigma}_{\beta}^{-1} \end{bmatrix},$$

$$\mathbf{H} = \mathbf{J}_T - \mathbf{J}_T\mathbf{F}'(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-1}\mathbf{F}\mathbf{J}_T,$$

$$\mathbf{G} = \mathbf{J}_T\mathbf{F}'(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-1},$$

$$\mathbf{g} = \frac{1}{T}\mathbf{1}_T - \mathbf{G}\boldsymbol{\lambda}_f^*,$$

$$\mathbf{M}_{\mathrm{D}} = \mathbf{g}\otimes\mathbf{G} + \frac{1}{1+(\boldsymbol{\lambda}_f^*)'\mathbf{P}_f\boldsymbol{\lambda}_f^*}(\mathbf{g}\otimes\mathbf{g})(\boldsymbol{\lambda}_f^*)'\mathbf{P}_f.$$

- The matrix $\boldsymbol{\Pi}_{\mathrm{D}}$ can be estimated consistently.

Structure and Estimation of $\mathbf{V}_e$: Clustering

- We adapt the Clustering approach of Kim and Skoulakis (JE 2018) to the present one-shot estimation and inference context
- M_N clusters with N_m stocks in the m-th cluster, $m = 1, \ldots, M_N$, so that $\sum_{m=1}^{M_N} N_m = N$ (N_m, $m = 1, \ldots, M_N$ bounded).
- As $N \to \infty$, $\frac{N}{M_N} \to G$, where G is to be interpreted as the average cluster size.
- For $m = 1, \ldots, M_N$, let I_m be the set of all indices i for which the i-th stock belongs to the m-th cluster, and define the aggregate cluster shocks

$$\boldsymbol{\eta}_m = \begin{bmatrix} \boldsymbol{\eta}'_{m,1} & \boldsymbol{\eta}'_{m,2} & \boldsymbol{\eta}'_{m,3} \end{bmatrix}' = \sum_{i \in I_m} \mathbf{e}_i,$$

so that $\boldsymbol{\eta}_{m,k} = \sum_{i \in I_m} \mathbf{e}_{i,k}$, for $k = 1, 2, 3$.

Structure and Estimation of $\mathbf{V}_e$: Clustering

- Under the Clustering structure, we have $\mathbf{V}_e = \frac{1}{G}\mathbf{V}_\eta$ where

$$\mathbf{V}_\eta = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{V}_{13} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{V}_{23} \\ \mathbf{V}_{31} & \mathbf{V}_{32} & \mathbf{V}_{33} \end{bmatrix},$$

with

$$\mathbf{V}_{k\ell} = \text{p-}\lim \frac{1}{M_N} \sum_{m=1}^{M_N} \boldsymbol{\eta}_{m,k} \boldsymbol{\eta}'_{m,\ell}, \quad k, \ell = 1, 2, 3.$$

Structure and Estimation of $\mathbf{V}_e$: Clustering

- To capture all possible interactions across clusters, we make a number of assumptions involving cross-sectional moments
- Under these assumptions, we have

$$\mathbf{V}_{11} = v_c \mathbf{I}_T, \quad \mathbf{V}_{21} = \xi_c \boldsymbol{\mathcal{M}},$$
$$\mathbf{V}_{22} = \kappa_c^{[1]} \mathcal{K}^{[1]} + \kappa_c^{[2]} \mathcal{K}^{[2]} + \kappa_c^{[3]} \mathcal{K}^{[3]} - (2G^2 - H) v^2 \mathbf{i}_T \mathbf{i}_T',$$
$$\mathbf{V}_{31} = \mathbf{I}_T \otimes \mathbf{c}_{31}, \quad \mathbf{V}_{32} = \boldsymbol{\mathcal{M}}' \otimes \mathbf{c}_{32}, \quad \mathbf{V}_{33} = \mathbf{I}_T \otimes \mathbf{C}_{33},$$

where $\frac{N}{M_N} \to G$, $\frac{1}{M_N} \sum_{m=1}^{M_N} N_m^2 \to H$, $\mathbf{i}_T = \mathrm{vec}(\mathbf{I}_T)$, v_c is a second-moment parameter, ξ_c is a third-moment parameter, $\kappa_c^{[k]}$, $k = 1, 2, 3$ are fourth-moment parameters (involving only shocks $u_{i,t}$), $\mathbf{c}_{31}$ and $\mathbf{c}_{32}$ are $K \times 1$ vectors of fourth-moment parameters (involving $u_{i,t}$ and $\boldsymbol{\beta}_i$), $\mathbf{C}_{33}$ is a $K \times K$ matrix of fourth-moment parameters (involving $u_{i,t}$ and $\boldsymbol{\beta}_i$), and $\mathcal{K}^{[k]}$, $k = 1, 2, 3$ and $\boldsymbol{\mathcal{M}}$ are suitable selection matrices.

Structure and Estimation of $\mathbf{V}_e$: Clustering

- Consistent estimators $\hat{v}_c$, $\hat{\xi}_c$, $\hat{\kappa}_c^{[k]}$, $k = 1, 2, 3$, $\hat{\mathbf{c}}_{31}$, $\hat{\mathbf{c}}_{32}$, $\hat{\mathbf{C}}_{33}$ are developed.

- We can then develop feasible tests for asset pricing model implications
  - zero CSR intercept: $\lambda_0^* = 0$
  - for traded factors: $\boldsymbol{\lambda}_f^* = \bar{\mathbf{f}}$
  - average squared pricing error:

$$\frac{1}{\sqrt{N}}\left(\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\lambda}})'\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\lambda}}) - Q\right) \xrightarrow{d} \mathrm{N}\left(0, \boldsymbol{\pi}_\alpha' \mathbf{V}_e \boldsymbol{\pi}_\alpha\right)$$

  where Q and $\boldsymbol{\pi}_\alpha$ can also be estimated consistently.

Conclusion

- Regularization is used to motivate and develop a class of consistent ex-post risk premia estimators in the small T-large N context
- Clustering is used to realistically model the cross-sectional interactions of the various shocks involved
- Long To-Do list
  - Comparison of the estimators in terms of asymptotic efficiency
  - Monte Carlo evidence on finite-sample performance
  - Revisit empirical evidence in a clustering-robust setting