Joint Estimation of Conditional Mean and Covariance for Unbalanced Panels

Damir Filipović Paul Schneider

EPFL, Swiss Finance Institute, USI Lugano

Research Seminar
WU Vienna, 19 March 2025

Overview

- Most asset pricing models build on conditional first two moments
- **Finance millenium problem** (Cochrane's 2011 address): "How to model conditional covariance as function of characteristics?"
- We provide a nonparametric consistent joint estimator of conditional mean and covariance for unbalanced panels as function of characteristics
- Satisfies asymptotic consistency and finite-sample guarantees
- Implies a conditional factor model representation
- Achieves maximal possible Sharpe ratio, i.e., "spans the stochastic discount factor" (Kozak and Nagel (2024))

Outline

Conditional mean and covariance functions

- Consider discrete time periods $t = 0, 1, \dots$ (e.g., months)
- Over period $[t, t+1]$ there are N_t assets with excess returns $x_{t+1,i}$
- Asset i is characterized by **covariates** $z_{t,i}$ in some covariate space $\mathcal{Z}$ observable at t
- **Goal**: model conditional moments given information at t,

$$\mathbb{E}_t[x_{t+1,i}] = \mu(z_{t,i}),$$
$$\mathbb{E}_t[x_{t+1,i}x_{t+1,j}] = q(z_{t,i}, z_{t,j}),$$

by **conditional moment functions** $\mu : \mathcal{Z} \to \mathbb{R}$ and $q : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$

Challenge

- Denote arrays

$$\boldsymbol{x}_{t+1} := [x_{t+1,1}, \ldots, x_{t+1,N_t}]^\top \in \mathbb{R}^{N_t}, \qquad \boldsymbol{z}_t := [z_{t,1}, \ldots, z_{t,N_t}]^\top \in \mathcal{Z}^{N_t}$$

and corresponding arrays of function values

$$\mu(\boldsymbol{z}_t) := [\mu(z_{t,i}) : 1 \le i \le N_t], \qquad q(\boldsymbol{z}_t, \boldsymbol{z}_t^\top) := [q(z_{t,i}, z_{t,j}) : 1 \le i, j \le N_t]$$

- **Challenge**: find functions μ and q such that implied conditional covariance matrix

$$q(\boldsymbol{z}_t, \boldsymbol{z}_t^\top) - \mu(\boldsymbol{z}_t)\mu(\boldsymbol{z}_t)^\top \text{ is symmetric and positive semidefinite.} \tag{1}$$

- Property (1) implies that $q(\boldsymbol{z}_t, \boldsymbol{z}_t^\top)$ is symmetric and positive semidefinite
- This is the defining property of a **kernel function**

Kernel functions and Schur complement

- Extended covariate spac $\mathcal{Z}_\Delta := \mathcal{Z} \cup \{\Delta\}$ for external point $\Delta \notin \mathcal{Z}$
- Assume $q : \mathcal{Z}_\Delta \times \mathcal{Z}_\Delta \to \mathbb{R}$ is a kernel function such that $q(\Delta, \Delta) = 1$
- Define $\mu(z) := q(z, \Delta)$
- Then the implied covariance function

$$c(z, z') := q(z, z') - \mu(z)\mu(z')$$

  is the Schur complement of q with respect to Δ and therefore a kernel function on $\mathcal{Z}_\Delta \times \mathcal{Z}_\Delta$
- Problem boils down to specify q

Systemic and idiosyncratic components

- Decompose $q(z,z') = q^{\mathrm{sy}}(z,z') + q^{\mathrm{id}}(z,z')$ into sum of two kernel functions:
- Systematic component $q^{\mathrm{sy}}(z,z')$ captures
  - conditional cross-sectional dependence, and
  - risk premium (conditional mean), and hence structural condition $q^{\mathrm{sy}}(\Delta,\Delta) = 1$
- Idiosyncratic component $q^{\mathrm{id}}(z,z') = q^{\mathrm{id}}(z,z')1_{z=z'}$ is supported on the diagonal of $\mathcal{Z} \times \mathcal{Z}$

Moment kernel specification

- **Need**: flexible nonparametric specification of moment kernel q on $\mathcal{Z}_\Delta \times \mathcal{Z}_\Delta$
- **Approach**: let $\mathcal{C} = \ell^2$ be auxiliary Hilbert space, fix unit vector $p \in \mathcal{C}$, i.e., $\langle p, p \rangle_{\mathcal{C}} = 1$
- Any **feature maps** $h^{\mathrm{sy}}, h^{\mathrm{id}} : \mathcal{Z} \to \mathcal{C}$ define a desired kernel function on $\mathcal{Z}_\Delta \times \mathcal{Z}_\Delta$ by

$$q_h(z, z') := \underbrace{\langle h^{\mathrm{sy}}(z) + p 1_{z=\Delta}, h^{\mathrm{sy}}(z') + p 1_{z'=\Delta} \rangle_{\mathcal{C}}}_{\text{systematic component } q_h^{\mathrm{sy}}(z,z')} + \underbrace{\| h^{\mathrm{id}}(z) \|_{\mathcal{C}}^2 \, 1_{z=z'}}_{\text{idiosyncratic component } q_h^{\mathrm{id}}(z,z')}$$

where we **extend** $h^\tau(\Delta) := 0$, for $\tau \in \{\mathrm{sy}, \mathrm{id}\}$

- This implies the conditional mean and covariance functions

$$\mu_h(z) = \langle h^{\mathrm{sy}}(z), p \rangle_{\mathcal{C}},$$
$$c_h(z, z') = \langle h^{\mathrm{sy}}(z), h^{\mathrm{sy}}(z') \rangle_{\mathcal{C}} - \langle h^{\mathrm{sy}}(z), p \rangle_{\mathcal{C}} \langle h^{\mathrm{sy}}(z'), p \rangle_{\mathcal{C}} + \| h^{\mathrm{id}}(z) \|_{\mathcal{C}}^2 \, 1_{z=z'}.$$

Outline

Estimation problem

- Estimate h via matrix-valued regression

$$\begin{bmatrix} 1 & \boldsymbol{x}_{t+1}^\top \\ \boldsymbol{x}_{t+1} & \boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}^\top \end{bmatrix} = \begin{bmatrix} 1 & \langle p, h^{\mathrm{sy}}(\boldsymbol{z}_t)\rangle_{\mathcal{C}} \\ \langle h^{\mathrm{sy}}(\boldsymbol{z}_t), p\rangle_{\mathcal{C}} & \langle h^{\mathrm{sy}}(\boldsymbol{z}_t), h^{\mathrm{sy}}(\boldsymbol{z}_t)\rangle_{\mathcal{C}} \end{bmatrix} + \begin{bmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & \mathrm{diag}(\|h^{\mathrm{id}}(\boldsymbol{z}_t)\|_{\mathcal{C}}^2) \end{bmatrix} + \boldsymbol{E}_{t+1}$$

  with residual matrix $\mathbb{E}_t[\boldsymbol{E}_{t+1}] = 0$

- Denote **data point** $\xi_t := (N_t, \boldsymbol{x}_{t+1}, \boldsymbol{z}_t)$ and weight function $w(N_t) := 1/N_t$

- Leads to **quadratic loss function** for $h = (h^{\mathrm{sy}}, h^{\mathrm{id}})$,

$$\begin{aligned} \mathcal{L}(h, \xi_t) &:= w(N_t)\, \|\boldsymbol{E}_{t+1}\|_F^2 \\ &= 2\, \underbrace{w(N_t)\, \|\boldsymbol{x}_{t+1} - \langle h^{\mathrm{sy}}(\boldsymbol{z}_t), p\rangle_{\mathcal{C}}\|_2^2}_{\text{first moment error}} \\ &\quad + \underbrace{w(N_t)\, \left\|\boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}^\top - \langle h^{\mathrm{sy}}(\boldsymbol{z}_t), h^{\mathrm{sy}}(\boldsymbol{z}_t)^\top\rangle_{\mathcal{C}} - \mathrm{diag}(\|h^{\mathrm{id}}(\boldsymbol{z}_t)\|_{\mathcal{C}}^2)\right\|_F^2}_{\text{second moment error}}, \end{aligned}$$

Hypothesis space for $h = (h^{\mathrm{sy}}, h^{\mathrm{id}})$

- **Need**: a hypothesis space for feature maps $h^{\mathrm{sy}}, h^{\mathrm{id}} : \mathcal{Z} \to \mathcal{C}$
- **Flexible choice**: $\mathcal{C}$-valued **reproducing kernel Hilbert spaces (RKHS)** $\mathcal{H}^{\mathrm{sy}}$, $\mathcal{H}^{\mathrm{id}}$ of functions $h^{\mathrm{sy}}, h^{\mathrm{id}} : \mathcal{Z} \to \mathcal{C}$ with operator-valued reproducing kernels $K^{\mathrm{sy}}, K^{\mathrm{id}}$ on $\mathcal{Z} \times \mathcal{Z}$
- **Tractable choice**: assume separable $K^{\mathrm{sy}}(z, z') = k^{\mathrm{sy}}(z, z') I_{\mathcal{C}}$, $K^{\mathrm{id}}(z, z') = k^{\mathrm{id}}(z, z') I_{\mathcal{C}}$ for scalar reproducing kernels $k^{\mathrm{sy}}, k^{\mathrm{id}}$ of separable RKHS $\mathcal{G}^{\mathrm{sy}}, \mathcal{G}^{\mathrm{id}}$ on $\mathcal{Z}$
- Note: $\mathcal{H}^{\mathrm{sy}} \cong \mathcal{G}^{\mathrm{sy}} \otimes \mathcal{C}$, $\mathcal{H}^{\mathrm{id}} \cong \mathcal{G}^{\mathrm{id}} \otimes \mathcal{C}$ can be identified with tensor products
- This is a fully flexible nonparametric setup

Non-convex kernel ridge regression

- Regularize loss function $\mathcal{L}(h, \xi_t)$ with parameters $\lambda^{\mathrm{sy}}, \lambda^{\mathrm{id}} > 0$,

$$\mathcal{R}(h, \xi_t) := \mathcal{L}(h, \xi_t) + \underbrace{\lambda^{\mathrm{sy}} \| h^{\mathrm{sy}} \|^2_{\mathcal{H}^{\mathrm{sy}}} + \lambda^{\mathrm{id}} \| h^{\mathrm{id}} \|^2_{\mathcal{H}^{\mathrm{id}}}}_{\text{regularization}}$$

- Sample average: empirical loss minimization $\Rightarrow$ kernel ridge regression problem

$$\boxed{\underset{h \in \mathcal{H}^{\mathrm{sy}} \times \mathcal{H}^{\mathrm{id}}}{\text{minimize}} \ \frac{1}{T} \sum_{t=0}^{T-1} \mathcal{R}(h, \xi_t)} \qquad (2)$$

- Problem (2) is not convex in $h = (h^{\mathrm{sy}}, h^{\mathrm{id}})$, there exist infinitely many minimizers h

- However, we can characterize their structure explicitly:

Lemma 2.1 (Representer theorem).

Every minimizer of (2) is of the form $h^\tau(\cdot) = \sum_{t=0}^{T-1} \sum_{i=1}^{N_t} k^\tau(\cdot, z_{t,i}) \gamma^\tau_{t,i}$ *for some* $\gamma^\tau_{t,i} \in \mathcal{C}$, *for both components* $\tau \in \{\mathrm{sy}, \mathrm{id}\}$.

Low-rank approximation: feature learning

- **Problem**: full sample $\boldsymbol{Z} := \begin{bmatrix} \boldsymbol{z}_0 \\ \vdots \\ \boldsymbol{z}_{T-1} \end{bmatrix} \in \mathcal{Z}^N$ for $N := \sum_{t=0}^{T-1} N_t$ may be too large
- **Nyström method** selects subsample $\Pi^\tau \subset \{1, \dots, N\}$ of size $m^\tau \leq N$ such that kernel gap

$$\operatorname{tr}\left(k^\tau(\boldsymbol{Z}, \boldsymbol{Z}^\top) - k^\tau(\boldsymbol{Z}, \boldsymbol{Z}_{\Pi^\tau}^\top) k^\tau(\boldsymbol{Z}_{\Pi^\tau}, \boldsymbol{Z}_{\Pi^\tau}^\top)^{-1} k^\tau(\boldsymbol{Z}_{\Pi^\tau}, \boldsymbol{Z}^\top)\right) \leq \epsilon_{\text{tolerance}}$$

- Gives m^τ linearly independent **feature maps** $\boldsymbol{\phi}^\tau(\cdot) := [\phi_1^\tau(\cdot), \dots, \phi_{m^\tau}^\tau(\cdot)] := k^\tau(\cdot, \boldsymbol{Z}_{\Pi^\tau}^\top)$
- **Low-rank approximation**: restrict to subspace of $\mathcal{H}^\tau$ consisting of

$$h^\tau(\cdot) = \sum_{i=1}^{m^\tau} \phi_i^\tau(\cdot)\gamma_i^\tau =: \boldsymbol{\phi}^\tau(\cdot)\boldsymbol{\gamma}^\tau$$

for coefficients $\boldsymbol{\gamma}^\tau \in \mathcal{C}^{m^\tau}$, … for both components $\tau \in \{\mathrm{sy}, \mathrm{id}\}$

Reparametrization I

Reparametrize loss function $\mathcal{R}(h,\xi_t)$ in terms of coefficients $\gamma=(\gamma^{\mathrm{sy}},\gamma^{\mathrm{id}})\in\mathcal{C}^{m^{\mathrm{sy}}}\times\mathcal{C}^{m^{\mathrm{id}}}$

$$
\begin{aligned}
\mathcal{R}(\gamma,\xi_t) := w(N_t)\Bigg\| \begin{bmatrix} 1 & \boldsymbol{x}_{t+1}^\top \\ \boldsymbol{x}_{t+1} & \boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}^\top \end{bmatrix} - \boldsymbol{\Psi}^{\mathrm{sy}}(\boldsymbol{z}_t)\boldsymbol{U}^{\mathrm{sy}}(\gamma^{\mathrm{sy}})\boldsymbol{\Psi}^{\mathrm{sy}}(\boldsymbol{z}_t)^\top \\
- \mathrm{Diag}(\boldsymbol{\Psi}^{\mathrm{id}}(\boldsymbol{z}_t)\boldsymbol{U}^{\mathrm{id}}(\gamma^{\mathrm{id}})\boldsymbol{\Psi}^{\mathrm{id}}(\boldsymbol{z}_t)^\top)\Bigg\|_F^2 \\
+ \lambda^{\mathrm{sy}}\mathrm{tr}(\boldsymbol{G}^{\mathrm{sy}}\boldsymbol{U}^{\mathrm{sy}}(\gamma^{\mathrm{sy}})) + \lambda^{\mathrm{id}}\mathrm{tr}(\boldsymbol{G}^{\mathrm{id}}\boldsymbol{U}^{\mathrm{id}}(\gamma^{\mathrm{id}})),
\end{aligned}
$$

for the matrices

$$
\boldsymbol{U}^{\mathrm{sy}}(\gamma^{\mathrm{sy}}) := \begin{bmatrix} 1 & \langle p,\gamma^{\mathrm{sy}\top}\rangle_{\mathcal{C}} \\ \langle p,\gamma^{\mathrm{sy}}\rangle_{\mathcal{C}} & \langle \gamma^{\mathrm{sy}},\gamma^{\mathrm{sy}\top}\rangle_{\mathcal{C}} \end{bmatrix} \in \mathbb{S}_+^{m^{\mathrm{sy}}+1}, \quad \boldsymbol{U}^{\mathrm{id}}(\gamma^{\mathrm{id}}) := \langle \gamma^{\mathrm{id}},\gamma^{\mathrm{id}\top}\rangle_{\mathcal{C}} \in \mathbb{S}_+^{m^{\mathrm{id}}},
$$

$$
\boldsymbol{\Psi}^{\mathrm{sy}}(\boldsymbol{z}_t) := \begin{bmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \phi^{\mathrm{sy}}(\boldsymbol{z}_t) \end{bmatrix} \in \mathbb{R}^{(N_t+1)\times(m^{\mathrm{sy}}+1)}, \quad \boldsymbol{\Psi}^{\mathrm{id}}(\boldsymbol{z}_t) := \begin{bmatrix} \boldsymbol{0}^\top \\ \phi^{\mathrm{id}}(\boldsymbol{z}_t) \end{bmatrix} \in \mathbb{R}^{(N_t+1)\times m^{\mathrm{id}}},
$$

$$
\boldsymbol{G}^{\mathrm{sy}} := \begin{bmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \langle \phi^{\mathrm{sy}\top},\phi^{\mathrm{sy}}\rangle_{\mathcal{G}^{\mathrm{sy}}} \end{bmatrix} \in \mathbb{S}_+^{m^{\mathrm{sy}}+1}, \quad \boldsymbol{G}^{\mathrm{id}} := \langle \phi^{\mathrm{id}\top},\phi^{\mathrm{id}}\rangle_{\mathcal{G}^{\mathrm{id}}} \in \mathbb{S}_+^{m^{\mathrm{id}}}.
$$

Reparametrization II gives convex problem

- Define convex **feasible set** $\mathcal{D} := \left\{ \boldsymbol{U} = (\boldsymbol{U}^{\mathrm{sy}}, \boldsymbol{U}^{\mathrm{id}}) \in \mathbb{S}_+^{m^{\mathrm{sy}}+1} \times \mathbb{S}_+^{m^{\mathrm{id}}} : \boldsymbol{U}_{11}^{\mathrm{sy}} = 1 \right\}$
- **Lemma**: $(\gamma^{\mathrm{sy}}, \gamma^{\mathrm{id}}) \mapsto (\boldsymbol{U}^{\mathrm{sy}}(\gamma^{\mathrm{sy}}), \boldsymbol{U}^{\mathrm{id}}(\gamma^{\mathrm{id}})) : \mathcal{C}^{m^{\mathrm{sy}}} \times \mathcal{C}^{m^{\mathrm{id}}} \to \mathcal{D}$ is surjective

$\Rightarrow$ Can reparametrize: gives **convex** (quadratic) loss function in terms of $\boldsymbol{U} \in \mathcal{D}$

$$\mathcal{R}(\boldsymbol{U}, \xi_t) := w(N_t) \left\| \begin{bmatrix} 1 & \boldsymbol{x}_{t+1}^\top \\ \boldsymbol{x}_{t+1} & \boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}^\top \end{bmatrix} - \boldsymbol{\Psi}^{\mathrm{sy}}(\boldsymbol{z}_t)\boldsymbol{U}^{\mathrm{sy}}\boldsymbol{\Psi}^{\mathrm{sy}}(\boldsymbol{z}_t)^\top - \mathrm{Diag}(\boldsymbol{\Psi}^{\mathrm{id}}(\boldsymbol{z}_t)\boldsymbol{U}^{\mathrm{id}}\boldsymbol{\Psi}^{\mathrm{id}}(\boldsymbol{z}_t)^\top) \right\|_F^2 + \lambda^{\mathrm{sy}}\mathrm{tr}(\boldsymbol{G}^{\mathrm{sy}}\boldsymbol{U}^{\mathrm{sy}}) + \lambda^{\mathrm{id}}\mathrm{tr}(\boldsymbol{G}^{\mathrm{id}}\boldsymbol{U}^{\mathrm{id}})$$

- Estimation boils down to constrained convex optimization problem over $\boldsymbol{U} \in \mathcal{D}$

Moment kernel estimator in reparametrization

- Estimator of moment kernel in terms of $\boldsymbol{U} = (\boldsymbol{U}^{\mathrm{sy}}, \boldsymbol{U}^{\mathrm{id}}) \in \mathcal{D}$ given by

$$q_{\boldsymbol{U}}(z,z') = \begin{bmatrix} 1_{z=\Delta} & \phi^{\mathrm{sy}}(z) \end{bmatrix} \boldsymbol{U}^{\mathrm{sy}} \begin{bmatrix} 1_{z'=\Delta} & \phi^{\mathrm{sy}}(z') \end{bmatrix}^\top + \phi^{\mathrm{id}}(z)\boldsymbol{U}^{\mathrm{id}}\phi^{\mathrm{id}}(z')^\top 1_{z=z'}.$$

- This implies the conditional mean and covariance functions

$$\begin{aligned} \mu_{\boldsymbol{U}}(z) &= \phi^{\mathrm{sy}}(z)\boldsymbol{b}, \\ c_{\boldsymbol{U}}(z,z') &= \phi^{\mathrm{sy}}(z)\big(\boldsymbol{V} - \boldsymbol{b}\boldsymbol{b}^\top\big)\phi^{\mathrm{sy}}(z')^\top + \phi^{\mathrm{id}}(z)\boldsymbol{U}^{\mathrm{id}}\phi^{\mathrm{id}}(z')^\top 1_{z=z'}, \end{aligned}$$

for $\begin{bmatrix} 1 & \boldsymbol{b}^\top \\ \boldsymbol{b} & \boldsymbol{V} \end{bmatrix} := \boldsymbol{U}^{\mathrm{sy}}$

Resulting conditional mean-covariance estimator

- Estimator of conditional mean and covariance matrix in terms of $\begin{bmatrix} 1 & \boldsymbol{b}^\top \\ \boldsymbol{b} & \boldsymbol{V} \end{bmatrix} := \boldsymbol{U}^{\mathrm{sy}}$ given by

$$\boldsymbol{\mu}_t = \phi^{\mathrm{sy}}(\boldsymbol{z}_t)\boldsymbol{b},$$
$$\boldsymbol{\Sigma}_t = \underbrace{\phi^{\mathrm{sy}}(\boldsymbol{z}_t)\big(\boldsymbol{V} - \boldsymbol{b}\boldsymbol{b}^\top\big)\phi^{\mathrm{sy}}(\boldsymbol{z}_t)^\top}_{=:\boldsymbol{\Sigma}_t^{\mathrm{sy}}} + \underbrace{\mathrm{Diag}(\phi^{\mathrm{id}}(\boldsymbol{z}_t)\boldsymbol{U}^{\mathrm{id}}\phi^{\mathrm{id}}(\boldsymbol{z}_t)^\top)}_{=:\boldsymbol{\Sigma}_t^{\mathrm{id}}}$$

$\Rightarrow$ No-arbitrage $\boldsymbol{\mu}_t \in \mathrm{Im}(\boldsymbol{\Sigma}_t)$ holds if either $\boldsymbol{V} - \boldsymbol{b}\boldsymbol{b}^\top$ or $\boldsymbol{\Sigma}_t^{\mathrm{id}}$ is invertible

Example (Isotropic idiosyncratic specification).

Dimension $m^{\mathrm{id}} = 1$, *constant feature map* $\phi^{\mathrm{id}}(\cdot) = \phi_1^{\mathrm{id}}(\cdot) := 1$, *and* $\boldsymbol{U}^{\mathrm{id}} = u^{\mathrm{id}} \in [0, \infty)$. *The idiosyncratic component becomes isotropic* $\boldsymbol{\Sigma}_t^{\mathrm{id}} = u^{\mathrm{id}} \boldsymbol{I}_{N_t}$.

Spanning factor model representation

Theorem.

Assume $\boldsymbol{\Sigma}_t^{\mathrm{id}}$ *is invertible. Then the* m^{sy} *(GLS) factor portfolios*

$$\boldsymbol{f}_{t+1} := \left((\boldsymbol{\Sigma}_t^{\mathrm{id}})^{-1/2}\phi^{\mathrm{sy}}(\boldsymbol{z}_t)\right)^{+}(\boldsymbol{\Sigma}_t^{\mathrm{id}})^{-1/2}\boldsymbol{x}_{t+1}$$

are conditionally uncorrelated with the residuals $\epsilon_{t+1} := \boldsymbol{x}_{t+1} - \phi^{\mathrm{sy}}(\boldsymbol{z}_t)\boldsymbol{f}_{t+1}$, *i.e.,*

$$\boldsymbol{x}_{t+1} = \phi^{\mathrm{sy}}(\boldsymbol{z}_t)\boldsymbol{f}_{t+1} + \epsilon_{t+1},$$

and span the conditionally mean-variance efficient (cMVE) portfolio with weights $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t$.

Example (Isotropic idiosyncratic specification).

In this case, $\boldsymbol{f}_{t+1} = \phi^{\mathrm{sy}}(\boldsymbol{z}_t)^{+}\boldsymbol{x}_{t+1}$ *are simply the OLS factors.*

Outline

Vectorization

- Assume sample $\xi_t = (N_t, \boldsymbol{x}_{t+1}, \boldsymbol{z}_t) \sim \xi = (N, \boldsymbol{x}, \boldsymbol{z})$ is i.i.d. for $t = 0, 1, \dots, T-1$
- Vectorize parameter $\boldsymbol{u} := \begin{bmatrix} \mathrm{vech}(\boldsymbol{U}^{\mathrm{sy}}) \\ \mathrm{vech}(\boldsymbol{U}^{\mathrm{id}}) \end{bmatrix}$ and feasible set $\mathcal{U} := \mathrm{vech}\,\mathcal{D}$
- Can express quadratic loss function in terms of $\boldsymbol{u} \in \mathcal{U}$

$$\boxed{\mathcal{R}(\boldsymbol{u}, \xi) := \frac{1}{2}\boldsymbol{u}^\top \boldsymbol{A}(\xi)\boldsymbol{u} + \boldsymbol{b}(\xi)^\top \boldsymbol{u} + c(\xi)}$$

- Denote **population loss** $\mathcal{E}(\boldsymbol{u}) := \mathbb{E}[\mathcal{R}(\boldsymbol{u}, \xi)] = \frac{1}{2}\boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{u} + \boldsymbol{b}^\top \boldsymbol{u} + c$
- Denote **empirical (sample average) loss** $\mathcal{R}_T(\boldsymbol{u}) := \frac{1}{T}\sum_{t=0}^{T-1} \mathcal{R}(\boldsymbol{u}, \xi_t)$
- How do empirical minimizers of $\mathcal{R}_T(\boldsymbol{u})$ compare to population minimizer of $\mathcal{E}(\boldsymbol{u})$?

Consistency

- Assume that the following moments are finite

$$\mathbb{E}[w(N)\|\phi^{\mathrm{sy}}(\boldsymbol{z})\|_F^4], \quad \mathbb{E}[w(N)\|\phi^{\mathrm{id}}(\boldsymbol{z})\|_F^4], \quad \mathbb{E}[w(N)\|\boldsymbol{x}\|_2^4] < \infty. \tag{3}$$

Theorem (Consistency).

Assume that $\boldsymbol{A}$ is non-singular, so that $\mathcal{E}$ is strictly convex and there exists a unique minimizer $\boldsymbol{u}^* := \arg\min_{\boldsymbol{u}\in\mathcal{U}} \mathcal{E}(\boldsymbol{u})$. Then any sequence of minimizers $\boldsymbol{u}_T^* \in \arg\min_{\boldsymbol{u}\in\mathcal{U}} \mathcal{R}_T(\boldsymbol{u})$ converges, $\boldsymbol{u}_T^* \to \boldsymbol{u}^*$ as $T \to \infty$, with probability 1.

Mean squared error bound

- A function $f(\boldsymbol{u})$ is α-**strongly convex** if $f(\boldsymbol{u}) - \frac{\alpha}{2}\|\boldsymbol{u}\|_2^2$ is convex.

Theorem (Mean squared error bound).

Assume further that $\mathcal{R}(\boldsymbol{u}, \xi)$ is α-strongly convex in $\boldsymbol{u}$ for $\mathbb{P}$-a.e. ξ, for some $\alpha > 0$, and

$$\mathbb{E}[\|(\boldsymbol{A}(\xi) - \boldsymbol{A})\boldsymbol{u}^* + \boldsymbol{b}(\xi) - \boldsymbol{b}\|_2^2] \leq \sigma^2, \tag{4}$$

for some $\sigma > 0$. Then $\mathcal{E}$ and $\mathcal{R}_T$ are α-strongly convex, so that the minimizers $\boldsymbol{u}_T^* = \arg\min_{\boldsymbol{u} \in \mathcal{U}} \mathcal{R}_T(\boldsymbol{u})$ are unique, and

$$\mathbb{E}[\|\boldsymbol{u}_T^* - \boldsymbol{u}^*\|_2^2] \leq \sigma^2/(\alpha^2 T).$$

Finite-sample guarantees

Theorem (Finite-sample guarantees).

Assume further that

$$\mathbb{E}[\exp(\tau^{-2}\|(\boldsymbol{A}(\xi)-\boldsymbol{A})\boldsymbol{u}^* + \boldsymbol{b}(\xi)-\boldsymbol{b}\|_2^2)] \le \exp(1), \tag{5}$$

for some $\tau > 0$. *Then for all* $\epsilon > 0$, $\mathbb{P}[\|\boldsymbol{u}_T^* - \boldsymbol{u}^*\|_2 \ge \epsilon] \le 2\exp(-\tau^{-2}T\epsilon^2\alpha^2/3)$. *This can equivalently be expressed as: for any* $\delta \in (0,1)$, *with sample probability of at least* $1-\delta$, *it holds that*

$$\|\boldsymbol{u}_T^* - \boldsymbol{u}^*\|_2 \le \sqrt{\log(2/\delta)}\sqrt{3}\tau/(\alpha\sqrt{T}).$$

- Condition (5) implies (4) for $\sigma^2 = \tau^2$.
- A sufficient condition for (5) is that ϕ^{sy} and ϕ^{id} are uniformly bounded on $\mathcal{Z}$, and individual returns x_i and $N^2 w(N)$ are uniformly bounded, $\mathbb{P}$-a.s.

Outline

Data

- Sample period 1962–2021
- Returns from CRSP for firms in NYSE, AMEX, and NASDAQ
- Macroeconomic predictors from Welch and Goyal (2008)
- Updated sample from Gu, Kelly, Xiu (2020)
- In total over 100 characteristics/macro factors
- Over 6000 firms on average per month
- Rolling out-of-sample (OOS) testing: **training sample** 96 months + **validation sample** 1 month + **test sample** 1 month
- Plot expanding or rolling 24-month averages

Specification and benchmarking

- **Specification**: cosine kernel $k^{cos}(z, z') = \frac{\langle z, z' \rangle_2}{\|z\|_2 \, \|z'\|_2}$
- Number of factors $m = 5, 10, 20, 40$
- Statistical **scoring rule** Dawid and Sebastiani (1999) to jointly benchmark first and second moments

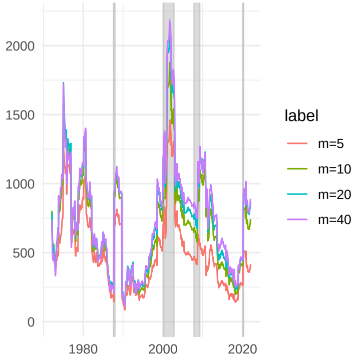
$$\mathcal{S}_c(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) := \log\det \boldsymbol{\Sigma} + (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$$

- Validation of hyperparameters by minimizing scoring rule
- **Benchmark**: purely idiosyncratic model $\boldsymbol{\mu}_t^{\text{bm}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_t^{\text{bm}} = \sigma_{\text{bm}}^2 \boldsymbol{I}_{N_t}$ estimated by minimizing loss function: $\sigma_{\text{bm}}^2 = \frac{\sum_{t=0}^{T-1} w(N_t) \|\boldsymbol{x}_{t+1}\|_2^2}{\sum_{t=0}^{T-1} w(N_t) N_t}$
- Statistical benchmarking by rolling difference ($T - t = 24$ months)

$$\frac{1}{T-t} \sum_{s=t}^{T-1} \left(\mathcal{S}_c(\boldsymbol{x}_{s+1}, \mathbf{0}, \sigma_{\text{bm}}^2 \boldsymbol{I}_{N_s}) - \mathcal{S}_c(\boldsymbol{x}_{s+1}, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) \right)$$

OOS scoring rule

- Scoring difference $\frac{1}{T-t}\sum_{s=t}^{T-1}\left(\mathcal{S}_c(\boldsymbol{x}_{s+1}, \boldsymbol{0}, \sigma_{bm}^2 \boldsymbol{I}_{N_s}) - \mathcal{S}_c(\boldsymbol{x}_{s+1}, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)\right)$

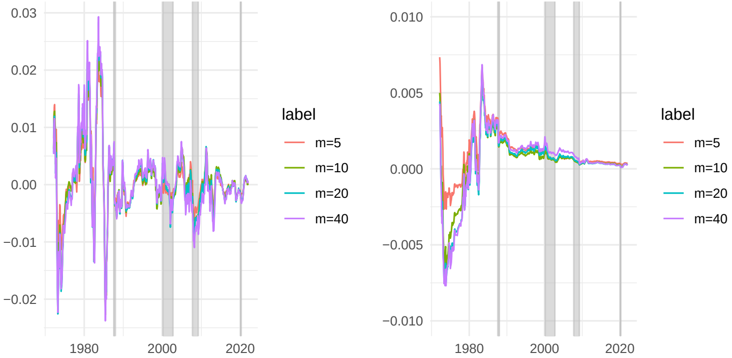


- Significant outperformance during **major market crashes** (shaded areas): 1987 Crash, Dot-Com Bubble, Global Financial Crisis, COVID-19 Pandemic

OOS first moment prediction

$$R^2_{t,T,\mathrm{OOS}} := 1 - \frac{\sum_{s=t}^{T-1} w(N_s)\|\boldsymbol{x}_{s+1} - \phi^{\mathrm{sy}}(\boldsymbol{z}_s)\boldsymbol{b}\|_2^2}{\sum_{s=t}^{T-1} w(N_s)\|\boldsymbol{x}_{s+1}\|_2^2},$$

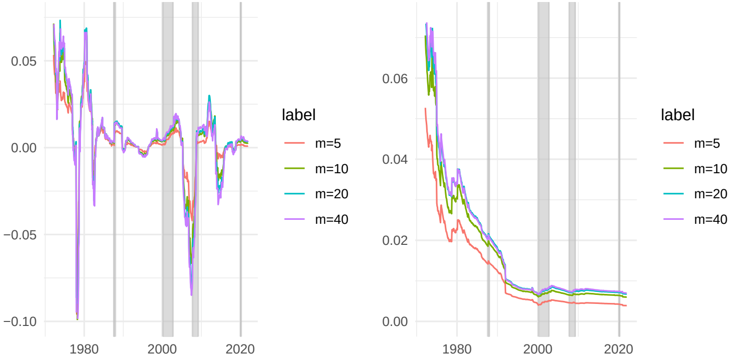
- Left panel predictive rolling. Right panel expanding



OOS second moment prediction

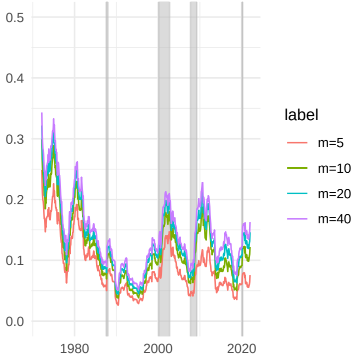
$$R^{2,2}_{t,T,\text{OOS}} := 1 - \frac{\sum_{s=t}^{T-1} w(N_s)\|\boldsymbol{x}_{s+1}\boldsymbol{x}_{s+1}^\top - \phi^{\text{sy}}(\boldsymbol{z}_s)\boldsymbol{V}\phi^{\text{sy}}(\boldsymbol{z}_s)^\top - u^{\text{id}}\boldsymbol{I}_{N_s}\|_F^2}{\sum_{s=t}^{T-1} w(N_s)\|\boldsymbol{x}_{s+1}\boldsymbol{x}_{s+1}^\top - \sigma^2_{\text{bm}}\boldsymbol{I}_{N_s}\|_F^2},$$

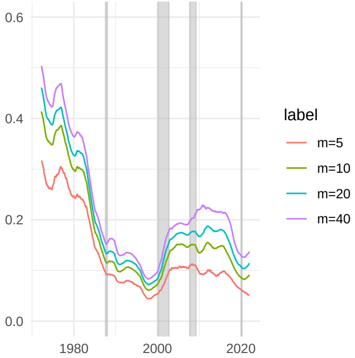
- Left panel rolling. Right panel expanding. Outperforms during major market crashes



OOS realized and predicted explained variance by factors

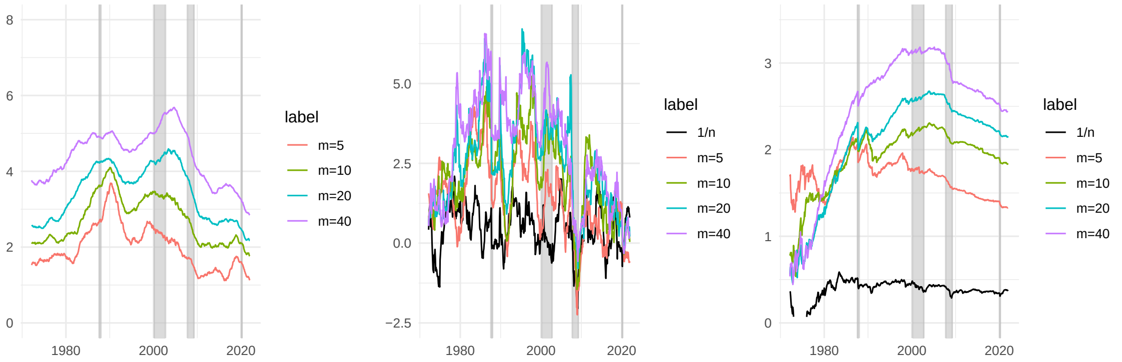
- Left panel: total $R^{2,\boldsymbol{f}}_{t,T,\mathrm{OOS}} := 1 - \frac{1}{T-t}\sum_{s=t}^{T-1} \frac{\|\boldsymbol{x}_{s+1} - \phi^{\mathrm{sy}}(\boldsymbol{z}_s)\boldsymbol{f}_{s+1}\|_2^2}{\|\boldsymbol{x}_{s+1}\|_2^2}$
- Right panel: factor-explained variance $\frac{\mathrm{tr}(\phi^{\mathrm{sy}}(\boldsymbol{z}_t)\mathrm{Cov}_t[\boldsymbol{f}_{t+1}]\phi^{\mathrm{sy}}(\boldsymbol{z}_t)^\top)}{\mathrm{tr}\boldsymbol{\Sigma}_t}$
- Increased factor-explained variance (dependence) during major market crases.
- Idiosyncratic risk explains **more than 75%** of variance on average





OOS realized and predicted Sharpe ratio rolling and expanding

- Realized Sharpe ratio of cMVE portfolio $\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}_t$ vs. $1/N_t$ portfolio
- Left panel: predicted maximal Sharpe ratio.
- Middle panel: realized Sharpe ratio rolling. Right panel: realized Sharpe ratio expanding



Conclusion and Outlook

- Novel nonparametric joint estimator of conditional mean and covariance for (possibly large) unbalanced panels
- Based on moment kernel, given as solution to convex semidefinite problem
- Consistency: estimated covariance matrix is positive semidefinite
- Gives direct estimate of conditional mean-variance efficient portfolio and max Sharpe ratio
- Asymptotic consistency and finite-sample guarantees of empirical estimator
- Simulation study confirms findings
- Further research: higher-order moments

References I

Dawid, A. P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *The Annals of Statistics*, 27(1):65 – 81.

Kozak, S. and Nagel, S. (2024). When do cross-sectional asset pricing factors span the stochastic discount factor? Working Paper 31275, National Bureau of Economic Research.